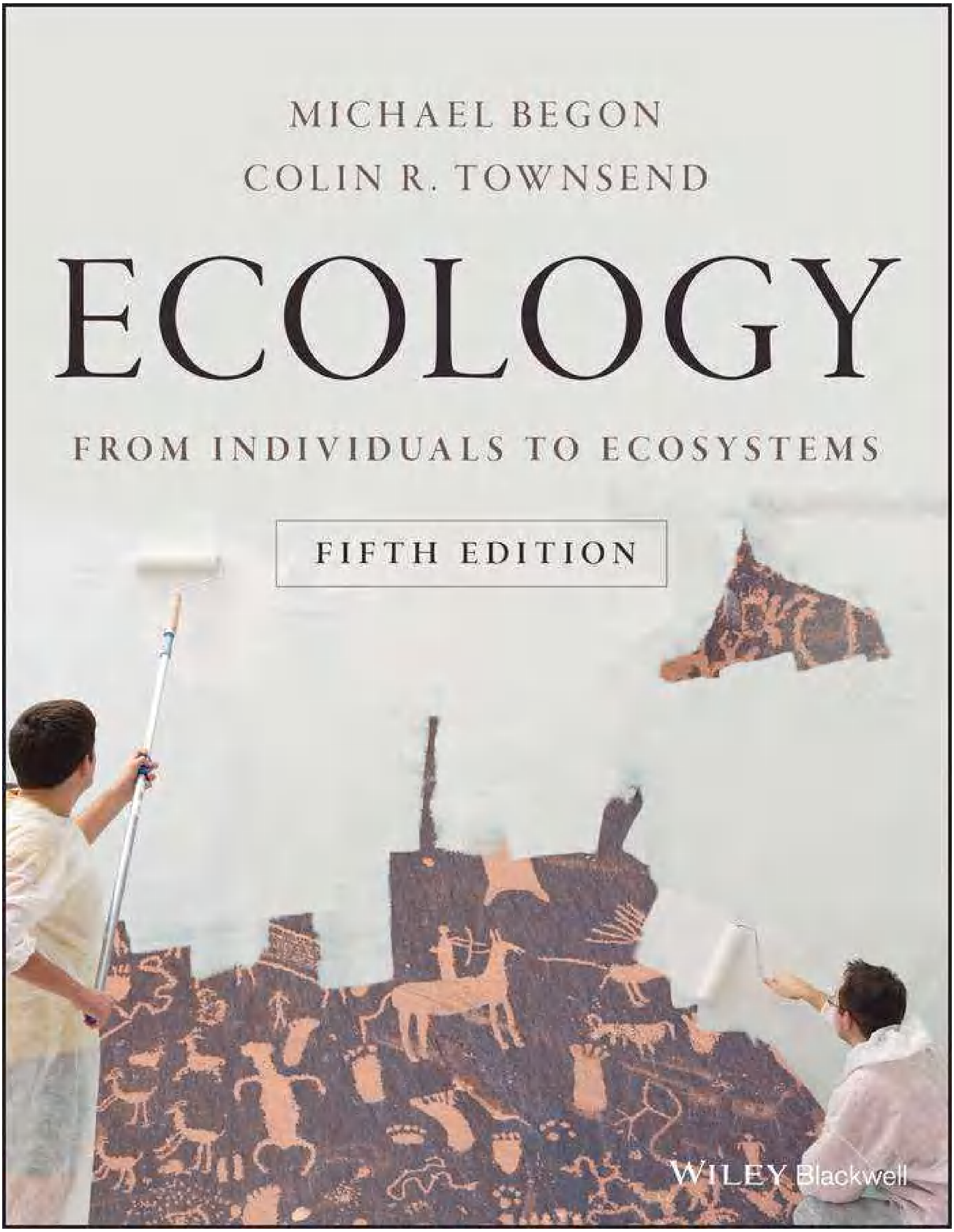
MICHAEL BEGON

COLIN R. TOWNSEND

ECOLOGY

FROM INDIVIDUALS TO ECOSYSTEMS

FIFTH EDITION



WILEY Blackwell

Table of Contents

List of Tables

List of Illustrations

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

Chapter 9

Chapter 10

Chapter 11

Chapter 12

Chapter 13

Chapter 15

Chapter 16

Chapter 17

Chapter 18

Chapter 19

Chapter 20

Chapter 21

Chapter 22

ECOLOGY

From Individuals to Ecosystems

MICHAEL BEGON
University of Liverpool
Liverpool, UK

COLIN R. TOWNSEND
University of Otago
Dunedin, New Zealand

FIFTH EDITION

WILEY

This fifth edition first published 2021 © 2021 by John Wiley & Sons Ltd

Edition History
4e, Wiley, 2006; 3e, Wiley, 1996; 2e, Wiley 1990; 1e, Wiley, 1986

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at http://www.wiley.com/go/permissions.

The right of Michael Begon and Colin R. Townsend to be identified as the authors of this work has been asserted in accordance with law.

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
9600 Garsington Road, Oxford, OX4 2DQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty
While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data

Names: Begon, Michael, author. | Townsend, Colin R., author.
Title: Ecology : from individuals to ecosystems / Michael Begon, The University of Liverpool, Liverpool, UK, Colin Townsend, University of Otago, Dunedin, New Zealand.
Description: Fifth edition. | Hoboken, NJ : Wiley, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020012775 (print) | LCCN 2020012776 (ebook) | ISBN 9781119279358 (paperback) | ISBN 9781119279372 (adobe pdf) | ISBN 9781119279310 (epub)
Subjects: LCSH: Ecology.
Classification: LCC QH541 .B415 2020 (print) | LCC QH541 (ebook) | DDC 577–dc23
LC record available at https://lccn.loc.gov/2020012775
LC ebook record available at https://lccn.loc.gov/2020012776

Cover Design: Wiley
Cover Credit: Decorators painting wall © Dougal Waters/Getty Images
Panel of Native American Newspaper Rock petroglyphs © milehightraveler/Getty Image

This book is dedicated to our families, and especially to our children and grandchildren,
who will inherit the ecosystems we hand on to them –
by Mike to Linda, Jessica, Robert, Carl, Ria and Erica,
and by Colin to Dominic, Brennan, Amelie and Ella

Preface

A science for everybody – but not an easy science

This book is about the distribution and abundance of living organisms, and about the physical, chemical and especially the biological features and interactions that determine these distributions and abundances.

Unlike some other sciences, the subject matter of ecology is apparent to everybody: most people have observed and pondered the natural world. In this sense most people are ecologists of sorts. But ecology is not an easy science. It must deal explicitly with three separate levels of the biological hierarchy – the individual organisms, the populations of organisms, and the communities of populations. What's more, as we shall see, it can't ignore the detailed biology of individuals, or the pervading influences of historical, evolutionary and geological events. It feeds on advances in our knowledge of biochemistry, behaviour, climatology, plate tectonics and so on, but it feeds back to our understanding of vast areas of biology too. One of the fathers of modern evolutionary biology, T. H. Dobzhansky famously claimed, in the middle of the 20th century, that 'Nothing in biology makes sense, except in the light of evolution'. But equally, very little in evolution, and hence in biology as a whole, makes sense, except in the light of ecology.

Ecology, too, has the distinction of being peculiarly confronted with uniqueness: millions of different species, countless billions of genetically distinct individuals, all living and interacting in a varied and ever-changing world. Imagine doing chemistry if water molecules were not precisely the same, always and everywhere – or physics if the speed of light was different at different field sites. The challenge for ecologists, therefore, is to seek patterns and predictions in a way that recognises this uniqueness and complexity, rather than being swamped by it. The mathematician and philosopher Alfred North Whitehead's advice for scientists, offered 100 years ago, is never more apposite than when applied to ecology: 'Seek simplicity – and distrust it'.

Thirty-four years on: the urgent problems facing us

This fifth edition comes fully 14 years after its immediate predecessor and 34 years after the first edition. Much has changed over that time – in ecology, in the world around us, and also, unsurprisingly, in us authors. The first edition had a cave painting as its cover, which we justified in the Preface by arguing that 'ecology, if not the oldest profession, is probably the oldest science', since the most primitive humans had to understand, as a matter of necessity, the dynamics of the environment in which they lived. We pursued that cave painting theme for two further editions, but for the fourth edition, replaced it with its modern equivalent, urban graffiti. This captured the idea that we, as a species, are still driven to broadcast our feelings graphically and publicly, but the cave painting's celebration of nature's bounty had been replaced by an impassioned plea for its protection. Now, 14 years on, the image on our cover, and its message, are darker and more desperate. Instead of focusing on those who are pleading, on nature's behalf, for respect, this time we shine our light on man the destroyer – obliterating the heritage in our cave painting as he is threatening our natural heritage (the inspiration coming from graffiti artist Banksy, who used a similar idea in an art work involving a water blaster). The ecosystem on our front cover is still there – but it is disappearing.

This evolving image reminds us, too, that 34 years ago it seemed acceptable for ecologists to hold a comfortable, detached position, from which the ecological communities around us were simply material for which we sought a scientific understanding. Now, we must accept the immediacy of the environmental problems that threaten us and the responsibility of ecologists to come in from the sidelines and play their full part in addressing these problems. Applying the principles that emerge from the study of ecology is not only a practical necessity, but also as challenging as deriving those principles in the first place. Thus, in this edition, two whole chapters and one half of another are devoted entirely to applied topics, and examples of how ecological principles have been applied to problems facing us appear, and are highlighted, throughout the remaining 19 chapters. Nonetheless, we remain wedded to the belief that environmental action can only ever be as sound as the ecological principles on which it is based. Hence, while we have tried harder than ever to help improve preparedness for addressing the environmental problems of the years ahead, the book remains, in its essence, an exposition of the *science* of ecology.

About this fifth edition

Hence, we have aimed to make this fifth edition an up-to-date guide to ecology *now*. To this end, the results from around 1000 studies have been newly incorporated into the text, most of them published since the fourth edition. Nonetheless, we have resisted the temptation to lengthen the text, mindful that, clichéd as it may be, less is often more. We have also consciously attempted, while including so much modern work, to avoid bandwagons that seem likely to have run into the buffers by the time many will be using the book. Of course, we may also, sadly, have excluded bandwagons that go on to fulfil their promise.

Having said this, we hope, still, that this edition will be of value to all those whose degree programme includes ecology and all who are, in some way, practicing ecologists. Certain aspects of the subject, particularly the mathematical ones, will prove difficult for some, but our coverage is designed to ensure that wherever our readers' strengths lie – in the field or laboratory, in theory or in practice – a balanced and up-to-date view should emerge.

Different chapters of this book contain different proportions of descriptive natural history, physiology, behaviour, rigorous laboratory and field experimentation, careful field monitoring and censusing, and mathematical modelling (a form of simplicity that it is essential to seek but equally essential to distrust). These varying proportions to some extent reflect the progress made in different areas. They also reflect intrinsic differences in various aspects of ecology. Whatever progress is made, ecology will remain a meeting-ground for the naturalist, the experimentalist, the field biologist and the mathematical modeller. We believe that all ecologists should to some extent try to combine all these facets.

Technical and pedagogical features

An important technical feature is the incorporation of marginal notes as signposts throughout the text. These, we hope, will serve a number of purposes. In the first place, they constitute a series of subheadings highlighting the detailed structure of the text. However, because they are numerous and often informative in their own right, they can also be read in sequence along with the conventional subheadings, as an outline of each chapter. They should act too as a revision aid for students – indeed, they are similar to the annotations that students themselves often add to their textbooks. Finally, because the marginal notes generally summarise the take-home message of the paragraph or paragraphs that they accompany, they can act as a continuous assessment of comprehension: if you can see that the signpost is the take-home message of what you have just read, then you have understood.

To highlight the link between the pure science of ecology and the application of this knowledge to the many environmental problems that now face us we have introduced a new feature – a systematic presentation of ecological applications, highlighted in special boxes throughout the text.

Acknowledgments

This is the second major revision we have written as a twosome rather than a trio, and this time the authorship reflects this. Nonetheless, while little remains of the original text, we are deeply

conscious of the debt we owe to John Harper, who sadly died in 2009. We cannot promise to have absorbed or, to be frank, to have accepted, every one of his views, but we hope, in this fifth edition, that we have not strayed too far from the paths along which he guided us. If readers recognise any attempts to stimulate and inspire rather than simply to inform, to question rather than to accept, to respect our readers rather than to patronise them, and to avoid unquestioning obedience to current reputation while acknowledging our debt to the masters of the past, then they will have identified John's intellectual legacy still firmly imprinted on the text.

In previous editions we thanked the great many friends and colleagues who helped us by commenting on various drafts of the text. The effects of their contributions are still strongly evident in the present edition. This fifth edition was also read by a series of reviewers, and we are delighted to be able to acknowledge their help. Thanks to David Atkinson, Richard Bardgett, Rob Brooker, Dylan Childs, Tim Coulson, Hans de Kroon, Andy Fenton, Rob Freckleton, Cristina Garcia, Sue Hartley, Andy Hector, Alan Hildrew, Marcel Holyoak, Dave Hooper, Tony Ives, Xavier Lambin, Steve Long, Michel Loreau, Peter Morin, Asko Noormets, Ilik Saccheri, Ross Thomson, Jake Weiner and Rob Whittaker.

At our publisher Wiley, we are grateful to Ward Cooper (now moved on), who brought us back within the fold, and throughout the writing and production to Sarah Keegan, Jane Andrew, Jane Grisdale and Debbie Maizels.

Mike Begon
Colin Townsend

Introduction: Ecology and its Domain

Definition and scope of ecology

The word ‘ecology’ was first used by Ernest Haeckel in 1866. Paraphrasing Haeckel we can describe ecology as the scientific study of the interactions between organisms and their environment. The word is derived from the Greek *oikos*, meaning ‘home’. Ecology might therefore be thought of as the study of the ‘home life’ of living organisms. A less vague definition was suggested by Krebs (1972): ‘Ecology is the scientific study of the interactions that determine the distribution and abundance of organisms’. Notice that Krebs’ definition does not use the word ‘environment’; to see why, it is necessary to define the word. The environment of an organism consists of all those factors and phenomena outside the organism that influence it, whether these are physical and chemical (abiotic) or other organisms (biotic). The ‘interactions’ in Krebs’ definition are, of course, interactions with these very factors. The environment therefore retains the central position that Haeckel gave it.

Krebs’ definition has the merit of pinpointing the ultimate subject matter of ecology: the distribution and abundance of organisms – *where* organisms occur, *how many* occur there and *why*. This being so, it might be better still to define ecology as:

> the scientific study of the distribution and abundance of organisms and the interactions that determine distribution and abundance.

As far as the subject matter of ecology is concerned, ‘the distribution and abundance of organisms’ is pleasantly succinct. But we need to expand it. The living world can be viewed as a biological hierarchy that starts with subcellular particles, and continues up through cells, tissues and organs. Ecology deals with the next three levels: the individual *organism*, the *population* (consisting of individuals of the same species) and the *community* (consisting of a greater or lesser number of species populations). At the level of the organism, ecology deals with how individuals are affected by (and how they affect) their environment. At the level of the population, ecology is concerned with the presence or absence of particular species, their abundance or rarity, and with the trends and fluctuations in their numbers. Community ecology then deals with the composition and organisation of ecological communities. Ecologists also focus on the pathways followed by energy and matter as these move among living and non-living elements of a further category of organisation: the *ecosystem*, comprising the community together with its physical and chemical environment.

There are two broad approaches that ecologists can take at each level of ecological organisation. First, much can be gained by building from properties at the level below: physiology when studying organismal ecology; individual clutch size and survival probabilities when investigating the dynamics of individual populations of species; food consumption rates when dealing with interactions between predator and prey populations; limits to the similarity of coexisting species when researching communities, and so on. An alternative approach deals directly with properties of the level of interest – for example, niche breadth at the organismal level; relative importance of density-dependent processes at the population level; species diversity at the level of community; rate of biomass production at the ecosystem level – and tries to relate these to abiotic or biotic aspects of the environment. Both approaches have their uses, and both will be used throughout the book.

Explanation, description, prediction and control

At all levels of ecological organisation we can try to do a number of different things. In the first place we can try to *explain* or *understand*. This is a search for knowledge in the pure scientific tradition. Obviously, though, in order to understand something, we must first have a *description* of whatever it is that we wish to understand. This, in itself, adds to our knowledge of the living world. Note, however, that the most valuable descriptions are those carried out with a particular problem or 'need for understanding' in mind. All descriptions are selective: but undirected description, carried out for its own sake, is often found afterwards to have selected the wrong things.

Ecologists also often try to *predict* what will happen to an organism, a population, a community or an ecosystem under a particular set of circumstances: and on the basis of these predictions we may try to *control* the situation. We may try to minimise the effects of locust plagues by predicting when they are likely to occur and taking appropriate action. We may try to protect crops by predicting when conditions will be favourable to the crop and unfavourable to its enemies. We may try to maintain endangered species by predicting the conservation policy that will enable them to persist. And we may try to conserve biodiversity in order to maintain ecosystem 'services' such as the protection of chemical quality of natural waters. Some prediction and control can be carried out without explanation or understanding. But confident predictions, precise predictions and predictions of what will happen in unusual circumstances can be made only when we can explain what is going on. Mathematical modelling has played, and will continue to play, a crucial role in the development of ecology, particularly in our ability to predict outcomes. But it is the real world we are interested in, and the worth of models must always be judged in terms of the light they shed on the working of natural systems.

Note, too, that there are two different classes of explanation in biology: proximal and ultimate explanations. For example, the present distribution and abundance of a particular species of bird may be 'explained' in terms of the physical environment that the bird tolerates, the food that it eats and the parasites and predators that attack it. This is a *proximal* explanation. However, we may also ask how this species of bird comes to have these properties that now appear to govern its life. This question has to be answered by an explanation in evolutionary terms. The *ultimate* explanation of the present distribution and abundance of this bird lies in the ecological experiences of its ancestors. There are many problems in ecology that demand evolutionary, ultimate explanations: 'How have organisms come to possess particular combinations of size, developmental rate, reproductive output and so on?' (Chapter 7), 'What causes predators to adopt particular patterns of foraging behaviour?' (Chapter 9) and 'How does it come about that coexisting species are often similar but rarely the same?' (Chapters 8 and 16). These problems are as much part of modern ecology as are the prevention of plagues, the protection of crops and the preservation of rare species. Our ability to control and exploit ecosystems cannot fail to be improved by an ability to explain and understand. And in the search for understanding, we must combine both proximal and ultimate explanations.

Pure and applied ecology

Ecologists are concerned not only with ecosystems, communities, populations and organisms *in nature*, but also with man-made or human-influenced environments (plantation forests, wheat fields, grain stores, nature reserves and so on), and with the consequences of human influence *on* nature (pollution, overharvesting, the spread of invasive species, global climate change, etc.). Our influence is so pervasive that we would be hard pressed to find an environment that was totally unaffected by human activity. Indeed, moves are afoot to designate a new geological epoch – the Anthropocene – the latest slither of geological history during which people have become a major geological force and the major ecological force around the globe. Environmental problems are now high on the political agenda and ecologists clearly have a central role to play: a sustainable future depends fundamentally on ecological understanding and our ability to predict or produce outcomes under different scenarios.

When the first edition of this text was published in 1986, the majority of ecologists would have classed themselves as pure scientists, defending their right to pursue ecology for its own sake and not wishing to be deflected into narrowly applied projects. The situation has changed dramatically in the past three and a half decades, partly because governments have shifted the focus of grant-awarding bodies towards ecological applications, but also, and more fundamentally, because ecologists have themselves responded to the need to direct much of their research to the many environmental problems that have become ever more pressing. This is recognised in this new edition by a systematic treatment of ecological applications throughout the text. We believe strongly that the application of ecological theory must be based on a sophisticated understanding of the pure science. Thus, our treatment of ecological applications is organised alongside the theory in each chapter.

Chapter 1

Organisms in their Environments: the Evolutionary Backdrop

1.1 Introduction: natural selection and adaptation

From our definition of ecology in the Preface, and even from a layman's understanding of the term, it is clear that at the heart of ecology lies the relationship between organisms and their environments. In this opening chapter we explain how, fundamentally, this is an evolutionary relationship. The great Russian–American biologist Theodosius Dobzhansky famously said: 'Nothing in biology makes sense, except in the light of evolution'. This is as true of ecology as of any other aspect of biology. Thus, we try here to explain the processes by which the properties of different sorts of species make their life possible in particular environments, and also to explain their failure to live in other environments. In mapping out this evolutionary backdrop to the subject, we will also be introducing many of the questions that are taken up in detail in later chapters.

the meaning of adaptation

The phrase that, in everyday speech, is most commonly used to describe the match between organisms and environment is: 'organism X is adapted to' followed by a description of where the organism is found. Thus, we often hear that 'fish are adapted to live in water', or 'cacti are adapted to live in conditions of drought'. In everyday speech, this may mean very little: simply that fish have characteristics that allow them to live in water (and perhaps exclude them from other environments) or that cacti have characteristics that allow them to live where water is scarce. The word 'adapted' here says nothing about how the characteristics were acquired. For an ecologist or evolutionary biologist, however, 'X is adapted to live in Y' means that environment Y has provided forces of natural selection that have affected the life of X's ancestors and so have moulded and specialised the evolution of X. 'Adaptation' means that genetic change has occurred.

Regrettably, though, the word '*ad*aptation' implies that organisms are matched *to* their present environments, suggesting 'design' or even 'prediction'. But organisms have not been designed for, or fitted to, the present: they have been moulded (by *natural selection*) by past environments. Their characteristics reflect the successes and failures of ancestors. They appear to be apt for the environments that they live in at present only because present environments tend to be similar to those of the past.

evolution by natural selection

The theory of evolution by natural selection is an ecological theory. It was first elaborated by Charles Darwin (1859), though its essence was also appreciated by a contemporary and correspondent of Darwin’s, Alfred Russell Wallace (Figure 1.1). It rests on a series of propositions.

1. The individuals that make up a population of a species are *not identical*: they vary, although sometimes only slightly, in size, rate of development, response to temperature, and so on.
2. Some, at least, of this variation is *heritable*. In other words, the characteristics of an individual are determined to some extent by its genetic make-up. Individuals receive their genes from their ancestors and therefore tend to share their characteristics.
3. All populations have the *potential* to populate the whole earth, and they would do so if each individual survived and each individual produced its maximum number of descendants. But they do not: many individuals die prior to reproduction, and most (if not all) reproduce at a less than maximal rate.
4. Different ancestors leave *different numbers of descendants*. This means much more than saying that different individuals produce different numbers of offspring. It includes also the chances of survival of offspring to reproductive age, the survival and reproduction of the progeny of these offspring, the survival and reproduction of their offspring in turn, and so on.
5. Finally, the number of descendants that an individual leaves depends, not entirely but crucially, on *the interaction between the characteristics of the individual and its environment*.

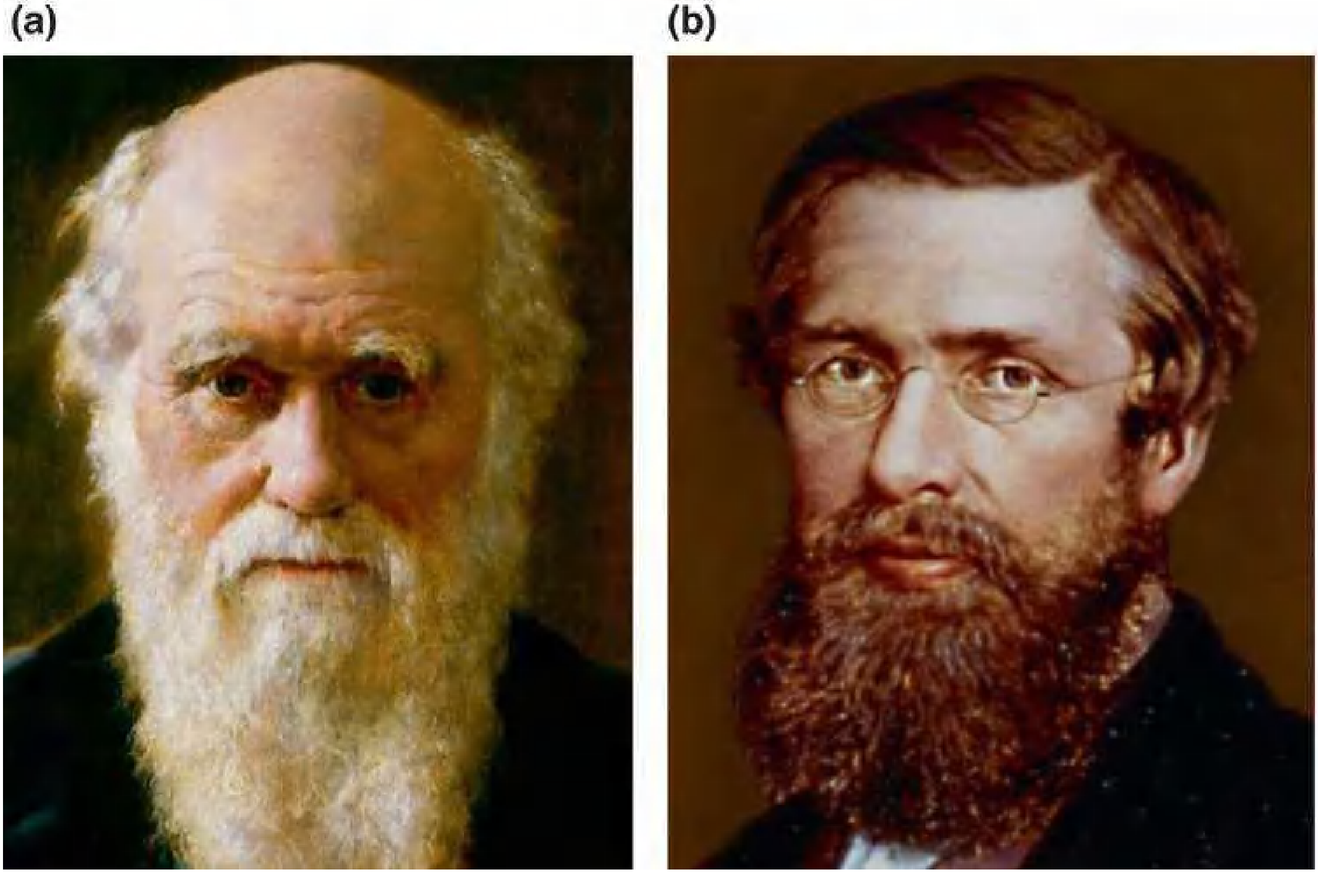


Figure 1.1 The fathers of evolution. (a) Charles Darwin. Detail from painting by John Collier 1883 (National Portrait Gallery RPG 1024). (b) Alfred Russell Wallace. Detail from photograph by Thomas Sims 1869, colourised by Paul Edwards, copyright G. W. Beccaloni.

In any environment, some individuals will tend to survive and reproduce better, and leave more descendants, than others. If, because of this, the heritable characteristics of a population change from generation to generation, then evolution by natural selection is said to have occurred. This is the sense in which nature may loosely be thought of as *selecting*. But nature does not select in the way that plant and animal breeders select. Breeders have a defined end in view – bigger seeds or a faster racehorse. But nature does not *actively* select in this way: it simply sets the scene within which the evolutionary play of differential survival and reproduction is played out.

fitness: it is all relative

The fittest individuals in a population are those that leave the greatest number of descendants. In practice, the term is often applied not to a single individual, but to a typical individual or a type. For example, we may say that in sand dunes, yellow-shelled snails are fitter than brown-shelled snails. *Fitness*, then, is a relative not an absolute term. The fittest individuals in a population are those that leave the greatest number of descendants *relative to* the number of descendants left by other individuals in the population.

evolved perfection? no

When we marvel at the diversity of complex specialisations, there is a temptation to regard each case as an example of evolved perfection. But this would be wrong. The evolutionary process works on the genetic variation that is available. It follows that natural selection is unlikely to lead to the evolution of perfect, 'maximally fit' individuals. Rather, organisms come to match their environments by being 'the fittest available' or 'the fittest yet': they are not 'the best imaginable'. Part of the lack of fit arises because the present properties of an organism have not all originated in an environment similar in every respect to the one in which it now lives. Over the course of its evolutionary history (its phylogeny), an organism's remote ancestors may have evolved a set of characteristics – evolutionary 'baggage' – that subsequently constrain future evolution. For many millions of years, the evolution of vertebrates has been limited to what can be achieved by organisms with a vertebral column. Moreover, much of what we now see as precise matches between an organism and its environment may equally be seen as constraints: koala bears live successfully on *Eucalyptus* foliage, but, from another perspective, koala bears cannot live without *Eucalyptus* foliage.

1.2 Specialisation within species

The natural world is not composed of a continuum of types of organism each grading into the next: we recognise boundaries between one type of organism and another. Nevertheless, within what we recognise as *species* (defined below), there is often considerable variation, and some of this is heritable. It is on such intraspecific variation, after all, that plant and animal breeders – and natural selection – work.

The word '*ecotype*' was first coined for plant populations (Turesson, 1922a, 1922b) to describe genetically determined differences between populations within a species that reflect local matches between the organisms and their environments. But evolution forces the characteristics of populations to diverge from each other only if: (i) there is sufficient heritable variation on which selection can act; and (ii) the forces favouring divergence are strong enough to counteract the mixing and hybridisation of individuals from different sites. Two populations will not diverge completely if their members (or, in the case of plants, their pollen) are continually migrating between them and mixing their genes.

Local, specialised populations become differentiated most conspicuously amongst organisms that are immobile for most of their lives. Motile organisms have a large measure of control over the environment in which they live; they can recoil or retreat from a lethal or unfavourable environment and actively seek another. Sessile, immobile organisms have no such freedom. They must live, or die, in the conditions where they settle. Populations of sessile organisms are therefore exposed to forces of natural selection in a peculiarly intense form.

This contrast is highlighted on the seashore, where the intertidal environment continually oscillates between the terrestrial and the aquatic. The fixed algae, sponges, mussels and barnacles all tolerate life somewhere along the continuum. But the mobile shrimps, crabs and fish track their aquatic habitat as it moves; whilst the shore-feeding birds track their terrestrial habitat. The mobility of such organisms enables them to match their environments to themselves. The immobile organism must match itself to its environment.

1.2.1 Geographic variation within species: ecotypes

geographic variation on a small scale

Differentiation within a species can occur over a remarkably small geographic scale. In the case of sweet vernal grass, *Anthoxanthum odoratum*, growing along a 90 m transition zone between mine and pasture soils at the Trelogan zinc and lead mine in Wales, there was a striking increase in evolved tolerance to zinc, at otherwise toxic concentrations, over a distance of only 3 m within the zone. In this case, any counteracting mixing and hybridisation of the ecotypes was reduced because plants growing on the mine soil tended to flower later than their counterparts in the pasture (Antonovics, 2006).

.... and a large scale

In a study with a much broader geographic range, common frogs (*Rana temporaria*) were monitored over a latitudinal gradient encompassing Sweden and Finland. Geographic variation within species is generally studied both *in situ* and using a ‘common garden’ approach, where individuals from different sites are transplanted and grown together, thus eliminating any influence of immediate environments. In this case, while there was considerable variation in tadpole development time (from complete gill absorption to emergence of the first foreleg), no consistent trend with latitude was evident (Figure 1.2a). However, when tadpoles from different sites were reared in a common environment, at a range of temperatures, those from higher latitudes developed significantly faster. There had clearly been local adaptation, and frogs experiencing colder temperatures (at higher latitudes) had evolved compensatory increases in development rate. The net result was that development times were similar at different latitudes.

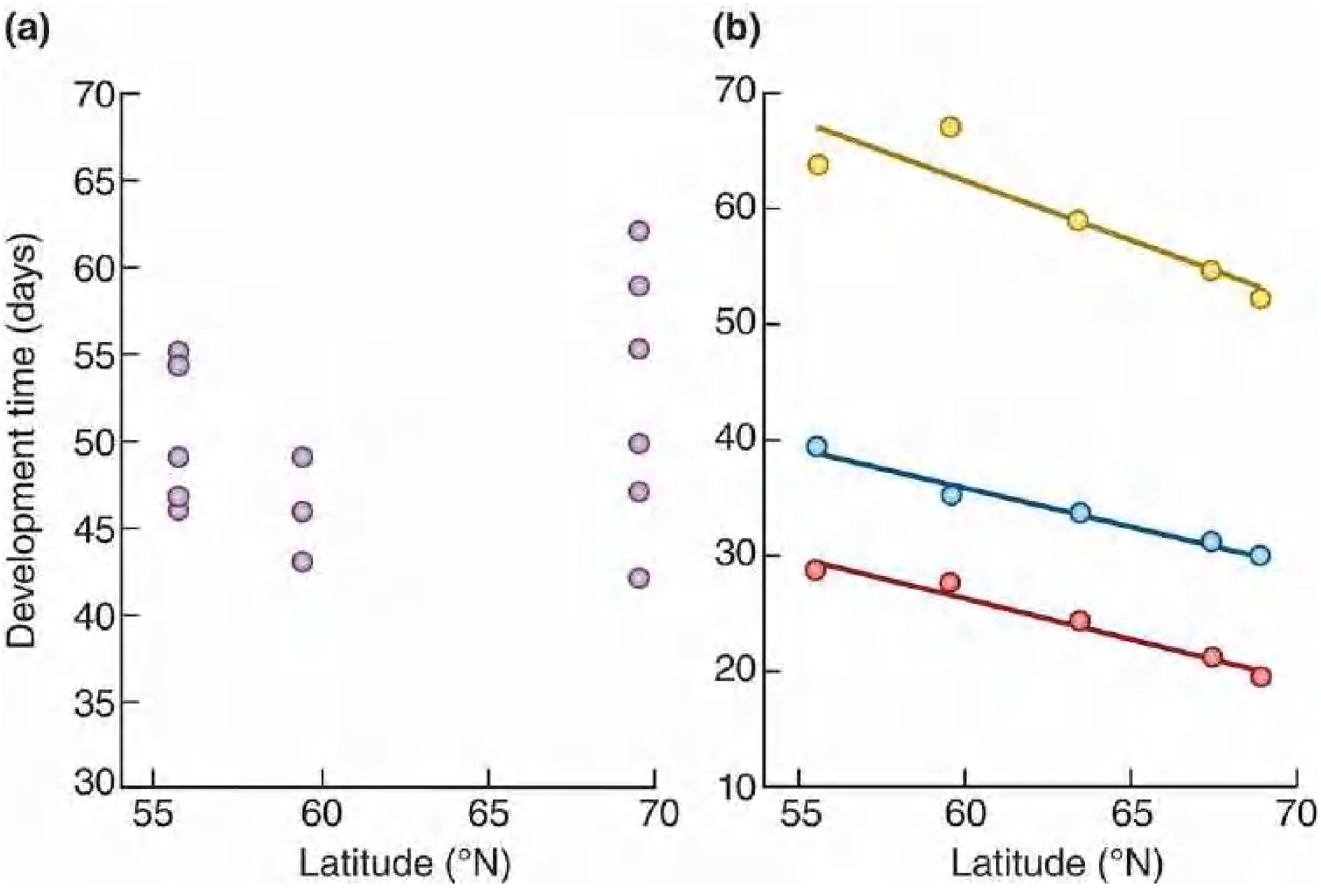


Figure 1.2 **At a given temperature, tadpoles from higher latitudes developed faster than those from lower latitudes.** (a) Tadpoles from ponds in two areas of Sweden, in the south, and from Finland, in the north, showed variation in development times but no consistent trend with latitude. (b) When tadpoles from sites at various latitudes were reared in the laboratory at different temperatures, those from higher latitudes consistently developed fastest. Temperatures: 14°C (yellow circles), 18°C (blue circles), and 22°C (red circles).

Source: From Laugen *et al*. (2003).

APPLICATION 1.1 Selection of ecotypes for conservation

The sapphire rockcress, *Arabis fecunda*, is a rare perennial herb restricted to calcareous soil outcrops in western Montana (USA) – so rare, in fact, that there are just 19 existing populations separated into two groups ('high elevation' and 'low elevation') by a distance of around 100 km. Whether there is local adaptation is of practical importance for conservation: four of the low-elevation populations are under threat from spreading urban areas and may require reintroduction from elsewhere if they are to be sustained. Reintroduction may fail if local adaptation is too marked. Observing plants in their own habitats and checking for differences between them would not tell us if there was local adaptation in the evolutionary sense. Differences may simply be the result of immediate responses to contrasting environments made by plants that are essentially the same. But once again, the 'common garden' approach circumvents this problem. The low-elevation sites were more prone to drought – both the air and the soil were warmer and drier – and the low-elevation plants in the common garden were indeed significantly more drought tolerant (Figure 1.3). More generally, we need to improve our understanding of local adaptation, and its genetic basis, because of their importance for the conservation and restoration of genetic resources, and for crop and animal production, and this is of particular significance in a changing climate (McKay *et al*., 2005; Savolainen *et al*., 2013).

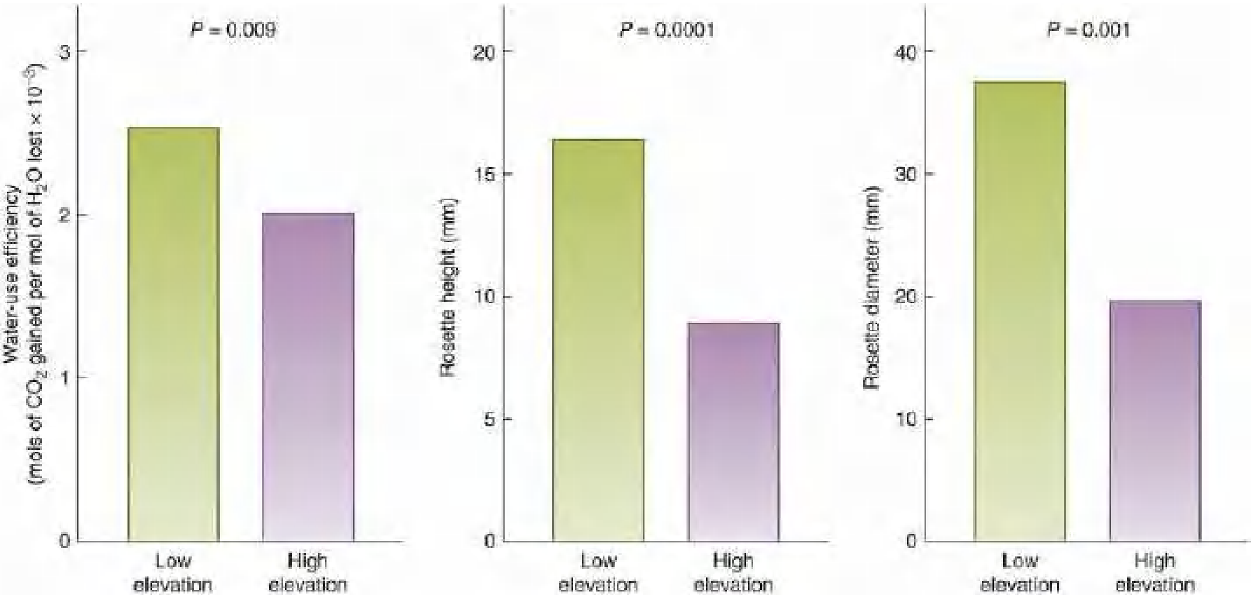


Figure 1.3 Local adaptation of rare sapphire rockcress plants. When plants of the rare sapphire rockcress from low-elevation (drought-prone) and high-elevation sites were grown together in a common garden, there was local adaptation: those from the low-elevation site had significantly better water-use efficiency as well as having both taller and broader rosettes.

Source: From McKay *et al.* (2001).

the balance between local adaptation and hybridisation

On the other hand, local selection by no means always overrides hybridisation. In a study of *Chamaecrista fasciculata*, an annual legume from disturbed habitats in eastern North America, plants were grown in a common garden that had been derived from the 'home' site or were transplanted from distances of 0.1, 1, 10, 100, 1000 and 2000 km (Galloway & Fenster, 2000). The study was replicated three times: in Kansas, Maryland and northern Illinois. Five characteristics were measured: germination, survival, vegetative biomass, fruit production and the number of fruit produced per seed planted. But for all characters in all replicates there was little or no evidence for local adaptation except for transplant distances of 1000 km or more. There is 'local adaptation' – but in this case it was clearly not *that* local.

We can also test whether organisms have evolved to become specialised to life in their local environment in *reciprocal transplant* experiments: comparing their performance when they are grown 'at home' (i.e. in their original habitat) with their performance 'away' (i.e. in the habitat of others). In his meta-analysis of 74 reciprocal transplant studies (50 concerning plants, 21 animals, two fungi and one protist), Hereford (2009) reported that local adaptation was common (71% of studies) but not ubiquitous. On average, local populations had 45% greater fitness than non-local populations. And crucially, there was a small but significant positive association between fitness differences and the magnitude of environmental differences between parental sites ('environmental distance' measured using composite values for up to four environmental variables, such as soil moisture, annual rainfall, elevation and frequency of predation) (Figure 1.4). The magnitude of local adaptation does not seem to be correlated with geographic distance (Leimu & Fischer, 2008), so Hereford's results emphasise the role of ecological factors, not separation itself, as drivers of adaptive differentiation.

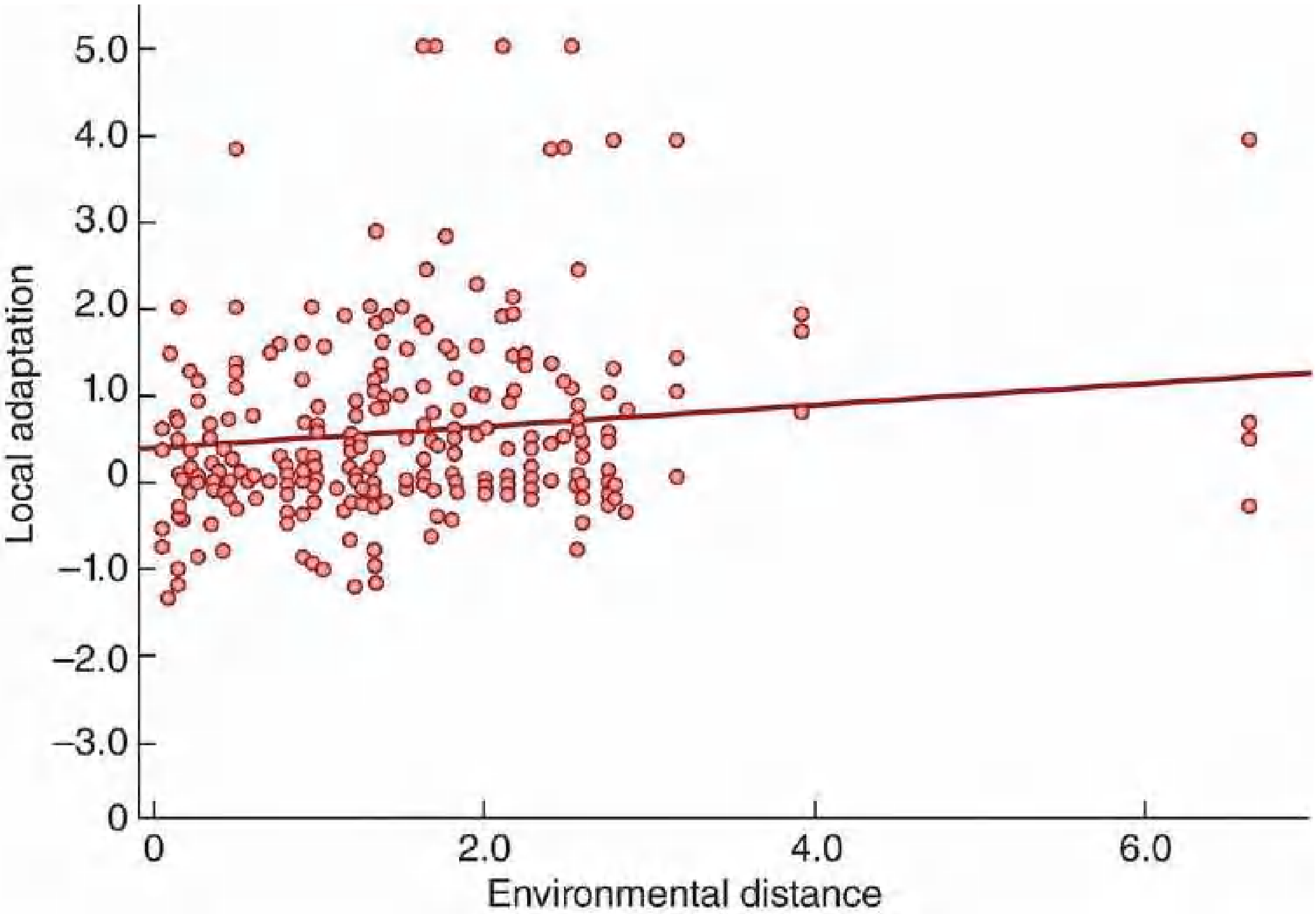


Figure 1.4 Meta-analyses reveal generalities about local adaptation. Regression of local adaptation on environmental distance between sites in a meta-analysis of reciprocal transplant experiments ($P = 0.003$). Local adaptation is the difference in relative fitness between a native population and a non-native population in the native's environment. To standardise measures of environmental difference between sites, Euclidean distances from the means of environmental variables were calculated for all sites in each study.

Source: From Hereford (2009).

1.2.2 Genetic polymorphism

transient polymorphisms

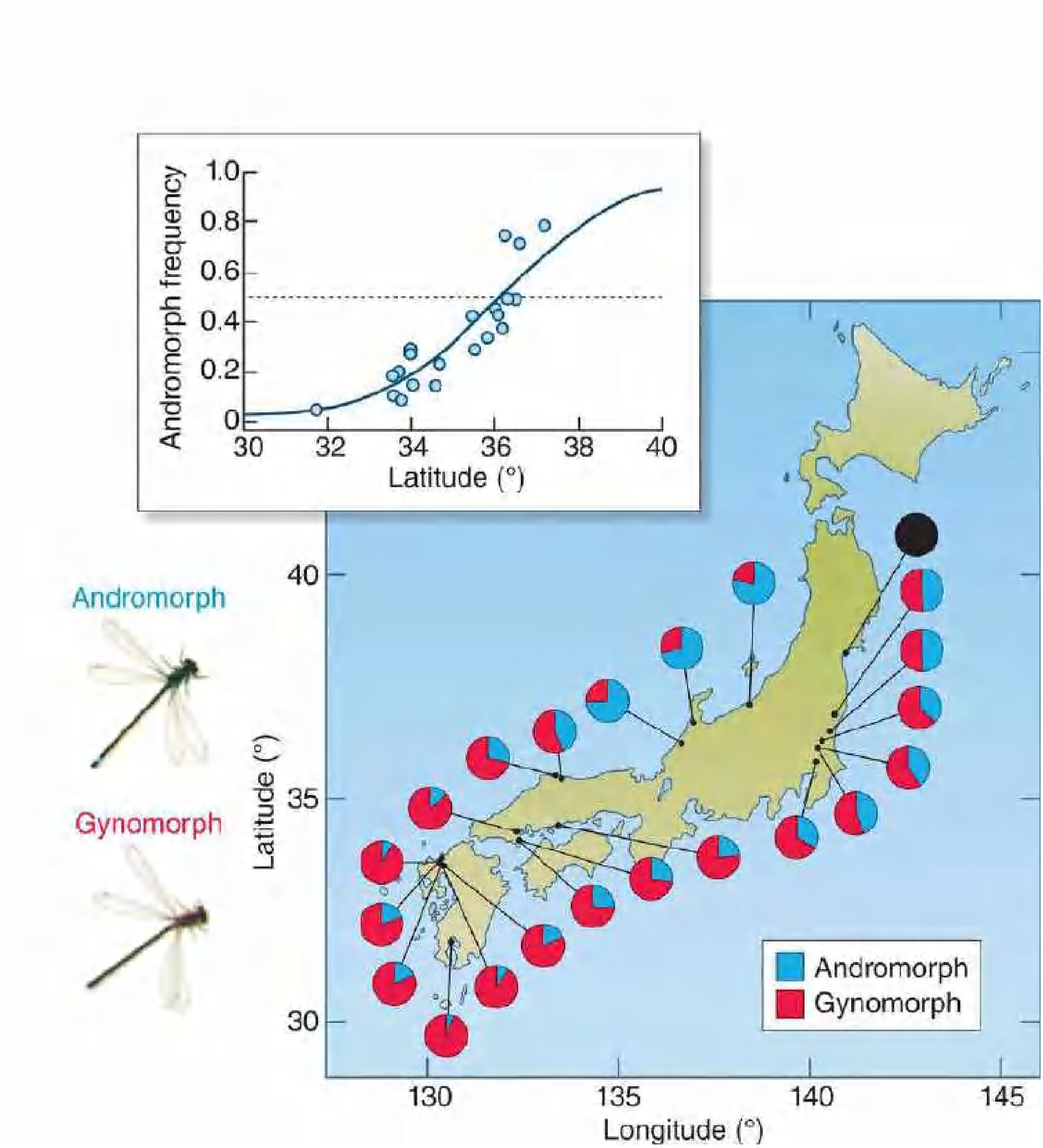
On a finer scale than ecotypes, it may also be possible to detect levels of variation *within* populations. Such variation is known as polymorphism. Specifically, genetic polymorphism is ‘the occurrence together in the same habitat of two or more discontinuous forms of a species in such proportions that the rarest of them cannot merely be maintained by recurrent mutation or immigration’ (Ford, 1940). Not all such variation represents a match between organism and environment. Indeed, some of it may represent a mismatch, if, for example, conditions in a habitat change so that one form is being replaced by another. Such polymorphisms are called transient. As all communities are always changing, much polymorphism that we observe in nature may be transient, representing the extent to which the genetic response of populations to environmental change will always be out of step with the environment and unable to anticipate changing circumstances.

the maintenance of polymorphisms

Many polymorphisms, however, are actively maintained in a population by natural selection, and there are a number of ways in which this may occur.

1. *Heterozygotes may be of superior fitness*, but because of the mechanics of Mendelian genetics they continually generate less fit homozygotes within the population. Such ‘heterosis’ is seen in human sickle-cell anaemia where malaria is prevalent. The malaria parasite attacks red blood cells. The sickle-cell mutation gives rise to red cells that are physiologically imperfect and misshapen. However, sickle-cell heterozygotes are fittest because they suffer only slightly from anaemia and are little affected by malaria, but they continually generate homozygotes who are either dangerously anaemic (two sickle-cell genes) or susceptible to malaria (no sickle-cell genes). Nonetheless, the superior fitness of the heterozygote maintains both types of gene in the population (that is, a polymorphism).
2. *There may be gradients of selective forces* favouring one form (morph) at one end of the gradient, and another form at the other. This can produce polymorphic populations at intermediate positions in the gradient. Females of some damselfly species come in distinct colour morphs: gynomorphs and male-mimicking andromorphs. The andromorph form may provide benefit by reducing harassment of the females by males, allowing more time for foraging, but this may be at the expense of being more obvious to predators (Huang & Reinhard, 2012). Takahashi *et al.* (2011) have described a geographic cline in this polymorphism in *Ischnura senegalensis* over a latitudinal range of 1100 km in Japan (Figure 1.5). Such clines suggest that the fitness advantage of each morph changes differentially along an environmental gradient such that the balance of advantage switches around a mid-point where each phenotype has equal fitness. In this case, Takahashi *et al.* (2011) determined that the reproductive potential of gynomorphs (related to ovariole number, body size and egg volume) was indeed higher in the south and lower in the north compared with andromorphs.
3. *There may be frequency-dependent selection* where each of the morphs of a species is fittest when it is rarest (Clarke & Partridge, 1988). This is believed to be the case when rare colour forms of prey are fit because they go unrecognised and are therefore ignored by their predators.
4. *Selective forces may operate in different directions within different patches* in the population. A striking example of this is provided by a reciprocal transplant study of white clover (*Trifolium repens*) in a field in north Wales. To determine whether the characteristics of individuals matched local features of their environment, Turkington and Harper (1979)

removed plants from marked positions in the field and multiplied them into clones in the common environment of a greenhouse. They then transplanted samples from each clone into the place in the sward of vegetation from which it had originally been taken (as a control), and also to the places from where all the others had been taken (a transplant). The plants were allowed to grow for a year before they were removed, dried and weighed. The mean weight of clover plants transplanted back into their home sites was 0.89 g but at away sites it was only 0.52 g, a statistically highly significant difference. This provides strong, direct evidence that clover clones in the pasture had evolved to become specialised, such that they performed best in their local environment. But all this was going on within a single population, which was therefore polymorphic.



[**Figure 1.5**]() **The frequency of andromorphs of local damselfly populations in Japan increases with latitude.** The inset shows the logistic regression with latitude ($t = 8.15$, df = 21, $P < 0.001$), excluding the northernmost population (solid black plot). This population had been recently established in a newly created pond by gynomorphs, and showed 100% gynomorph frequency because of the founder effect.

Source: From Takahashi *et al.* (2011).

no clear distinction between local ecotypes and a polymorphism

In fact, the distinction between local ecotypes and polymorphic populations is not always a clear one, as illustrated by a study involving the marine snail *Littorina saxatilis*. This common inhabitant of North Atlantic shores is remarkably polymorphic with reproductively isolated ecotypes in microhabitats where crabs are either present and wave action is weak (crab ecotype), or on wave-swept rocky surfaces where waves are strong and crabs are absent (wave ecotype) (Johannesson, 2015). The crab ecotype is large and robust, with a thick shell, a high spire and a relatively small aperture, while the wave ecotype is only about half the size of its crab counterpart, has a thin shell, a relatively large aperture and a low spire (Figure 1.6). The same pattern is observed in different parts of the snail's range and, for example, in both Sweden and Spain, snails of each ecotype are fitter in their native microhabitat than if moved to the other microhabitat. In contact zones, however, snail morphologies represent a continuum from one morph to the other, with all possible intermediate stages. Even though the spatial scale of distribution of the two ecotypes may be very small, the forces of selection are clearly able to outweigh the mixing forces of hybridisation – but it is a moot point whether we should describe this as a small-scale series of local ecotypes or a polymorphic population maintained by a gradient of selection.

Figure 1.6 Contrasting ecotypes of the periwinkle *Littorina saxatilis* from Sweden and Spain. Swedish crab ecotype (top left) and wave ecotype (top right), and Spanish wave ecotype (bottom left) and crab ecotype (bottom right).

Source: From Johannesson (2015).

APPLICATION 1.2 Variation within a species with man-made selection pressures

It is, perhaps, not surprising that some of the most dramatic examples of local specialisation within species (indeed of natural selection in action) have been driven by man-made ecological forces, especially those of environmental pollution. These can provide rapid change under the influence of powerful selection pressures. *Industrial melanism*, for example, is the phenomenon in which black or blackish forms of species have come to dominate populations in industrial areas. In the dark individuals, a dominant gene is typically responsible for producing an excess of the black pigment melanin. Industrial melanism has been reported in most industrialised countries and in more than 100 species of moth.

industrial melanism in the peppered moth

The earliest recorded species to evolve in this way was the peppered moth (*Biston betularia*); the first black specimen in an otherwise pale population was caught in Manchester (UK) in 1848. The mutation event giving rise to industrial melanism is reported to have been the insertion of a transposable element (a DNA sequence that can change its position in the genome) into a gene called *cortex*, which plays a role in early wing development, and is estimated to have occurred in about 1819 (van’t Hof *et al*., 2016). By 1895, about 98% of the Manchester peppered moth population was melanic. Following many more years of pollution, a large-scale survey of pale and melanic forms of the peppered moth in Britain recorded more than 20 000 specimens (Figure 1.7). The winds in Britain are predominantly westerlies, spreading industrial pollutants (especially smoke and sulphur dioxide) toward the east. Melanic forms were concentrated toward the east and were completely absent from the unpolluted western parts of England and Wales, northern Scotland and Ireland. Notice from the figure, though, that many populations were polymorphic: melanic and non-melanic forms coexisted. Thus, the polymorphism seems to be a result both of environments changing (becoming more polluted) – to this extent the polymorphism is transient – and of there being a gradient of selective pressures from the less polluted west to the more polluted east.

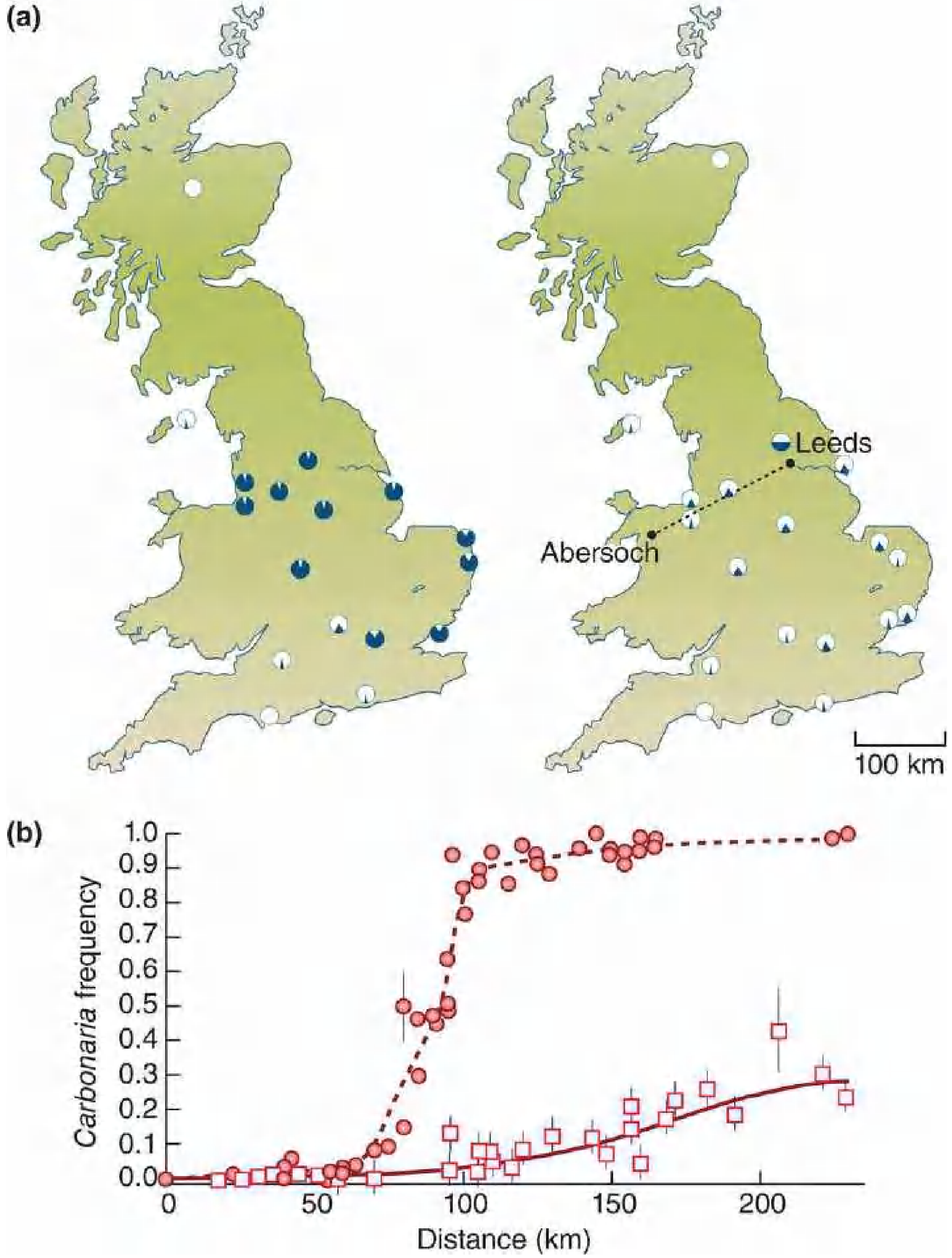

(a)
Leeds
Abersoch
100 km
(b)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Carbonaria frequency
0
50
100
150
200
Distance (km)

Figure 1.7 The frequency of melanic forms of the peppered moth in western Britain was high during the height of pollution from the burning of coal, but that frequency declined after the passing of smoke-free legislation. (a) The distribution of melanic (*carbonaria*) and pale forms (blue and white portions of the pie diagrams, respectively) of the peppered moth, *Biston betularia*, for 1952–56 (left) and 1996 (right), for sites where a comparison between the two periods could be made. The dotted line shows the transect examined in (b). (b) Clines in the frequency of the melanic form along a transect running WSW to NSE from Abersoch in Wales to Leeds in England for the periods 1964–75 (filled circles) and 2002 (open squares). Bars are SEs.

Source: (a) After Grant *et al.* (1998). (b) After Saccheri *et al.* (2008).

The overriding selective pressure appears to be applied by birds that prey on the moths. In field experiments, large numbers of melanic and pale ('typical') moths were reared and released in equal numbers. In a rural and largely unpolluted area of southern England, most of those captured by birds were melanic. In an industrial area near the city of Birmingham, most were typicals (Kettlewell, 1955). Any idea, however, that melanic forms were favoured simply because they were camouflaged against smoke-stained backgrounds in the polluted areas (and typicals were favoured in unpolluted areas because they were camouflaged against pale backgrounds) may be only part of the story. The moths rest on lateral branches or tree trunks during the day, and non-melanic moths are well hidden against a background of mosses and lichens, especially on tree trunks. Industrial pollution has not just blackened the moths' background; sulphur dioxide, especially, has also destroyed most of the moss and lichen on the tree trunks. Thus, sulphur dioxide pollution may have been as important as smoke in selecting melanic moths. The distribution patterns are probably also influenced to some extent by migration between sites that differ in pollution levels (male moths can move 2 km in a night while newly emerged larvae spin threads that might carry them away from the oviposition sites over even greater distances) and there may be some non-visual advantage of melanics over typicals, but this must be weaker than the visual disadvantage associated with predation in a polluted environment (Cook & Saccheri, 2013).

reversing man-made selection pressures

In the 1960s, industrialised environments in Western Europe and the USA started to change again, as oil and electricity began to replace coal, and legislation was passed to impose smoke-free zones and to reduce industrial emissions of sulphur dioxide. The frequency of melanic forms then fell back to near preindustrial levels with remarkable speed (Figure 1.7). Again, there was transient polymorphism – but this time populations were heading in the other direction as pollution was declining.

It is heartening to note that sometimes the consequences of anthropogenic pressures can be reversed if appropriate action is taken.

1.3 Speciation

It is clear, then, that natural selection can force populations of plants and animals to change their character – to evolve. But none of the examples we have considered so far has involved the evolution of a new species. What, then, justifies naming two populations as different species? And what is the process – 'speciation' – by which two or more new species are formed from one original species?

1.3.1 What do we mean by a 'species'?

biospecies: the Mayr–Dobzhansky test

Cynics have said, with some truth, that a species is what a competent taxonomist regards as a species. On the other hand, back in the 1930s two American biologists, Mayr and Dobzhansky, proposed an empirical test that could be used to decide whether two populations were part of the same species or of two different species. They recognised organisms as being members of a single species if they could, at least potentially, breed together in nature to produce fertile offspring. They called a species tested and defined in this way a *biological species* or *biospecies*. In the examples that we have used earlier in this chapter, we know that melanic and normal peppered moths can mate and that the offspring are fully fertile; this is also true of *Anthoxanthum* plants from different positions along the gradient at the Trelogan mine. They are all variations within species – not separate species.

In practice, however, biologists do not apply the Mayr–Dobzhansky test before they recognise every species: there is simply not enough time or resources, and in any case, there are vast portions of the living world – most microorganisms, for example – where an absence of sexual reproduction makes a strict interbreeding criterion inappropriate. What is more important is that the test recognises a crucial element in the evolutionary process that we have met already in considering specialisation within species. If the members of two populations are able to hybridise, and their genes are combined and reassorted in their progeny, then natural selection can never make them truly distinct. Although natural selection may tend to force a population to evolve into two or more distinct forms, sexual reproduction and hybridisation mix them up again.

1.3.2 Allopatric speciation

Allopatric speciation is speciation driven by divergent natural selection in distinct subpopulations in different places. The most orthodox scenario for this comprises a number of stages (Figure 1.8). First, two subpopulations become geographically isolated and natural selection drives genetic adaptation to their local environments. Next, as a *byproduct* of this genetic differentiation, a degree of reproductive isolation builds up between the two. This may be ‘pre-zygotic’, tending to prevent mating in the first place (e.g. differences in courtship ritual), or ‘post-zygotic’: reduced viability, perhaps inviability, of the offspring themselves. Then, in a phase of ‘*secondary contact*’, the two subpopulations re-meet. The hybrids between individuals from the different subpopulations are now of low fitness, because they are literally neither one thing nor the other. Natural selection will then favour any feature in either subpopulation that *reinforces* reproductive isolation, especially pre-zygotic characteristics, preventing the production of low-fitness hybrid offspring. These breeding barriers then cement the distinction between what have now become separate species.

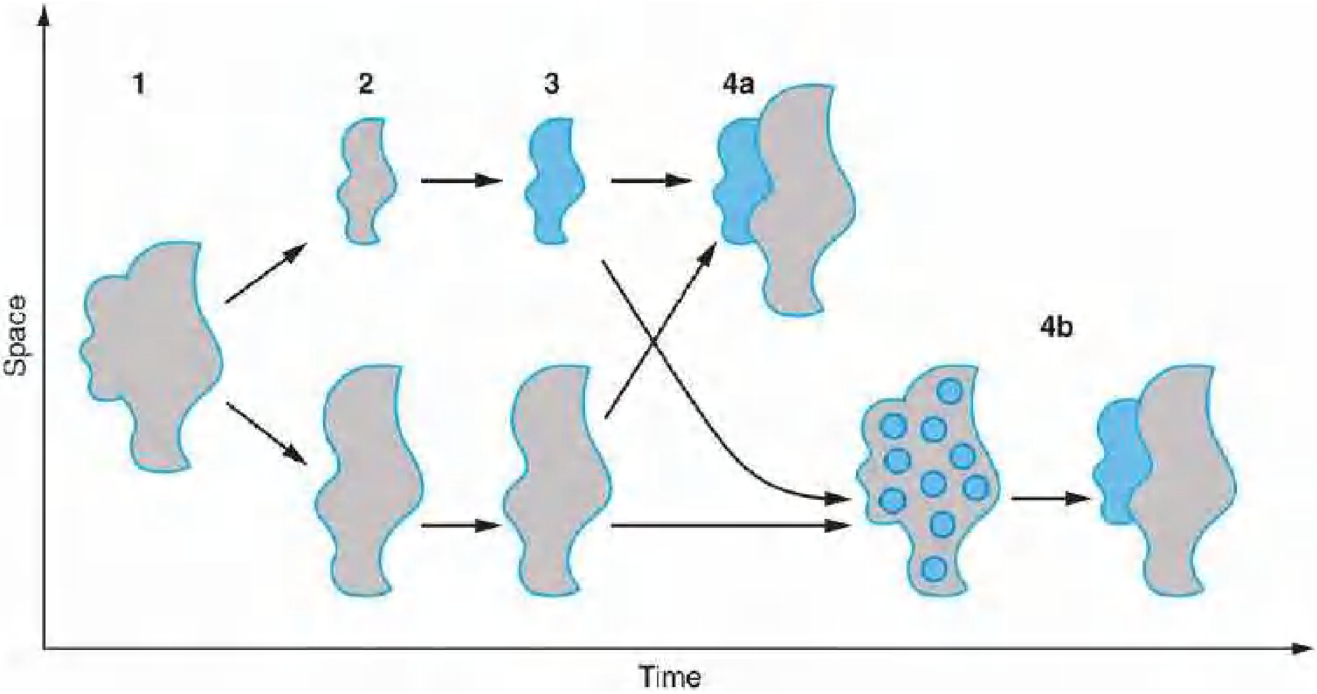


Figure 1.8 **The orthodox picture of ecological speciation.** A uniform species with a large range (1) differentiates (2) into subpopulations (for example, separated by geographic barriers or dispersed onto different islands), which become genetically isolated from each other. (3) After evolution in isolation they may meet again, when they are either already unable to hybridise (4a) and have become true biospecies, or they produce hybrids of lower fitness (4b), in which case evolution may favour features that prevent interbreeding between the 'emerging species' until they are true biospecies.

Darwin's finches

The isolation of islands provides arguably the most favourable scenario for populations to diverge into distinct species. The most celebrated example is the case of Darwin's finches in the Galápagos archipelago, a group of volcanic islands isolated in the Pacific Ocean about 1000 km west of Ecuador and 750 km from the island of Cocos, which is itself 500 km from Central America (Figure 1.9). At more than 500 m above sea level the vegetation is open grassland. Below this is a humid zone of forest that grades into a coastal strip of desert vegetation with some endemic species of prickly pear cactus (*Opuntia*). Fourteen species of finch are found on the islands. The evolutionary relationships amongst them have been traced by molecular techniques using microsatellite DNA that have confirmed the long-held view that the family tree of the Galápagos finches radiated from a single trunk: a single ancestral species that invaded the islands from the mainland of Central America. The molecular data also provide strong evidence that the warbler finch (*Certhidea olivacea*) was the first to split off from the founding group and is likely to be the most similar to the original colonist ancestors. The entire process of evolutionary divergence of these species appears to have happened in less than 3 million years.

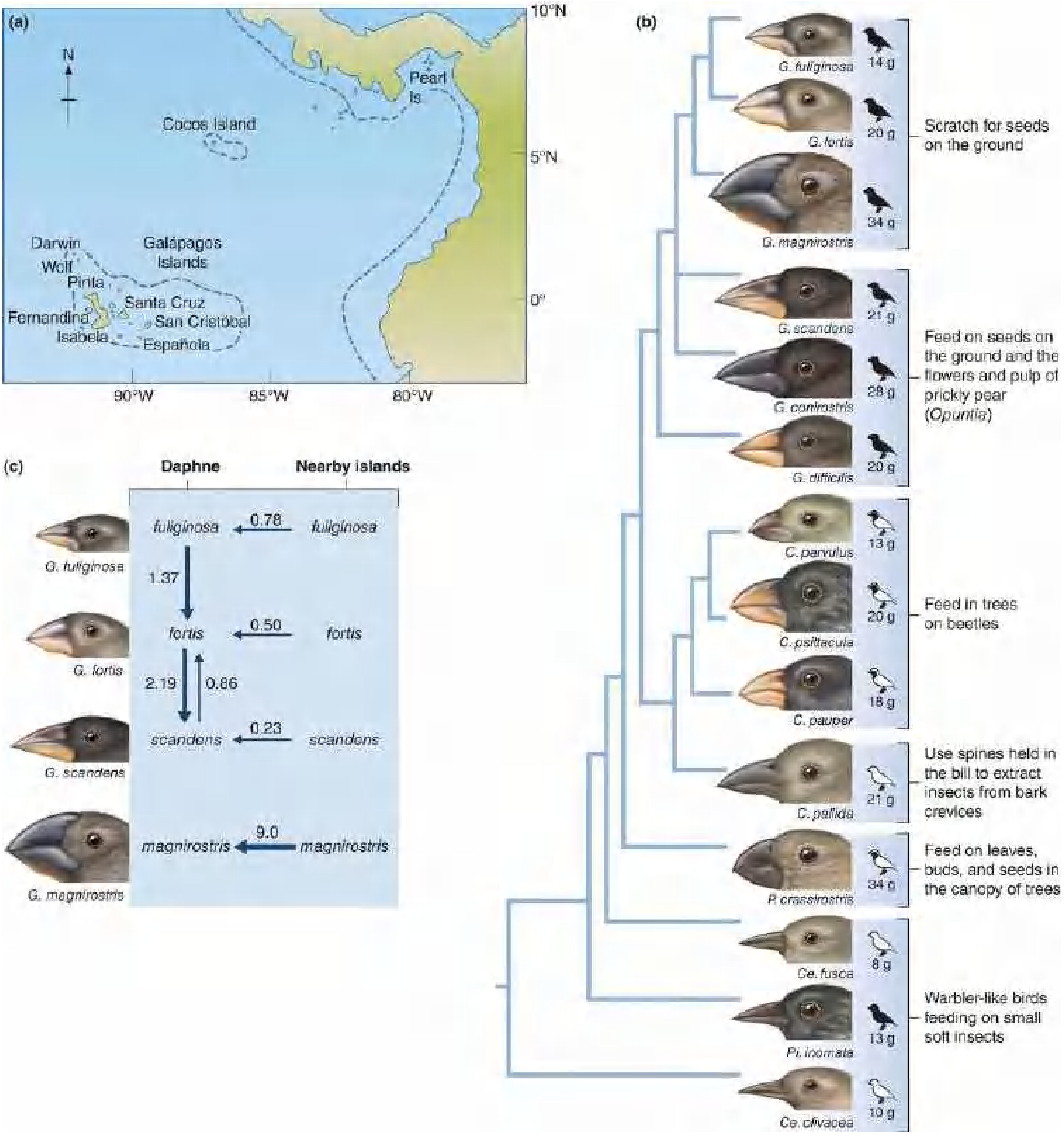

(a)
N
10°N
5°N
0°
90°W
85°W
80°W
Pearl Is
Cocos Island
Darwin
Wolf
Pinta
Galápagos Islands
Santa Cruz
Fernandina
San Cristóbal
Isabela
Española
(b)
G. fuliginosa
14 g
G. fortis
20 g
G. magnirostris
34 g
Scratch for seeds on the ground
G. scandens
21 g
G. conirostris
28 g
G. difficilis
20 g
Feed on seeds on the ground and the flowers and pulp of prickly pear (Opuntia)
C. parvulus
13 g
C. psittacula
20 g
C. pauper
18 g
Feed in trees on beetles
C. pallida
21 g
Use spines held in the bill to extract insects from bark crevices
P. crassirostris
34 g
Feed on leaves, buds, and seeds in the canopy of trees
Ce. fusca
8 g
Pi. inornata
13 g
Warbler-like birds feeding on small soft insects
Ce. olivacea
10 g
(c)
Daphne
Nearby islands
G. fuliginosa
fuliginosa
0.78
fuliginosa
1.37
G. fortis
fortis
0.50
fortis
2.19
0.66
G. scandens
scandens
0.23
scandens
G. magnirostris
magnirostris
9.0
magnirostris

Figure 1.9 Many different species of Darwin's finches have evolved on the Galápagos Islands. (a) Map of the Galápagos Islands showing their position relative to Central America; on the equator 5° equals approximately 560 km. (b) A reconstruction of the evolutionary history of the Galápagos finches based on variation in the length of microsatellite DNA. (A microsatellite is a tract of repetitive DNA in which certain DNA motifs, ranging in length from 2 to 5 base pairs, are repeated, with the number of repeats varying in alleles of individuals.) A measure of the genetic difference between species is shown by the length of the horizontal lines. The feeding habits of the various species are also shown. Drawings of the birds' heads are proportional to actual body size. The maximum amount of black colouring in male plumage and the average body mass are shown for each species. *C*, *Camarhynchus*; *Ce*, *Certhidea*; *G*, *Geospiza*; *P*, *Platyspiza*; *Pi*, *Pinaroloxias*. (c) Gene flow for the four species on Daphne Major, through interbreeding with other species on the island and with immigrants of the same and other species from the nearby islands. Flow is measured as the effective number of individuals per generation. For genes to flow, the first-generation hybrid offspring must themselves mate with one of the parental species. Genes flow from *G. fortis* to *G. scandens* when the hybrid sings the *G. scandens* song (because its father did) and vice versa for genes flowing from *G. scandens* to *G. fortis*. The population of *G. fulginosa* on Daphne Major is very small, and hence the flow of genes into *G. fortis* comes from immigrants from other islands.

Source: (b) After Petren *et al.* (1999). (c) After Grant & Grant (2010).

Isolation – both of the archipelago itself and of individual islands within it – has led to an original evolutionary line radiating into a series of species, each matching its own environment. Populations of ancestor species became reproductively isolated, most likely after chance colonisation of different islands within the archipelago, and evolved separately for a time. Secondary contact phases subsequently occurred as a result of movements between islands that brought non-hybridising biospecies together that then evolved to fill different niches that elsewhere in the world are filled by quite unrelated species. Members of one group, including *Geospiza fuliginosa* and *G. fortis*, have strong bills and hop and scratch for seeds on the ground. *G. scandens* has a narrower and slightly longer bill and feeds on the flowers and pulp of the prickly pears as well as on seeds. Finches of a third group have parrot-like bills and feed on leaves, buds, flowers and fruits, and a fourth group with a parrot-like bill (*Camarhynchus psittacula*) has become insectivorous, feeding on beetles and other insects in the canopy of trees. A so-called woodpecker finch, *Camarhynchus* (*Cactospiza*) *pallida*, extracts insects from crevices by holding a spine or a twig in its bill, while yet a further group includes the warbler finch, which flits around actively and collects small insects in the forest canopy and in the air.

However, the biospecies compartments are not water-tight. A study of the four species on the small island of Daphne Major, and of their possible interbreeding with birds from larger nearby islands, again using molecular techniques, is summarised in Figure 1.9c. The two most abundant species, *Geospiza fortis* and *G. scandens*, were subject to a greater flow of genes between one another than they were to genes from immigrants of their own species from other islands. Indeed, in the case of *G. fortis*, there was also a substantial flow of genes from *G. fuliginosa* immigrants from other islands. Thus, the 'ideal' of gene flow within a species but not between them is not borne out by the data. But the fact that there are 'grey areas' partway through the process does not diminish the importance of either the process of speciation or the concept of biospecies.

ring species – perfect examples of speciation in action, but why so rare?

That speciation is a process rather than an event is beautifully illustrated by the existence of ring species. In these, races or subspecies of a species that fall short of being full species themselves (i.e. distinct forms that are nonetheless capable of producing fertile hybrids) are arranged along a geographic gradient in such a way that the two ends of the gradient themselves meet, hence forming a ring, and where they do, they behave as good species despite being linked, back around the ring, by the series of interbreeding races. Thus, what would normally be a temporal sequence of events, that we can only presume to have happened, becomes frozen in space. That the phenomenon is theoretically feasible has been demonstrated using mathematical models (e.g. de Brito Martins & de Aguiar, 2016). But actual examples are rare, and several that have been proposed in the past have been called into question by modern molecular studies, leading Pereira and Wake (2015) to wonder whether ring species are an unfulfilled promise or, worse still, wish-fulfilment fantasy.

The classic example is the extraordinary case of two species of sea gull. The lesser black-backed gull (*Larus fuscus*) originated in Siberia and colonised progressively to the west, forming a chain or *cline* of different forms, spreading from Siberia to Britain and Iceland. The neighbouring forms along the cline are distinctive, but were assumed to hybridise readily in nature. Neighbouring populations are regarded as part of the same species and taxonomists give them only 'subspecific' status (e.g. *L. fuscus graellsii*, *L. fuscus fuscus*). Populations of the gull have, however, also spread east from Siberia, again forming a cline of freely hybridising forms. Together, the populations spreading east and west encircle the northern hemisphere. They meet and overlap in northern Europe. There, the eastward and westward clines have diverged so far that it is easy to tell them apart, and they are recognised as two different *species*, the lesser black-backed gull (*L. fuscus*) and the herring gull (*L. argentatus*). Moreover, the two species do not hybridise: they have become true biospecies. In this remarkable example, then, we can see how two distinct species seem to have evolved from one primal stock, and that the stages of their divergence remain frozen in the cline that connects them.

However, modern molecular techniques to determine genetic relationships have revealed a more complex picture. Thus, while ancestral populations expanded in a roughly circular fashion, there have been intermittent periods of allopatric fragmentation and subsequent range expansion, leading to areas of secondary contact where hybridisation currently occurs. Population divergence, therefore, proceeded at least partly in allopatry, not exclusively through isolation by distance throughout a contiguous range, as the ring species concept requires. Moreover, adjacent subspecies have been found not necessarily to be each other's closest relatives and evidence is lacking of closure of the circumpolar ring by colonisation of Europe by North American herring gulls, a cornerstone of the ring species concept (Martens & Packert, 2007).

A more convincing example involves bulbuls in the genus *Alophoixus* in montane habitats of the Indo-Malayan bioregion. Fuchs *et al.* (2015) have shown that diversification is consistent with most criteria expected for ring species (Figure 1.10a). First, molecular analysis shows that the seven taxa (Figure 1.10b) are all descendants of a single ancestral species, and probably derive from a single colonisation from Sundaland. Second, neighbouring taxa are most closely related, suggesting that taxa have diverged from a stepping stone colonisation of the high-elevation forest around Thailand's lowlands (lowland 'barriers' A and B in Figure 1.10a). The current distribution suggests that divergence can be explained by isolation by distance, as assumed by the ring species concept (but also, partly, by periods of geographic isolation that probably occurred during climatic cycles following initial diversification of the complex). Third, gene flow between neighbouring taxa suggests that divergence and secondary contact between taxa around the ring have resulted in genetic intergradation. And fourth, demographic analyses indicate a recent expansion and geographic overlap of the oldest taxon (1) and its most distant relative (7), leading to closure of the ring. However, hybrids sampled at the terminus of the ring (where taxon 1 meets taxon 7) indicate that divergence has not been sufficient for complete reproductive isolation to evolve.

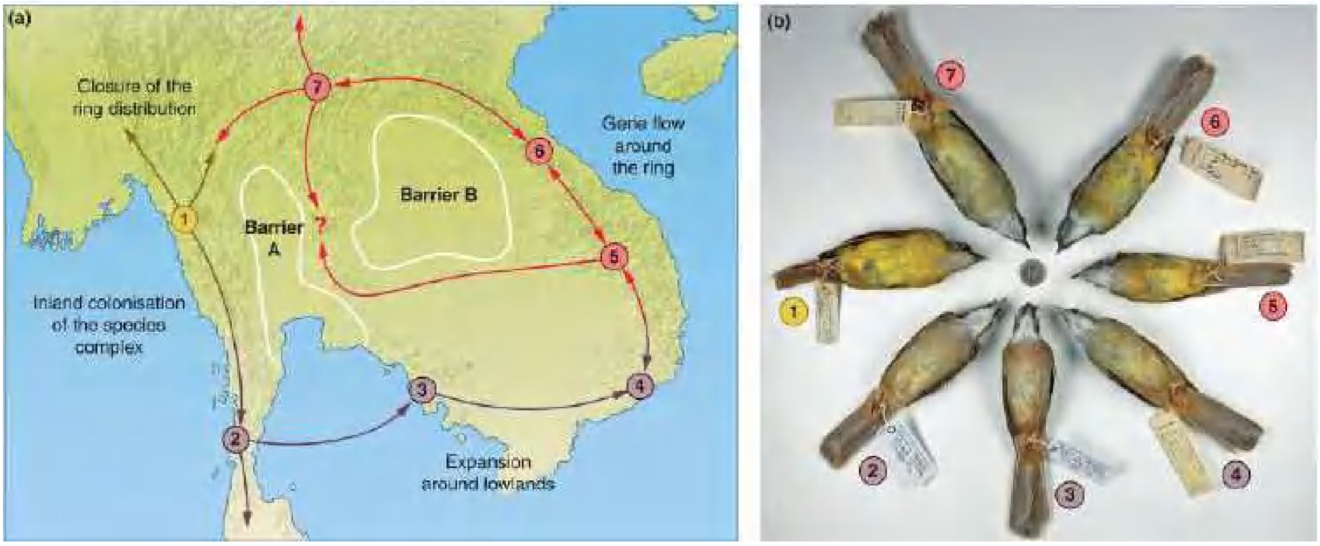


Figure 1.10 **Closure of a ring distribution of bulbul morphotypes.** (a) Distribution of *Alophoixus* bulbuls in the Indo-Malayan bioregion. Taxa composing the *Alophoixus* ring are represented by circles (colours distinguish three currently recognised species); single arrows represent colonisation around the barrier; double arrows represent zones of genetic intergradation; closure of the ring (involving taxa 1 and 7) is shown at the top left (the question mark indicates a possible secondary contact at the mid-ring involving taxa 5 and 7). (b) Eco-morphotypes: (1) *A. flaveolus*, (2) *A. ochraceus ochraceus*, (3) *A. o. cambodianus*, (4) *A. o. hallae*, (5) *A. pallidus khmerensis*, (6) *A. p. annamensis* and (7) *A. p. henrici*.

Source: From Fuchs *et al.* (2015), after Pereira & Wake (2015). (b) Photo credit: A. Previato, MNHN.

allopatric speciation without secondary contact

It would be wrong to imagine that all examples of speciation conform fully to the orthodox picture described in Figure 1.8. In fact, there may never be secondary contact. This would be pure ‘allopatric’ speciation; that is, with all divergence occurring in subpopulations in *different* places. This seems particularly likely for island populations and helps explain the preponderance of endemic species (those found nowhere else) on remote islands.

1.3.3 Sympatric speciation

Furthermore, the advent of modern molecular techniques has spurred interest in the view that an allopatric phase may not be necessary: that is, ‘sympatric’ speciation is possible, with subpopulations diverging despite not being geographically separated from one another. Sympatric speciation has long fascinated evolutionary biologists because it sets diversifying selection against the tendency of sexual reproduction to homogenise populations. There are few truly convincing cases in nature, and indeed it is to be expected that examples of such a process will be difficult to identify because, for most groups, range maps are incomplete, the patterns of habitat use are poorly known and phylogenies do not include all species (Santini *et al.*, 2012). Once again, however, mathematical models provide a way of testing the viability of alternative speciation scenarios and suggest the criteria that need to be satisfied (Bird *et al.*, 2012). There are at least five criteria for inferring that a particular case is best explained by sympatric speciation – four proposed by Coyne and Orr (2004): (1) the two species must have largely overlapping geographical distributions; (2) speciation must be complete; (3) the two species must be sister species (descended from a common ancestor); and (4) the biogeographic and evolutionary history of the groups must make the existence of an allopatric phase ‘very unlikely’; and a fifth, based on a population genetics rather than biogeographic perspective: (5) evidence must support panmixia of the ancestral population (i.e. mating must have been possible between all potential partners) (Fitzpatrick *et al*., 2008.

A good example is provided by two species of cichlid fish in Nicaragua: the Midas cichlid *Amphilophus citrinellus* and the arrow cichlid *A. zaliosus* (Figure 1.11a) (Barluenga *et al*., 2006). These species coexist in the small, isolated Lake Apoyo (satisfying criterion 1), which is relatively homogeneous in terms of habitat, and of recent origin (less than 23 000 years). *A. zaliosus* is found nowhere else, while *A. citrinellus* occurs in many water bodies in the region, including the largest. A variety of behavioural (mate choice) and genetic evidence, including that from microsatellite DNA, indicates that the two species in Lake Apoyo are reproductively isolated from one another (satisfying criterion 2) and indeed from *A. citrinellus* in other lakes (Figure 1.11b). Further genetic evidence from mitochondrial DNA (passed by mothers to their offspring) indicates that the cichlids of Lake Apoyo, of both species, had a single common ancestor arising from the much more widespread stock of *A. citrinellus* (Figure 1.11c) (satisfying criteria 3 and 5). The common ancestor was a high-bodied benthic species but *A. zaliosus*, the new elongated pelagic species, has evolved in less than 10 000 years. Now *A. citrinellus* and *A. zaliosus* in Lake Apoyo are morphologically distinct from one another and have substantially different diets: both feed on biofilm but *A. citrinellus* feeds more from the benthic environment (algae, insects and fish along the lake shore and bed) while *A. zaliosus* feeds more from open water and the surface (including winged insects; Figure 1.11d). There seems little doubt, therefore, that this speciation must have occurred sympatrically, presumably driven by the divergent selection to specialise on bottom-feeding in the one case and on open water-feeding in the other.

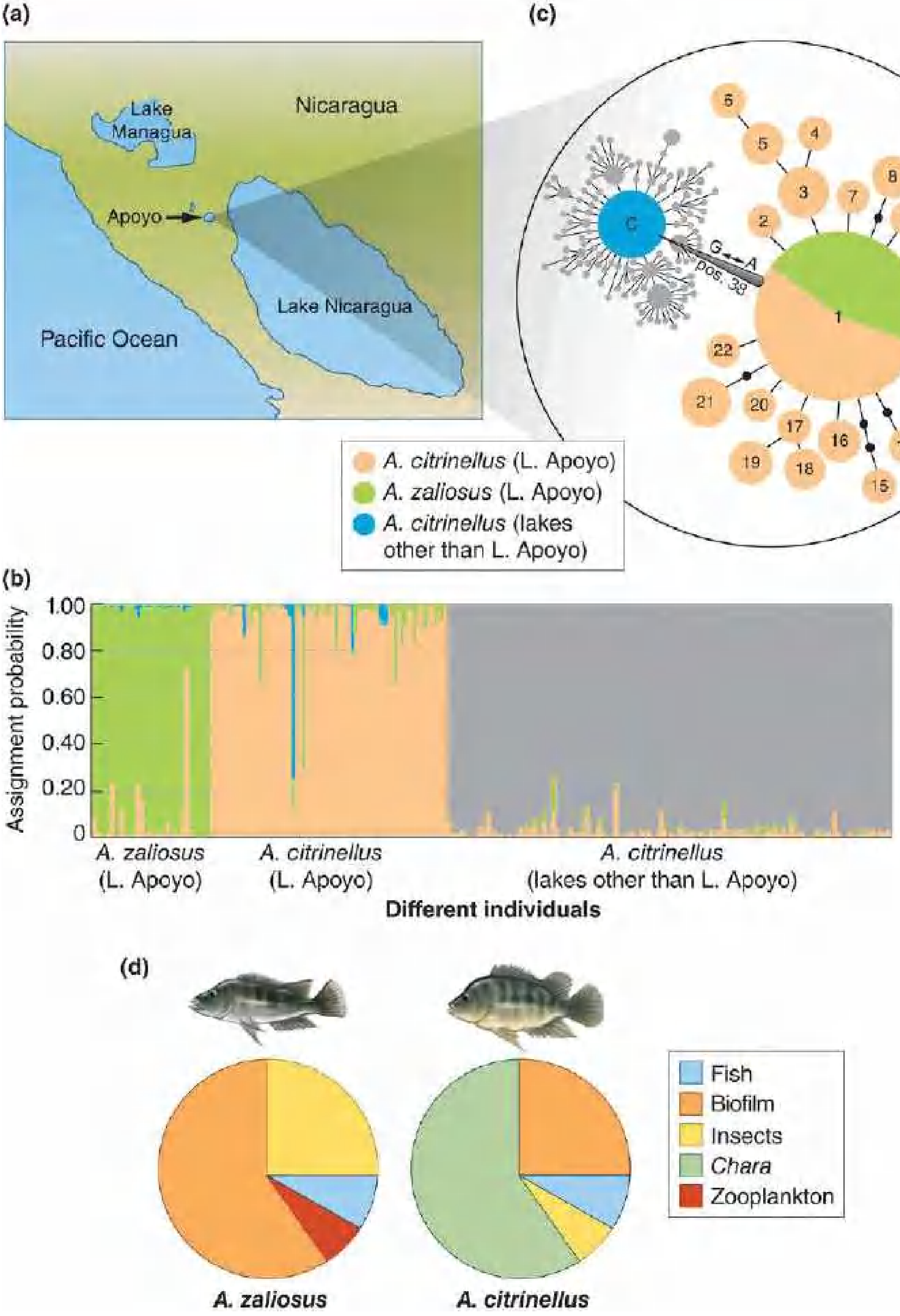

(a)
Lake Managua
Nicaragua
Apoyo
Lake Nicaragua
Pacific Ocean
(c)
C
G A
pos. 38
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
A. citrinellus (L. Apoyo)
A. zaliosus (L. Apoyo)
A. citrinellus (lakes other than L. Apoyo)
(b)
Assignment probability
1.00
0.80
0.60
0.40
0.20
0
A. zaliosus (L. Apoyo)
A. citrinellus (L. Apoyo)
A. citrinellus (lakes other than L. Apoyo)
Different individuals
(d)
Fish
Biofilm
Insects
Chara
Zooplankton
A. zaliosus
A. citrinellus

Figure 1.11 Sympatric speciation in the cichlid fish *Amphilophus citrinellus* and *A. zaliosus*. (a) The location of Lake Apoyo in Nicaragua. (b) The assignment of individuals to populations based on variation at 10 microsatellite DNA loci: *A. zaliosus* (green) and *A. citrinellus* (orange) from Lake Apoyo and *A. citrinellus* (blue) from other lakes. The clear separation of colours is indicative of partial or total reproductive isolation between them. (c) The relatedness network of 637 'haplotypes' of mitochondrial DNA sequences from individual fish using the same colour coding as in (b). (A haplotype is a set of markers on a single chromosome that tend to be inherited together from a single parent.) The size of a circle reflects the frequency with which a particular haplotype was found. The most common haplotype in Lake Apoyo ('1') is distinguished from the most common *A. citrinellus* haplotype from elsewhere ('C') by a single mutation (base substitution of guanine by adenine at position 38), but all Lake Apoyo haplotypes, of both species, share this mutation, indicating their origin from a single common ancestor. (d) Stomach content analyses of the two species in Lake Apoyo. (*Chara* is a multicellular alga.).

Source: After Barluenga *et al.* (2006).

where is sympatric speciation most likely?

Examples of species groups most likely to satisfy Coyne and Orr's (2004) criteria are organisms with a strong, genetically determined fidelity to a habitat in which mating will occur, such as insects that feed on more than one species of host plant, where each requires specialisation by the insects to overcome the plant's defenses, fish on coral reefs (and perhaps marine animals more generally; Bird *et al.*, 2012) and parasites (Santini *et al.*, 2012). And we have already seen how two lake fish conform to the scenario. Indeed, one of the most staggeringly rich examples of endemism has also been provided by cichlid fish: those of the East African Great Lakes, with more than 1500 endemic species in a relatively small, isolated geographic region. It remains to be discovered how important a role sympatric speciation plays in that case, and whether divergent selection to different niches is the main driving force.

sympatric speciation – divergence with gene flow or microallopatric speciation?

A final critical question is whether a case thought to have arisen by sympatric speciation is truly a result of species diverging while gene flow was occurring (sympatric) or merely 'microallopatric' speciation. A small-scale geographic barrier (analogous to ocean habitat between islands) may occur, for example, as an underwater ridge in a lake. Moreover, host-specific parasites and phytophagous insects might also have broadly overlapping geographic ranges and yet never encounter one another because of their distinct ecological niches. In other words, populations might overlap at a coarse grain if they occupy the same geographic region, but not co-occur at a finer grain if they occupy different habitats within that region. Thus, whether populations are described as sympatric is somewhat at the discretion of the observer (Fitzpatrick *et al.*, 2008).

a mechanism for sympatric speciation: AITs?

It is easy to see how geographically isolated populations have diverged, because they are also reproductively isolated, but not so straightforward to conceive how assortative mating can evolve sympatrically in populations that are not geographically isolated but experience divergent selective pressures. This may occur via 'automatic isolating traits' (AITs). An example would be where a particular locus or set of loci interacts with the environment to express different mating behaviours under different environmental conditions, regardless of genotype, such as the timing of flowering in plants. For example, the most recent common ancestor of two sympatric sister *Howea* palms on the tiny Lord Howe Island, 600 km off the coast of Australia, may have exhibited different flowering times when growing in different soil types so that a difference in physiology elicited by environmental differences, rather than a difference in genotype, could have enforced mating fidelity by soil type rather than genotype and increased the likelihood that divergence was possibly despite broad-scale sympatry (Figure 1.12). Papadopulos *et al.* (2011) describe other examples of sympatric speciation of plants in the genera *Metrosideros* and *Coprosma* on Lord Howe Island. Further possible cases where AITs may operate include fish with colour polymorphisms, genes responsible for insect hybrid male sterility, and cases involving chemical signalling (Bird *et al.*, 2012).

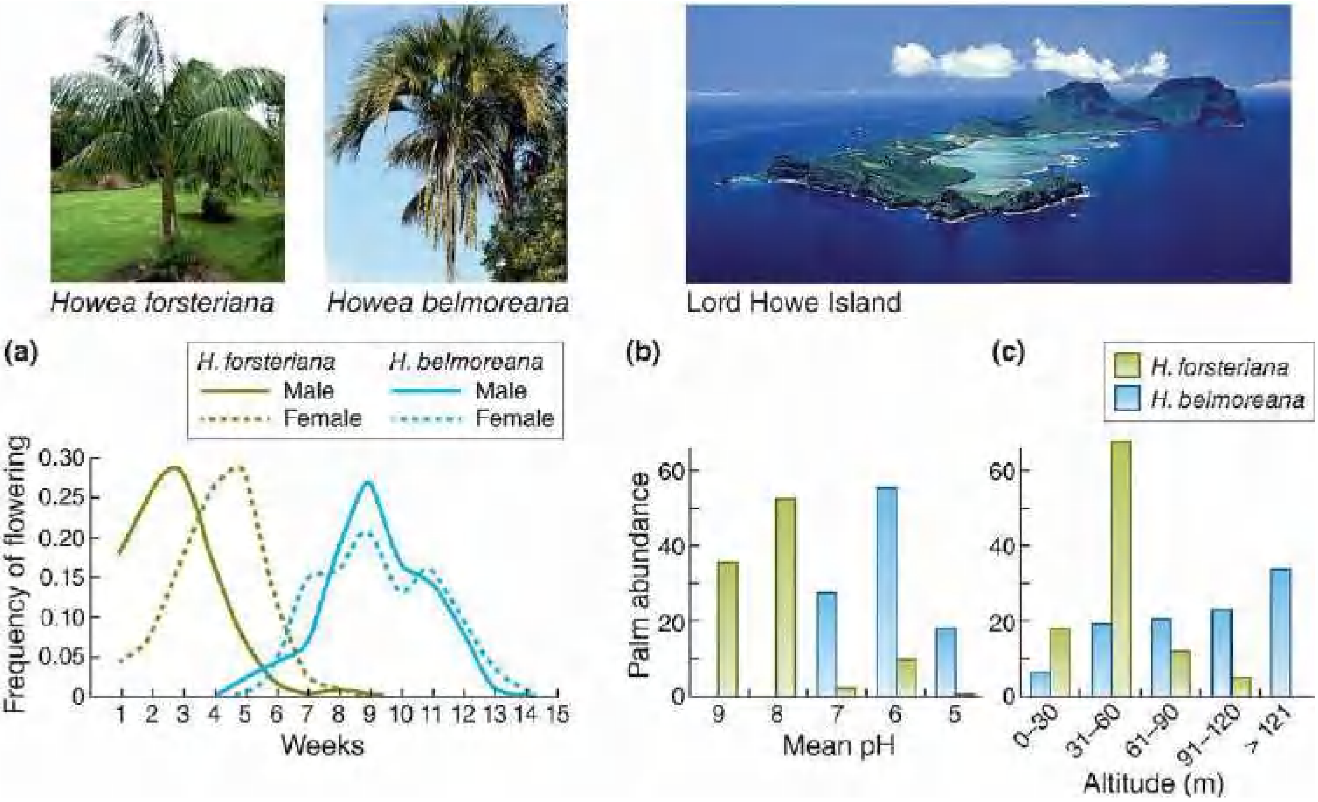


Figure 1.12 Sympatric speciation in *Howea* palms. Two species of *Howea* palm on the tiny and isolated Lord Howe Island off the coast of Australia. *Howea forsteriana* has straight leaves with drooping leaflets, while *H. belmoreana* has recurved leaves with ascending leaflets. (*H. forsteriana* is now one of the world's most widely traded house plants.) A comprehensive DNA-based phylogenetic tree indicates that these are sister species with their closest relative, *Laccospadix*, on the Australian mainland. Molecular dating methods show the two *Howea* species diverged 1–1.92 million years ago, long after Lord Howe Island was formed by volcanic activity 6.4–6.9 million years ago. *H. forsteriana* diverged from its sister species (an ancestor of *H. belmoreana*) by colonising widespread lowland calcarenite deposits. Extensive molecular evidence is consistent with Coyne and Orr's criteria for sympatric speciation (see earlier). (a) *H. forsteriana* (green lines) flowers early in the flowering season, with male flowering (solid line) peaking 2 weeks before female receptivity (dashed line); *H. belmoreana* male and female flowering is synchronous but later in the season. (b) *H. forsteriana* occurs in soils of higher pH and (c) lower altitude than *H. belmoreana*.

Source: After Savolainen *et al.* (2006).

While allopatric speciation is generally accepted to be much more common than sympatric speciation, sympatric lineage divergence due to selection has certainly come of age in the wake of the molecular biology revolution, which has allowed hypotheses that were once untestable to be critically evaluated. Evolutionary ecologists are not so focused now on whether or not sympatric speciation can happen, but rather how often and under what conditions.

APPLICATION 1.3 Conservation significance of hot spots of endemism

Conservationists have to make hard decisions in their quest to preserve biological diversity. Given limited resources, how can the most species be supported at minimum cost? One way is to focus attention on 'biodiversity hot spots' of species that are found nowhere else. Myers *et al.* (2000) took this approach when mapping the entire globe in terms of exceptional concentrations of endemic species coupled with exceptional loss of habitat (and therefore subject to a greater degree of threat to biodiversity than areas without such habitat loss). Hot-spot boundaries were set according to the characteristic biotas they contain: examples include island groups such as the Galápagos (Section 1.3.2) and Hawaii (Section 1.4.2), and 'ecological' islands such as the East African Great Lakes (Section 1.3.3) or clearly defined continental units such as the Cape Floristic Province in South Africa. The taxa included in the analysis consisted of vascular plants, mammals, birds, reptiles and amphibians. Figure 1.13 shows 25 identified hot spots that between them contain 133 149 plant species (that is, 44% of the world's plants) and 9645 vertebrate species (35% of the world's total). Or to put it in another way that emphasises their importance, we can say that this set of hot spots provide the sole remaining habitats of 44% of the world's plant species (and 35% of animals).

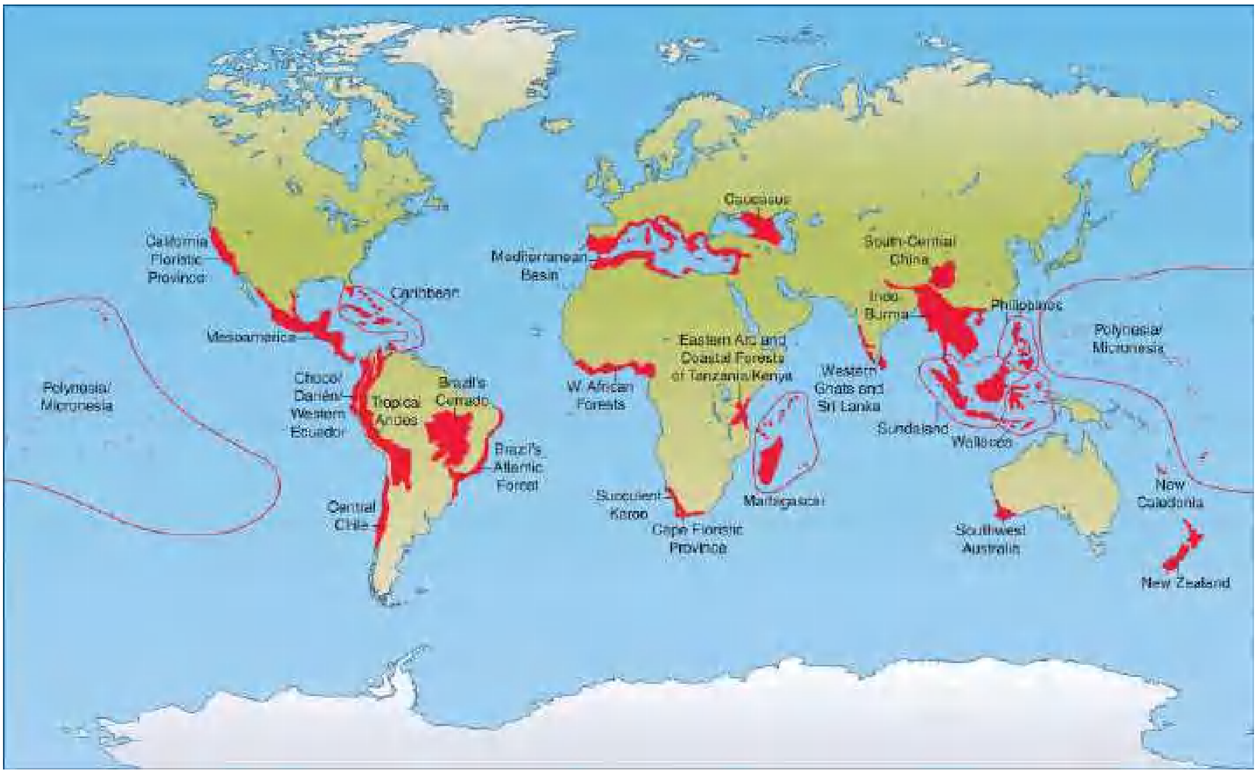


Figure 1.13 Biodiversity hot spots. Twenty-five biodiversity hot spots identified because of their exceptional concentrations of endemic species that are undergoing exceptional levels of human induced habitat loss.

Source: From Myers *et al.* (2000).

The five most prominent hot spots, the tropical Andes, Sundaland, Madagascar, Brazil's Atlantic Forest and the Caribbean, contain 20% of the world's vascular plants and 16% of vertebrate species but together they comprise only 0.4% of the world's surface. Moreover, they are subject to some of the heaviest levels of habitat loss: the Caribbean retains only 11.3% of its primary vegetation, Madagascar 9.9%, Sundaland 7.8% and Brazil's Atlantic Forest 7.5%. There was reasonable congruence between levels of endemism of plants and vertebrates in the hot spots, but note that no invertebrates were included in the analysis. In a geographically more restricted study in South Africa, Bazelet *et al.* (2016) showed that there was congruence between hot spots of the rather circumscribed diversity of katydids (bush crickets) and the biodiversity hot spots already recognised for much wider groupings, indicating that the conservation of biodiversity hot spots may often also protect non-target organisms.

Myers *et al.* (2000) called for a more than 10-fold increase in annual funding from governmental and international agencies to safeguard these hot spots.

1.4 The role of historical factors in the determination of species distributions

Our world has not been constructed by someone taking each species in turn, testing it against each environment, and moulding it so that every species finds its perfect place. It is a world in which species live where they do for reasons that are often, at least in part, accidents of history. We illustrate this first by considering continental drift, a process that operates over a timescale of tens of millions of years.

1.4.1 Movements of landmasses

Long ago, the curious distributions of species between continents, seemingly inexplicable in terms of dispersal over vast distances, led biologists, especially Wegener (1915), to suggest that the continents themselves must have moved. This was vigorously denied by geologists, until geomagnetic measurements required the same, apparently wildly improbable explanation. The discovery that the tectonic plates of the earth’s crust move and carry with them the migrating continents, reconciles geologist and biologist (Figure 1.14). Thus, whilst major evolutionary developments were occurring in the plant and animal kingdoms, populations were being split and separated, and land areas were moving across climatic zones.

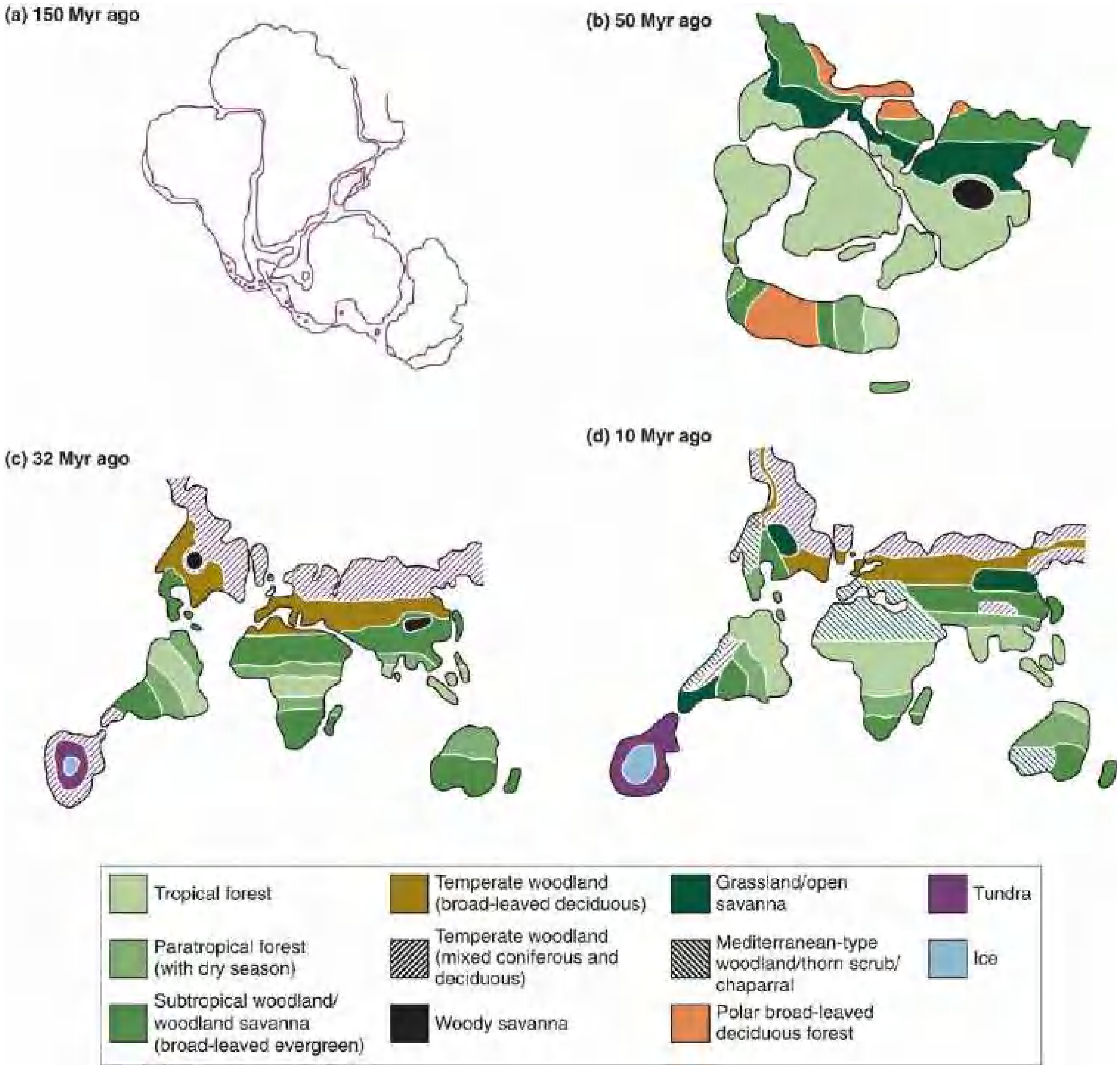


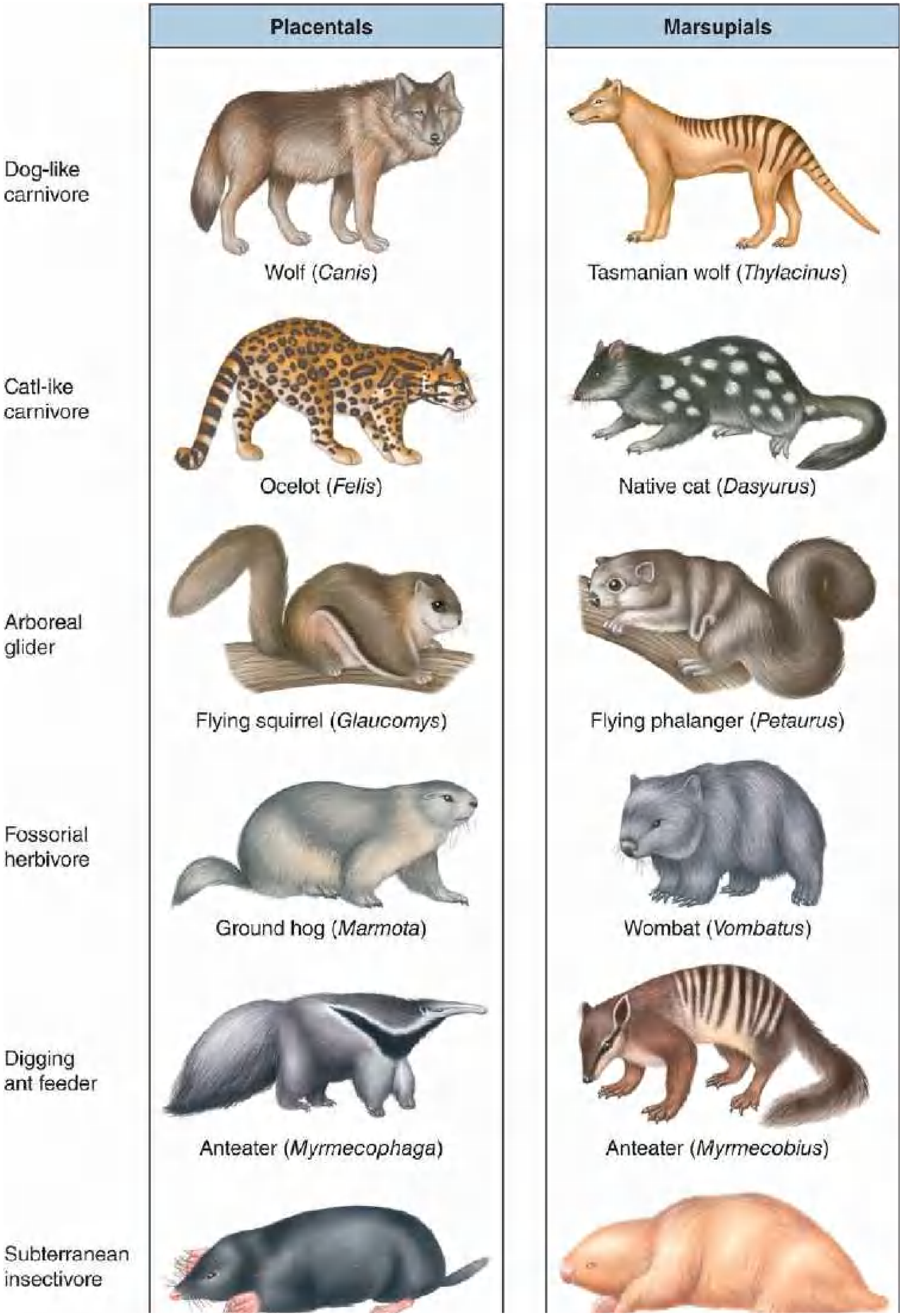
Figure 1.14 **Continental drift means that continents that are now separate were once joined to one another.** (a) The ancient supercontinent of Gondwanaland began to break up about 150 million years (Myr) ago. (b) About 50 Myr ago (early Middle Eocene) recognisable bands of distinctive vegetation had developed, and (c) by 32 Myr ago (early Oligocene) these had become more sharply defined. (d) By 10 Myr ago (early Miocene) much of the present geography of the continents had become established but with dramatically different climates and vegetation from today; the position of the Antarctic ice cap is highly schematic.

Source: After Norton & Sclater (1979), Janis (1993) and other sources.

placental and marsupial mammals

The drift of large landmasses over the face of the earth explains many patterns in the distribution of species that would otherwise be difficult to understand. A classic example is provided by the placental and marsupial mammals. Marsupials arrived on what would become the Australian continent about 90 million years ago (in the Cretaceous period), when the only other mammals present were the curious egg-laying monotremes (now represented only by the spiny anteaters

(*Tachyglossus aculeatus*) and the duckbill platypus (*Ornithorynchus anatinus*)). An evolutionary process of radiation then occurred that in many ways paralleled that of placental mammals on other continents (Figure 1.15). The subtlety of the parallels in both the form of the organisms and their lifestyle is so striking that it is hard to escape the view that the environments of placentals and marsupials provided similar opportunities to which the evolutionary processes of the two groups responded in similar ways. Because they started to diversify from a common ancestral line, and both inherited a common set of potentials and constraints, we refer to this as *parallel evolution* (as opposed to *convergent evolution*, where structures are analogous (similar in superficial form or function) but not homologous (i.e. not derived from an equivalent structure in a common ancestry), such as the wings of birds and bats). The important point here, though, is that the marsupials are found where they are not simply because they are the best fitted to those particular environments but also because of an accident of history – in this case, geological history.


Placentals
Marsupials
Dog-like carnivore
Wolf (*Canis*)
Tasmanian wolf (*Thylacinus*)
Catl-ike carnivore
Ocelot (*Felis*)
Native cat (*Dasyurus*)
Arboreal glider
Flying squirrel (*Glaucomys*)
Flying phalanger (*Petaurus*)
Fossorial herbivore
Ground hog (*Marmota*)
Wombat (*Vombatus*)
Digging ant feeder
Anteater (*Myrmecophaga*)
Anteater (*Myrmecobius*)
Subterranean insectivore

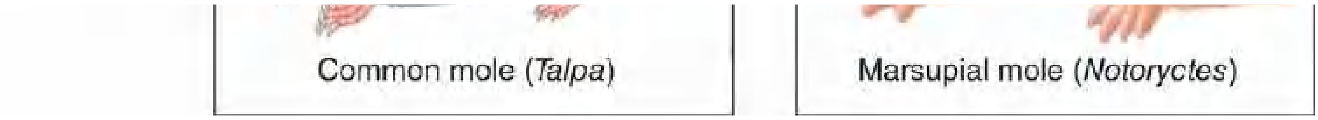


Figure 1.15 Parallel evolution of marsupial and placental mammals. The pairs of species are similar in both appearance and habit, and usually (but not always) in lifestyle.

1.4.2 Island history

Hawaii provides another remarkable example of a historical process that depends on the movement of a tectonic plate, but in this case in relation to volcanism and in a restricted geographic area. The Hawaiian chain of islands is volcanic in origin, having been formed gradually over the last 40 million years, as the centre of the Pacific tectonic plate moved steadily over a volcanic 'hot spot' in a south-easterly direction. Thus, Niihau and Kauai are the most ancient of the islands, and Hawaii itself the most recent.

Hawaiian *Drosophila*

The *Drosophila* 'fruit-flies' of Hawaii provide an especially spectacular example of species formation and endemism on islands. There are several thousand species of *Drosophila* worldwide (not all named yet) of which up to 1000 are found only in the Hawaiian Islands (Kang *et al*., 2016). Of particular interest are the 120 or so endemic species of 'picture-winged' *Drosophila*, very few of which occur on more than one island. The majority of the picture-winged species are specialised to oviposit and develop in the decaying bark of native trees in particular families. The lineages through which these species have evolved can be traced by analysing their DNA sequences, in this case using five nuclear genes with a total of 4260 nucleotides, to produce a comprehensive phylogeny of 93 of the species (Magnacca & Price, 2015). The evolutionary tree that emerges is shown in Figure 1.16, with each estimated species divergence date lined up alongside the island on which it occurred.

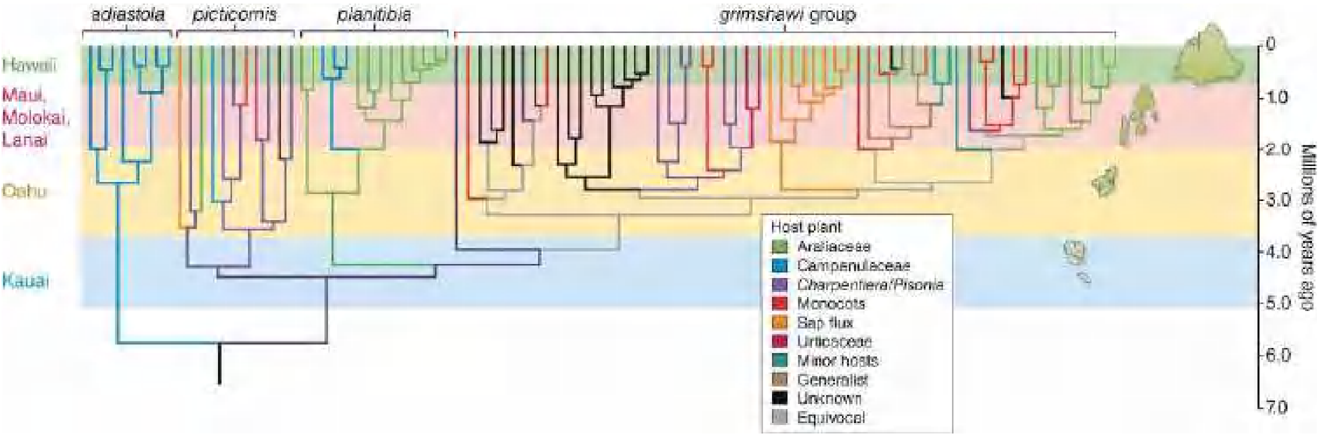


Figure 1.16 An evolutionary tree linking 93 species of picture-winged *Drosophila* on the Hawaiian Islands, traced by the analysis of DNA sequences on five nuclear genes, with species groups indicated. The shaded areas denote the time periods when the labelled island was the youngest available for colonisation. Species' divergence dates are lined up against the island on which they occurred. The islands of Maui, Molokai and Lanai are grouped together because they were recently linked by a land bridge. The earliest split (with separation of the *adiastola* group) occurred prior to the emergence of the mature Kauai, with subsequent splits into the other three species groups occurring on the ancient Kauai, and later dispersal to the younger islands as these emerged and matured. Speciation often involved specialisation on particular host plant families, indicated by different colours in the tree ('sap flux' is fermenting sap as it oozes from trees; 'minor hosts' are individual plant species, not in the families in the list, and used by only one or two species).

Source: From Magnacca & Price (2015).

The standard view for the biogeographic evolution of Hawaiian taxa is a 'progression-rule pattern', with the most basal species occurring on ancient Kauai and each lineage dispersing to younger islands as these emerged and matured to a deeply eroded topography with forests and well-developed soils. Historical elements in 'what lives where' are plainly apparent for many of the species groups in the figure. The earliest split within the picture-wing clade occurred prior to the emergence of the mature Kauai, with separation of the basal *adiastola* group (Figure 1.16). The figure also shows that the basal species of the *picticornis*, *planitibia* and *grimshawi* groups are found on the ancient island of Kauai, with these groups separating 3.8–4.7 million years ago (mya). A second stage of diversification can be seen when the *grimshawi* subgroups split at 2.1–3.1 mya on Oahu. The *planitibia* group provides a particularly clear example of progression-rule dispersal, with a split into two lineages on Kauai, followed by a split into three lineages on Oahu and subsequent dispersal to the younger islands. But such clear patterns are not always discernible, particularly in the *grimshawi* group. As new islands have been formed, rare dispersers have reached them and eventually evolved into new species, usually by becoming specialised on particular host plants. The arrival on Oahu around 3 mya of new plants upon which *grimshawi* species specialise (including *Charpentiera* and *Pisona* spp.) may have triggered a burst of speciation in the group. At least some of the picture-winged species appear to match the same environment as others on different islands. Of two closely related species, for example, *D. adiastola* is only found on Maui and *D. setosimentum* only on Hawaii, but the environments that they live in are apparently indistinguishable (Heed, 1968). What is most noteworthy, of course, is the power and importance of isolation (coupled with natural selection) in generating new species. Thus, this island biota illustrates two important, related points: (i) that there is a historical element in the match between organisms and environments; and (ii) that there is not just one perfect organism for each type of environment.

1.4.3 Climatic history

Climatic variations have occurred on shorter timescales than the movements of landmasses. Changes in climate during the Pleistocene ice ages, in particular, bear a lot of the responsibility for the present patterns of distribution of plants and animals. Techniques for analysing and dating biological remains (particularly buried pollen) increasingly allow us to detect just how much of the present distribution of organisms is a precise locally evolved match to present environments, and how much is a fingerprint of the hand of history. As climates have changed, species populations have advanced and retreated, been fragmented into isolated patches, and may then have rejoined. Much of what we see in the present distribution of species represents a phase in the recovery from past climatic change (Figure 1.17).

the Pleistocene glacial cycles ...

Techniques for the measurement of oxygen isotopes in ocean cores indicate that there may have been as many as 16 glacial cycles in the Pleistocene, each lasting for about 125 000 years (Figure 1.17a). Each cold (glacial) phase may have lasted for as long as 50 000–100 000 years, with brief intervals of only 10 000–20 000 years when the temperatures rose to, or above, those of today. From this perspective, present floras and faunas are unusual, having developed at the warm end of one of a series of unusual catastrophic warm periods.

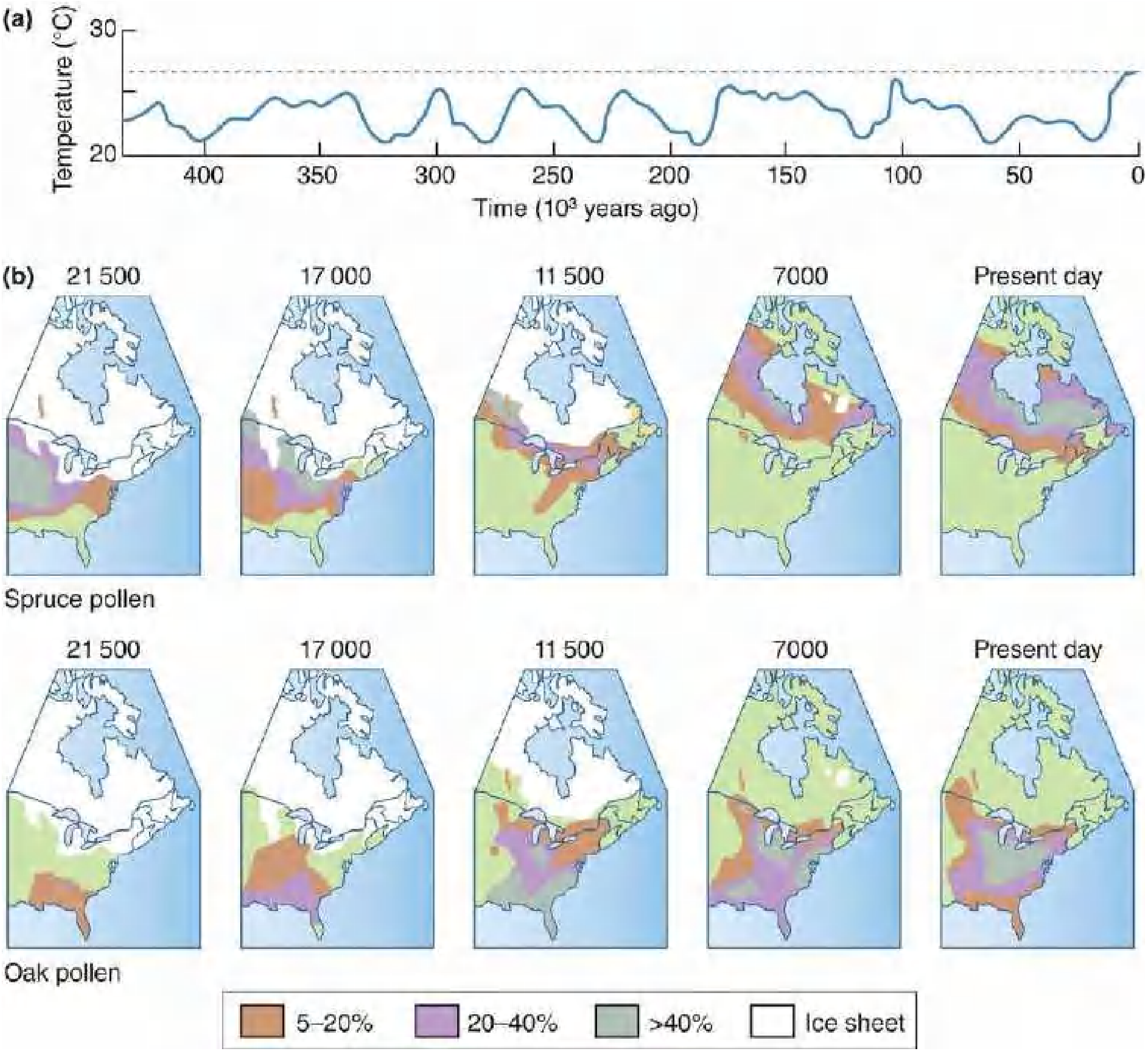


Figure 1.17 Contrasting changes in the distribution of spruce and oak species in relation to the waning of an ice age. (a) Estimates of temperature during glacial cycles over the past 400 000 years, obtained by comparing oxygen isotope ratios in fossils taken from ocean cores in the Caribbean. Periods as warm as the present have been rare events, and the climate during most of the past 400 000 years has been glacial. The dotted line represents the temperature 10 000 years ago at the beginning of the present period of warming (b) Ranges in eastern North America, as indicated by pollen percentages in sediments, of spruce species (above) and oak species (below) from 21 500 years ago to the present. Note how the ice sheet contracted during this period.

Source: (a) After Emiliani (1966) and Davis (1976). (b) After Davis & Shaw (2001).

During the 20 000 years since the peak of the last glaciation, global temperatures have risen by about 8°C. The analysis of buried pollen – particularly of woody species, which produce most of the pollen – can show how vegetation has changed (Figure 1.17b). As the ice retreated, different forest species advanced in different ways and at different speeds. For some, like the spruce of eastern North America, there was displacement to new latitudes; for others, like the oaks, the picture was more one of expansion.

We do not have such good records for the postglacial spread of animals associated with the changing forests, but it is certain that many species could not have spread faster than the trees on which they feed. Some of the animals may still be catching up with their plants, and tree species are still returning to areas they occupied before the last ice age. It is quite wrong to imagine that our present vegetation is in some sort of equilibrium with (adapted to) the present climate.

Even in regions that were never glaciated, pollen deposits record complex changes in distributions. In the mountains of the Sheep Range, Nevada, for example, different woody species of plant show different patterns of change in the ranges of elevations that they have occupied as climate has changed (Figure 1.18). The species composition of vegetation has continually been changing and is almost certainly still doing so.

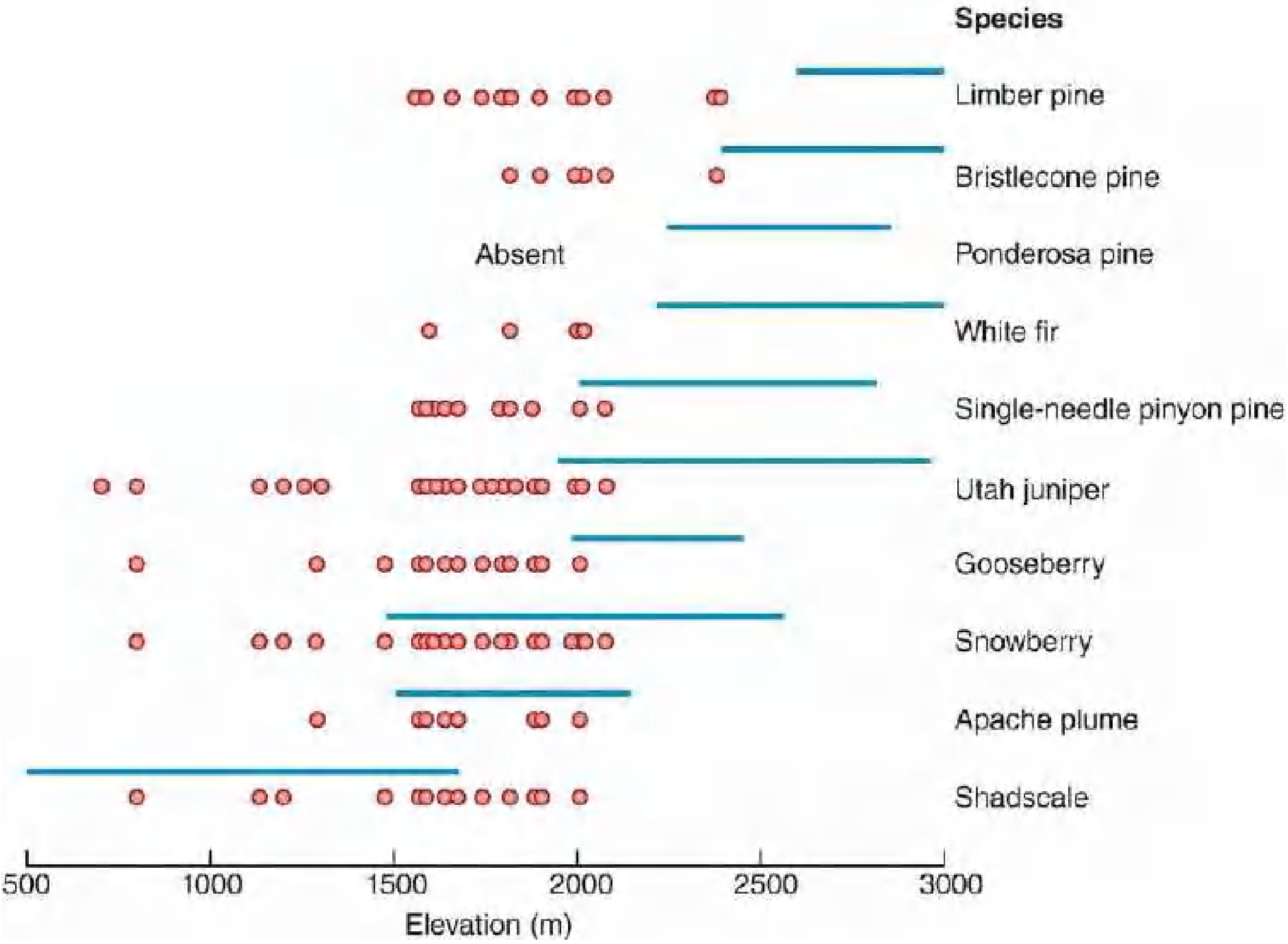


Figure 1.18 Contrasting changes between fossil and current distributions of 10 species of woody plant from the mountains of the Sheep Range, Nevada. The red dots represent fossil records, while the blue lines show current elevational ranges.

Source: After Davis & Shaw (2001).

The records of climatic change in the tropics are far less complete than those for temperate regions. It has been suggested, though, that during cooler, drier glacial periods, the tropical forests retreated to smaller patches, surrounded by a sea of savanna, within which speciation was intense, giving rise to present-day ‘hot spots’ of endemism. Evidence for this in, for example, the Amazonian rainforest now seems less certain than it once did, but there is support for the idea in other regions. In the Australian wet tropics of Queensland, north-eastern Australia, it has been possible to use present-day distributions of forest to predict distributions in the cool-dry climate of the last glacial maximum when forest contraction was greatest (about 18 000 years ago), the cool-wet period around 7000 years ago when a massive expansion was likely, and the warm-wet period around 4000 years ago when there was likely to have been another contraction (Figure 1.19a) (Graham *et al*., 2006). Putting the distributions together, then, allows each subregion of the forest to be assigned a 'stability' (Figure 1.19b) – the most stable being the one in which forest has been most constantly present – and these stabilities can in turn be compared with the species richness today, in each subregion, of mammals, birds, reptiles and frogs. Richness tends to be greatest where stability has been highest (Figure 1.19c), that is, where the forest refuges were in the past. On this interpretation, the present distributions of species may again be seen as largely accidents of history (where the refuges were) rather than precise matches between species and their differing environments.

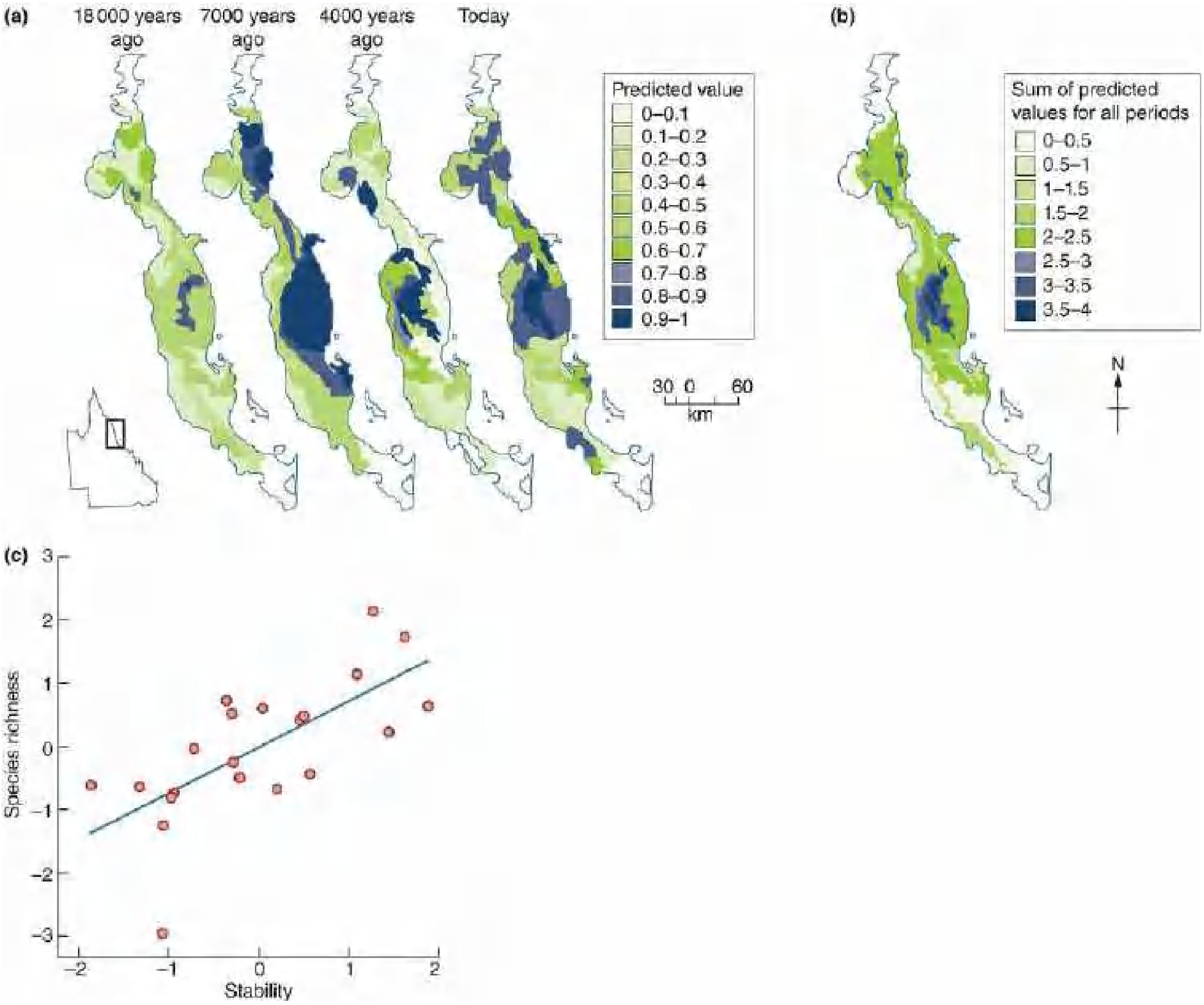


Figure 1.19 Forest species richness is positively related to forest 'stability' in north-east Australia. (a) A predictive study of the distribution of Australian 'wet forest' in Queensland (see inset), based on climatic conditions at 2000 random points in the current forested region. From left to right are the predicted overall distribution of forest across the region in cool-dry (18 000 years ago), cool-wet (7000 years ago), warm-wet (4000 years ago) and current climatic conditions. The values between 0 and 1 indicate the probability, according to the model, of forest being found at a given point. (b) Forest 'stability', which is calculated simply as the sum of the values in the four figures in (a). (c) Current species richness in 21 forest subregions (of mammals, birds, reptiles and frogs) increases with estimated 'stability'. The axes are scaled around zero because both stability and species richness were standardised to take account of the fact that both increase with subregion area.

Source: After Graham *et al*. (2006).

APPLICATION 1.4 Global warming and species distributions and extinctions

Evidence of changes in vegetation that followed the last retreat of the ice provides clues about the likely consequences of global warming associated with the continuing increases of carbon dioxide and other greenhouse gases in the atmosphere. Warming of 0.7°C in average global mean surface temperature was recorded between 1970 and 2010. Future climate will depend on warming still to occur but as a result of past anthropogenic emissions, future emissions, natural climate variability and whether or not major volcanic eruptions occur. Models that take into account various scenarios indicate that, relative to 1850–1900, global surface temperature increase by the end of the 21st century is likely to exceed 2°C (IPCC, 2014). But note that the scale of current temperature change is dramatically different from that which has occurred since the last ice age. Postglacial warming of 8°C over 20 000 years, or 0.04°C per century, must be compared with the current rate of global warming of about 1.75°C per century. It is disturbing to note that changes in the vegetation failed to keep pace even with a rise of 0.04°C per century. Projections for the 21st century require range shifts for trees at rates of 300–500 km per century compared with typical rates in the past of 20–40 km per century (and exceptional rates of 100–150 km). It is striking that the only precisely dated extinction of a tree species in the Quaternary period, that of *Picea critchfeldii*, occurred around 15 000 years ago at a time of especially rapid postglacial warming (Jackson & Weng, 1999). Clearly, even more rapid change in the future could result in extinctions of many additional species (Davis & Shaw, 2001).

The Pleistocene ice ages undoubtedly eliminated biota from many mid- to high-latitude areas of the planet. However, in the case of alpine species of the Pyrenees, Himalayas, Andes and Southern Alps, evidence is accumulating that glaciation may sometimes have promoted allopatric speciation by severing continuously distributed populations along the length of mountain ranges (Wallis *et al*., 2016). In the Southern Alps of New Zealand, for example, comparative phylogeographic analysis, based on mitochondrial and nuclear DNA, has revealed a phylogenetic split 2 mya (the date of the first major glacial epoch), found in both insect and bird biotas, that bisects each into northern and southern assemblages (Figure 1.20).

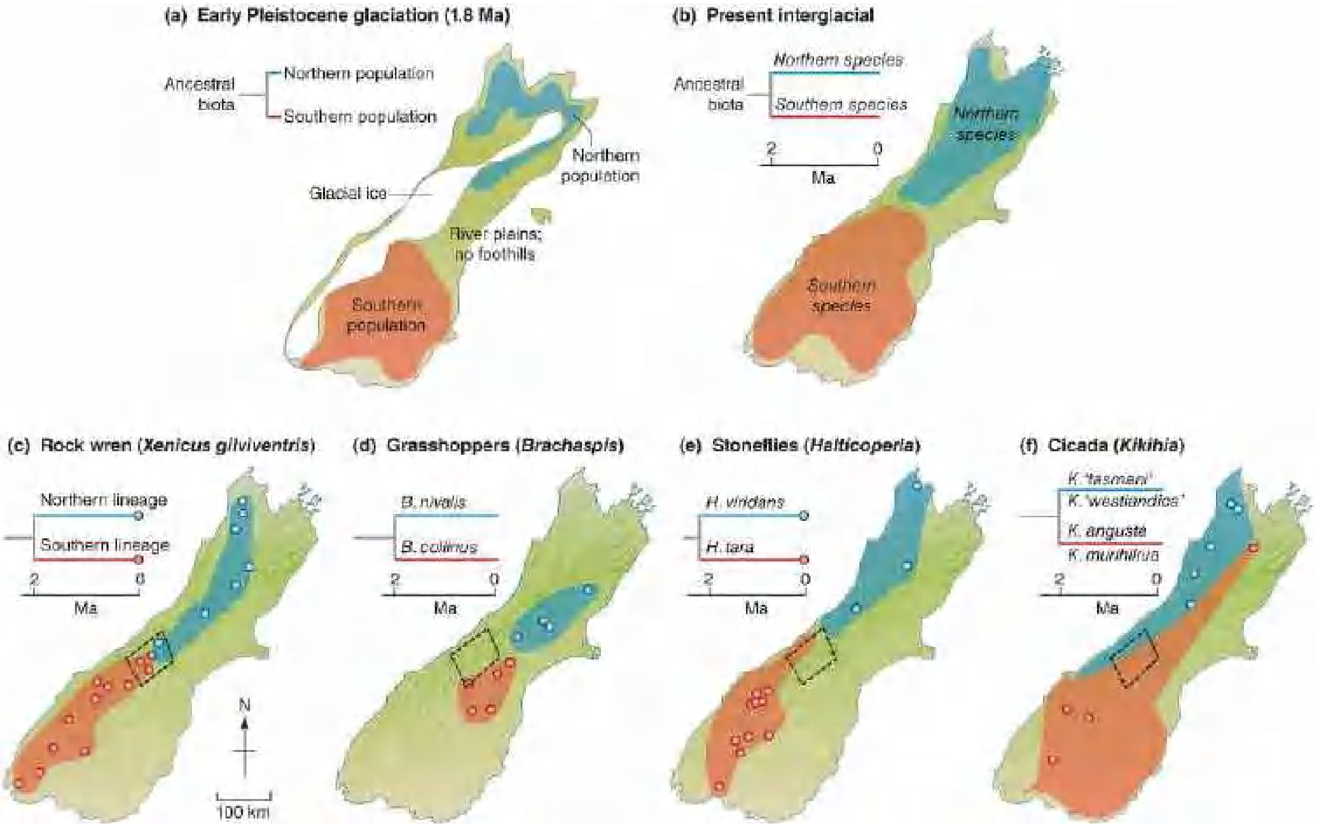


Figure 1.20 A phylogenetic split bisecting insect and bird biotas into northern and southern assemblages in New Zealand. (a, b) Glaciation of the alpine region of the South Island of New Zealand which is hypothesised to be responsible for fracturing the ancestral biota into north and south lineages that have subsequently diverged. Green shading represents lowlands. (c–f) North–south phylogeographic breaks for alpine birds and insects with dendrograms showing phylogenetic relationships and approximate divergence times. Sampling sites are shown as circles, the shading shows approximate taxon ranges, and the dotted square represents a particularly highly glaciated narrow alpine neck.

Source: After Wallis *et al.* (2016).

APPLICATION 1.5 Human history and species invasions

In this section on the role of history in the determination of species distributions, it would be misleading only to consider 'natural' historical factors. Human history has had equally profound effects on the distribution of species around the globe. The world has shrunk as travel has boomed and, just like us, animals, plants and microorganisms have become globetrotters, introduced to new locations on purpose or as accidental tourists.

Only about 10% of invaders become established, and of these, about 10% spread and have significant, sometimes disastrous, consequences (Townsend, 2008). The introduction in the 1950s of Nile perch (*Lates niloticus*) to establish a commercial fishery in Lake Victoria in East Africa, for example, is held responsible for the extinction or near-extinction of several hundred native fish species. An equally striking plant example is provided by the neotropical *Parthenium* weed (*Parthenium hysterophorus*) that has invaded large areas of a number of continents (Figure 1.21), generally by accident, and is responsible for the degradation of grasslands and of many cropping and livestock systems. Its success as an invader can be attributed to tolerance of a wide range of physicochemical conditions, escape from natural predators in non-native regions, high competitiveness against other plants (including the release of allelopathic chemicals into the soil) and high genetic diversity, among other factors (Bajwa *et al*., 2016). Finally, prominent among imported human disease organisms are HIV, influenza and zika viruses.

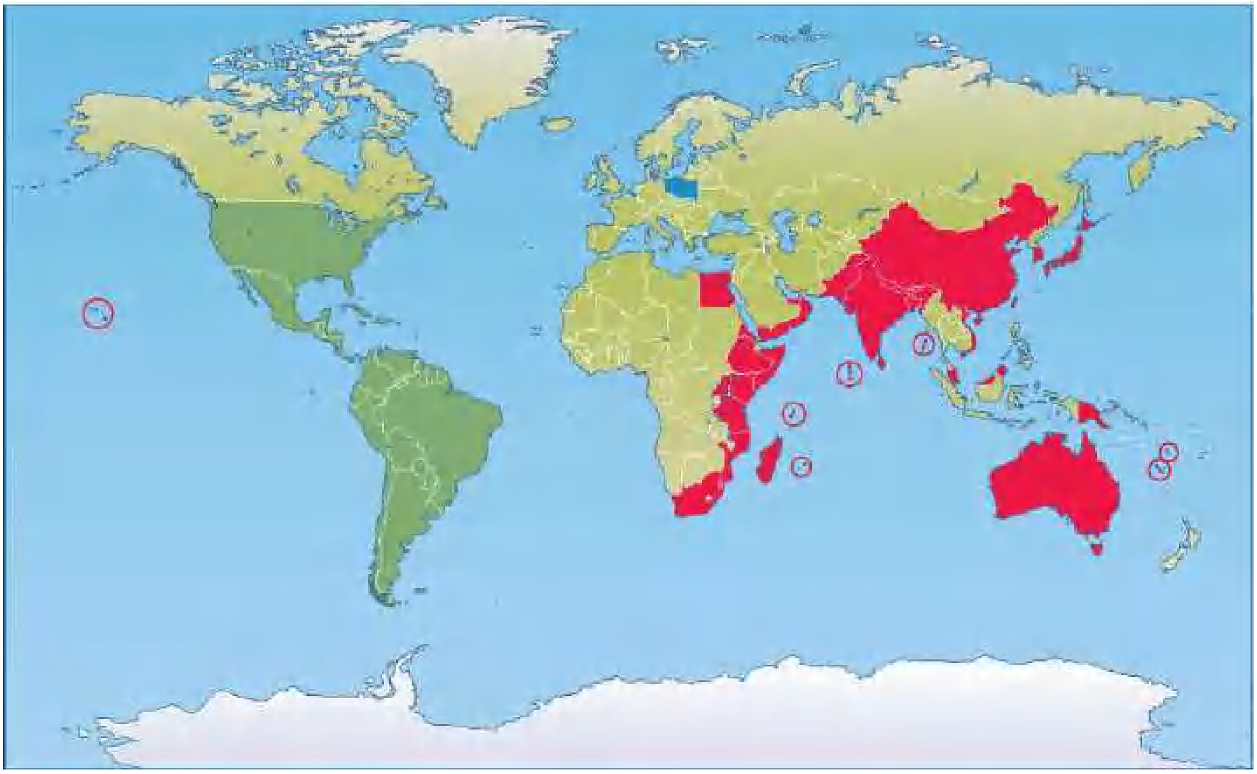


Figure 1.21 The invasion of *Parthenium* weed. This weed is invasive in the countries shaded or circled in red. Blue-shaded countries represent populations of the weed that may be establishing in Europe. Countries shaded green are considered to be within its invasive range.

Source: From Bajwa *et al*. (2016).

homogenisation of the biota

A profound consequence of global transport and past colonialism has been homogenisation of the biota. The same set of human camp followers now occur in widely separated regions – house sparrows, cockroaches, rats and mice, salmonid fish and game animals, domestic animals and crop plants (with their associated pests and diseases) (Townsend, 2008). Because native species often fare badly in the face of this onslaught, many parts of North America and the southern hemisphere now reflect a European legacy more closely than their native heritage. A graphic example of biotic homogenisation is provided at both ends of the trade link between the Great Lakes of North America and the Baltic Sea. Often spread in the ballast water of ships moving along this route, a third of the 170 species of invasive fish, molluscs and crustaceans in the Great Lakes are from the Baltic, and a third of the 100 invaders in the Baltic Sea come from the Great Lakes. Furthermore, enhanced shipping promoted by recent expansions of the Suez and Panama Canals can be expected to escalate marine invasions at regional and continental scales, while increases in internet-based trade in pets and other exotic organisms is providing novel pathways that will be difficult to regulate (Ricciardi *et al.*, 2017). We take this case up again in Section 6.6.2.

economic consequences of invasions

Invaders can have far-reaching economic as well as ecological consequences. Table 1.1 categorises the tens of thousands of exotic invaders in the USA into taxonomic groupings and lists these in order of their estimated annual costs (in terms of economic damage done and costs of control). Overall, pests of crop plants, including weeds, insects and pathogens are the most expensive, with rats some way behind but responsible for the destruction of stored grains, starting fires by gnawing through electrical wires, polluting foodstuffs, spreading diseases and preying on native animals. Imported human disease organisms are also very expensive to treat and result in 40 000 deaths per year. Ecological knowledge is needed to predict future invasions that are likely to have damaging consequences so that the invaders can be confronted, particularly via biosecurity at national borders. Despite their obvious deleterious effects, the importance of biological invasions seems not to be well understood by the public or fully accepted by decision makers (Courchamp *et al.*, 2017). Ecologists have much to contribute in this arena.

Table 1.1 Estimated annual costs (billions of US dollars) associated with damage caused and management costs of invaders in the USA. Taxonomic groupings are ordered in terms of the total costs associated with them. *Source*: After Pimentel *et al.* (2000).

Type of organism	Number of invaders	Major culprits	Loss and damage	Control costs	Total costs
Microbes (pathogens)	>20 000	Crop pathogens	32.1	9.1	41.2
Mammals	20	Rats and cats	37.2	Na	37.2
Plants	5 000	Crop weeds	24.4	9.7	34.1
Arthropods	4 500	Crop pests	17.6	2.4	20.0
Birds	97	Pigeons	1.9	NA	1.9
Molluscs	88	Asian clams, zebra mussels	1.2	0.1	1.3
Fishes	138	Grass carp, etc.	1.0	NA	1.0
Reptiles, amphibians	53	Brown tree snake	0.001	0.005	0.006

NA, data not available

1.5 The match between communities and their environments

1.5.1 Terrestrial biomes of the earth

Before we examine the differences and similarities between communities, we need to consider the larger groupings, ‘biomes’, in which biogeographers recognise marked differences in the flora and fauna of different parts of the world. The number of terrestrial biomes that are distinguished is a matter of taste. They certainly grade into one another, and sharp boundaries are a convenience for cartographers rather than a reality of nature. We describe eight terrestrial biomes and illustrate their global distribution in Figure 1.22, and show how they may be related to annual temperature and precipitation (Figure 1.23). Apart from anything else, understanding the terminology that describes and distinguishes these biomes is necessary when we come to consider key questions later in the book. Why are there more species in some communities than in others? Are some communities more stable in their composition than others, and if so why? Do more productive environments support more diverse communities? Or do more diverse communities make more productive use of the resources available to them?

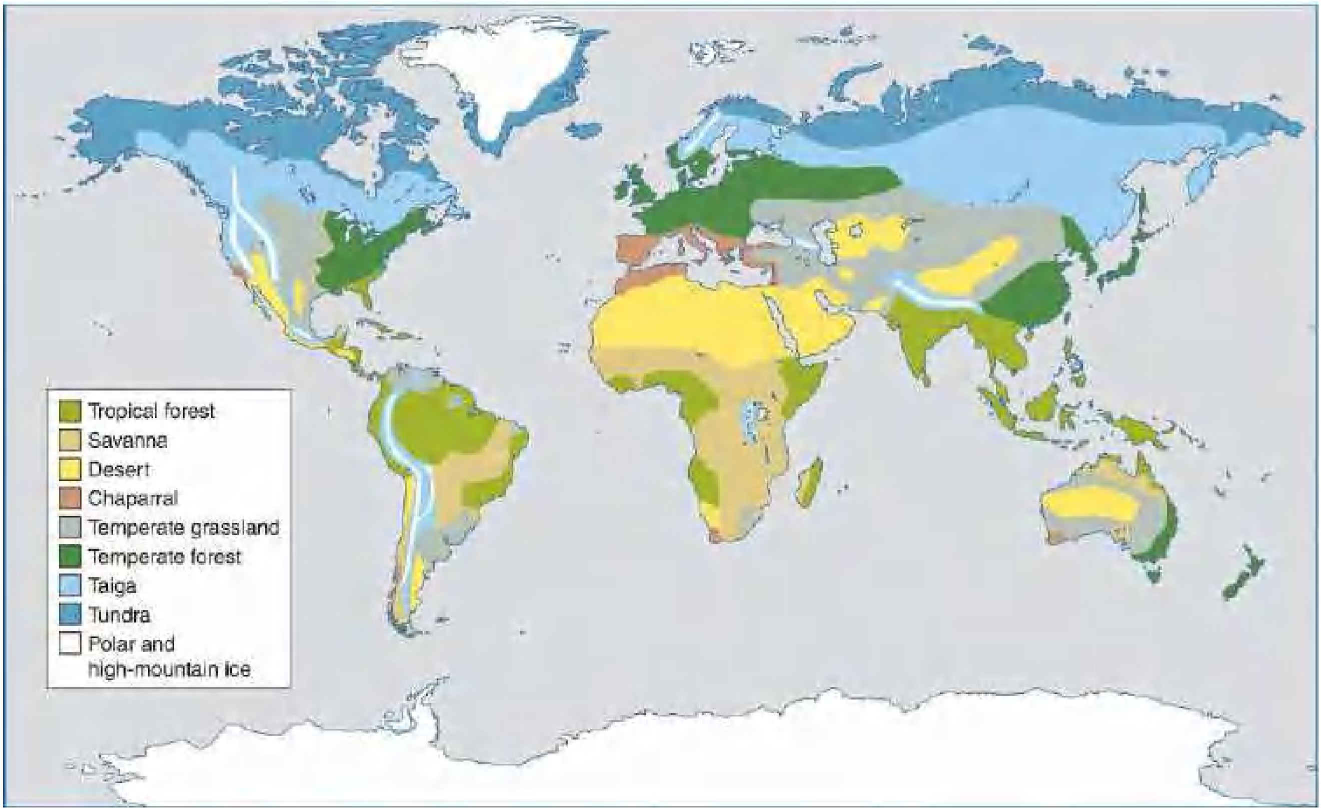
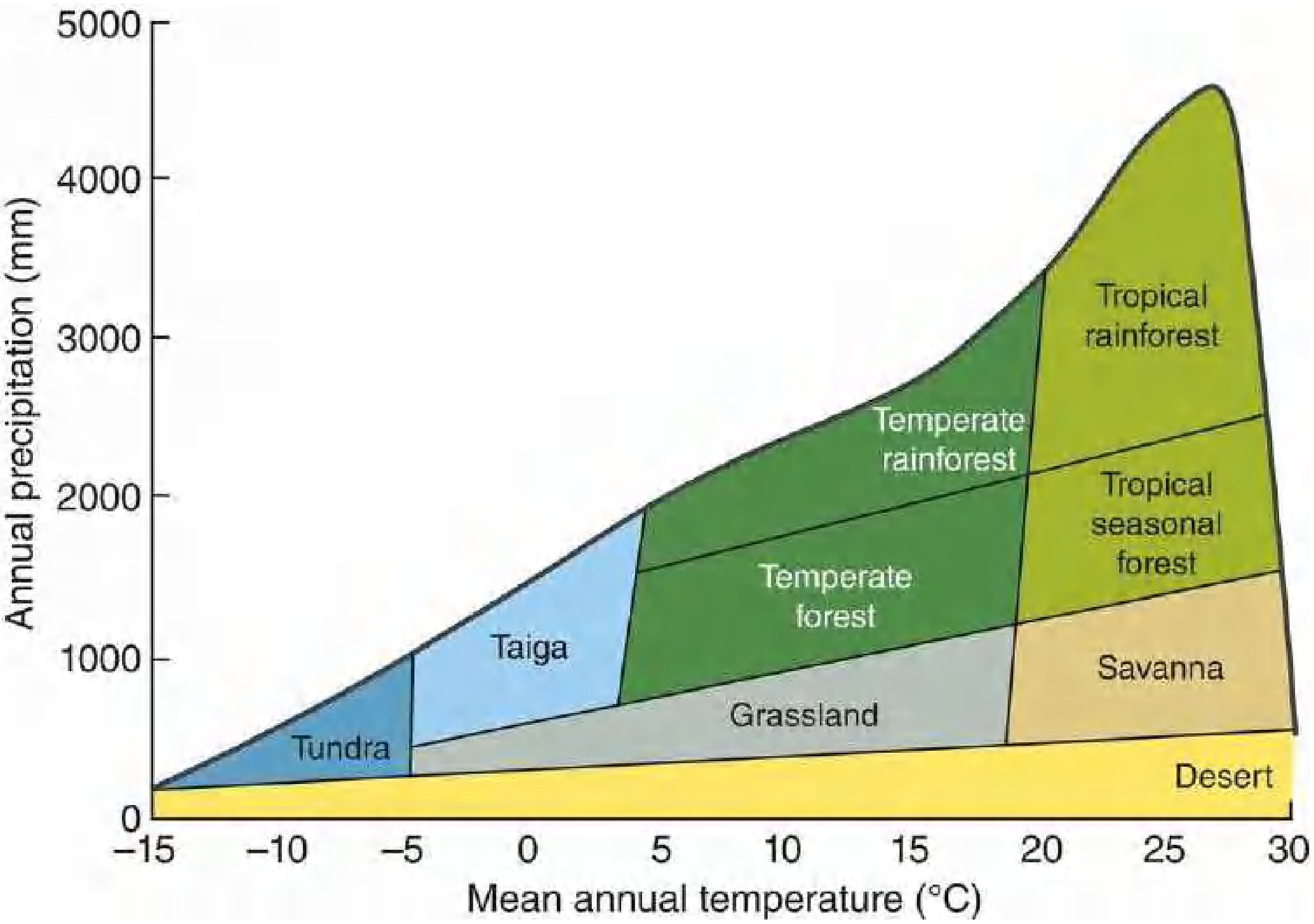


Figure 1.22 **World distribution of the major biomes of vegetation.**

Source: From http://www.zo.utexas.edu/faculty/sjasper/images/50.24.gif.



[Figure 1.23](#) **Biomes in relation to rainfall and temperature.** The variety of environmental conditions experienced in terrestrial biomes can be described in terms of their annual rainfall and mean annual temperatures.

Source: After Woodward & Lomas (2004).

tundra

Tundra occurs around the Arctic Circle, beyond the tree line. Small areas also occur on sub-Antarctic islands in the southern hemisphere. 'Alpine' tundra is found under similar conditions but at high altitude. The environment is characterised by the presence of permafrost – water permanently frozen in the soil – while liquid water is present for only short periods of the year. The typical flora includes lichens, mosses, grasses, sedges and dwarf trees. Insects are extremely seasonal in their activity, and the native bird and mammal fauna is enriched by species that migrate from warmer latitudes in the summer. In the colder areas, grasses and sedges disappear, leaving nothing rooted in the permafrost. Ultimately, vegetation that consists only of lichens and mosses gives way, in its turn, to the polar desert. The number of species of higher plants (i.e. excluding mosses and lichens) decreases from the Low Arctic (around 600 species in North America) to the High Arctic (north of 83°, e.g. around 100 species in Greenland and Ellesmere Island). In contrast, the flora of Antarctica contains only two native species of vascular plant and some lichens and mosses that support a few small invertebrates. The biological productivity and diversity of Antarctica are concentrated at the coast and depend almost entirely on resources harvested from the sea.

taiga

Taiga or northern coniferous forest occupies a broad belt across North America and Eurasia. Liquid water is unavailable for much of the winter, and plants and many of the animals have a conspicuous winter dormancy in which metabolism is very slow. Generally, the tree flora is very limited. In areas with less severe winters, the forests may be dominated by pines (*Pinus* species, which are all evergreens) and deciduous trees such as larch (*Larix*), birch (*Betula*) or aspens (*Populus*), often as mixtures of species. Farther north, these species give way to single-species forests of spruce (*Picea*) covering immense areas. The overriding environmental constraint in northern spruce forests is the presence of permafrost, creating drought except when the sun warms the surface. The root system of spruce can develop in the superficial soil layer, from which the trees derive all their water during the short growing season.

temperate forests

Temperate forests range from the mixed conifer and broad-leaved forests of much of North America and northern Central Europe (where there may be 6 months of freezing temperatures), to the moist, dripping forests of broad-leaved evergreen trees found at the biome's low-latitude limits in, for example, Florida and New Zealand. In most temperate forests, however, there are periods of the year when liquid water is in short supply, because potential evaporation exceeds the sum of precipitation and water available from the soil. Deciduous trees, which dominate in most temperate forests, lose their leaves in the autumn and become dormant. On the forest floor, diverse floras of perennial herbs often occur, particularly those that grow quickly in the spring before the new tree foliage has developed. Temperate forests also provide food resources for animals that are usually very seasonal in their occurrence. Many of the birds of temperate forests are migrants that return in spring but spend the remainder of the year in warmer biomes.

grassland

Grassland occupies the drier parts of temperate and tropical regions. Temperate grassland has many local names: the steppes of Asia, the prairies of North America, the pampas of South America and the veldt of South Africa. Tropical grassland or savanna is the name applied to tropical vegetation ranging from pure grassland to some trees with much grass. Almost all of these temperate and tropical grasslands experience seasonal drought, but the role of climate in determining their vegetation is almost completely overridden by the effects of grazing animals that limit the species present to those that can recover from frequent defoliation. In the savanna, fire is also a common hazard in the dry season and, like grazing animals, it tips the balance in the vegetation against trees and towards grassland. Nonetheless, there is typically a seasonal glut of food, alternating with shortage, and as a consequence the larger grazing animals suffer extreme famine (and mortality) in drier years. A seasonal abundance of seeds and insects supports large populations of migrating birds, but only a few species can find sufficiently reliable resources to be resident year-round.

Many of these natural grasslands have been cultivated and replaced by arable annual 'grasslands' of wheat, oats, barley, rye and corn. Such annual grasses of temperate regions, together with rice in the tropics, provide the staple food of human populations worldwide. At the drier margins of the biome, many of the grasslands are 'managed' for meat or milk production, sometimes requiring a nomadic human lifestyle. The natural populations of grazing animals have been driven back in favour of cattle, sheep and goats. Of all the biomes, this is the one most coveted, used and transformed by humans.

chaparral

Chaparral or *maquis* occurs in Mediterranean-type climates (mild, wet winters and summer drought) in Europe, California and north-west Mexico, and in a few small areas in Australia, Chile and South Africa. Chaparral develops in regions with less rainfall than temperate grasslands and is dominated mainly by a drought-resistant, hard-leaved scrub of low-growing woody plants. Annual plants are also common in chaparral regions during the winter and early spring, when rainfall is more abundant. Chaparral is subject to periodic fires; many plants produce seeds that will only germinate after fire while others can quickly resprout because of food reserves in their fire-resistant roots.

desert

Deserts are found in areas that experience extreme water shortage: rainfall is usually less than about 25 cm year^{-1}, is usually very unpredictable and is considerably less than potential evaporation. The desert biome spans a very wide range of temperatures, from hot deserts, such as the Sahara, to very cold deserts, such as the Gobi in Mongolia. In their most extreme form, the hot deserts are too arid to bear any vegetation; they are as bare as the cold deserts of Antarctica. Where there is sufficient rainfall to allow plants to grow in arid deserts, its timing is always unpredictable. Desert vegetation falls into two sharply contrasted patterns of behaviour. Many species have an opportunistic lifestyle, stimulated into germination by the unpredictable rains. They grow fast and complete their life history by starting to set new seed after a few weeks. These are the species that can occasionally make a desert bloom. A different pattern of behaviour is to be long-lived with sluggish physiological processes. Cacti and other succulents, and small shrubby species with small, thick and often hairy leaves, can close their stomata (pores through which gas exchange takes place) and tolerate long periods of physiological inactivity. The relative poverty of animal life in arid deserts reflects the low productivity of the vegetation and the indigestibility of much of it.

tropical rainforest

Tropical rainforest is the most productive of the earth's biomes – a result of the coincidence of high solar radiation received throughout the year and regular and reliable rainfall. The productivity is achieved, overwhelmingly, high in the dense forest canopy of evergreen foliage. It is dark at ground level except where fallen trees create gaps. Often, many tree seedlings and saplings remain in a suppressed state from year to year and only leap into action if a gap forms in the canopy above them. Apart from the trees, the vegetation is largely composed of plant forms that reach up into the canopy vicariously; they either climb and then scramble in the tree canopy (vines and lianas, including many species of fig) or grow as epiphytes, rooted on the damp upper branches. Most species of both animals and plants in tropical rainforest are active throughout the year, though the plants may flower and ripen fruit in sequence. Dramatically high species richness is the norm for tropical rainforest, and communities rarely if ever become dominated by one or a few species. The diversity of rainforest trees provides for a corresponding diversity of resources for herbivores, and so on up the food chain.

aquatic biomes?

All of these biomes are terrestrial. Aquatic ecologists could also come up with a set of biomes, although the tradition has largely been a terrestrial one. We might distinguish springs, rivers, ponds, lakes, estuaries, coastal zones, coral reefs and deep oceans, among other distinctive kinds of aquatic community. For present purposes, we recognise just two aquatic biomes, *marine* and *freshwater*. The oceans cover about 71% of the earth's surface and reach depths of more than 10 000 m. They extend from regions where precipitation exceeds evaporation to regions where the opposite is true. There are massive movements within this body of water that prevent major differences in salt concentrations developing (the average concentration is about 3%). Two main factors influence the biological activity of the oceans. Photosynthetically active radiation is absorbed in its passage through water, so photosynthesis is confined to the surface region. Mineral nutrients, especially nitrogen and phosphorus, are commonly so dilute that they limit the biomass that can develop. Shallow waters (e.g. coastal regions and estuaries) tend to have high biological activity because they receive mineral input from the land and less incident radiation is lost than in passage through deep waters. Intense biological activity also occurs where nutrient-rich waters from the ocean depths come to the surface; this accounts for the concentration of many of the world's fisheries in Arctic and Antarctic waters.

Freshwater biomes occur mainly on the route from land drainage to the sea. The chemical composition of the water varies enormously, depending on its source, its rate of flow and the inputs of organic matter from vegetation that is rooted in or around the aquatic environment. In water catchments where the rate of evaporation is high, salts leached from the land may accumulate and the concentrations may far exceed those present in the oceans; brine lakes or even salt pans may be formed in which little life is possible. Even in aquatic situations liquid water may be unavailable, as is the case in the polar regions.

Differentiating between biomes allows only a very crude recognition of the sorts of differences and similarities that occur between communities of organisms. Within biomes there are both small- and large-scale patterns of variation in the structure of communities and in the organisms that inhabit them. Moreover, as we see next, what characterises a biome is not necessarily the particular species that live there.

1.5.2 The 'life form spectra' of communities

We pointed out earlier the crucial importance of geographic isolation in allowing populations to diverge under selection. The geographic distributions of species, genera, families and even higher taxonomic categories of plants and animals often reflect this geographic divergence. All species of lemurs, for example, are found on the island of Madagascar and nowhere else. Similarly, 230 species in the genus *Eucalyptus* (gum tree) occur naturally in Australia (and two or three in Indonesia and Malaysia). The lemurs and the gum trees occur where they do because they evolved there – not because these are the only places where they could survive and prosper. Indeed, many *Eucalyptus* species grow with great success and spread rapidly when they have been introduced to, for example, California, Spain and Kenya. A map of the natural world distribution of lemurs tells us quite a lot about the evolutionary history of this group. But as far as its relationship with a biome is concerned, the most we can say is that lemurs happen to be one of the constituents of the tropical rainforest biome in Madagascar.

Similarly, particular biomes in Australia include certain marsupial mammals, while the *same* biomes in other parts of the world are home to their placental counterparts. A map of biomes, then, is not usually a map of the distribution of species. Instead, we recognise different biomes and different types of aquatic community from the *types* of organisms that live in them. How can we describe their similarities so that we can classify, compare and map them? In addressing this question, the Danish biogeographer Raunkiaer developed, in 1934, his idea of 'life forms', a deep insight into the ecological significance of plant forms (Figure 1.24). He then used the spectrum of life forms present in different types of vegetation as a means of describing their ecological character.

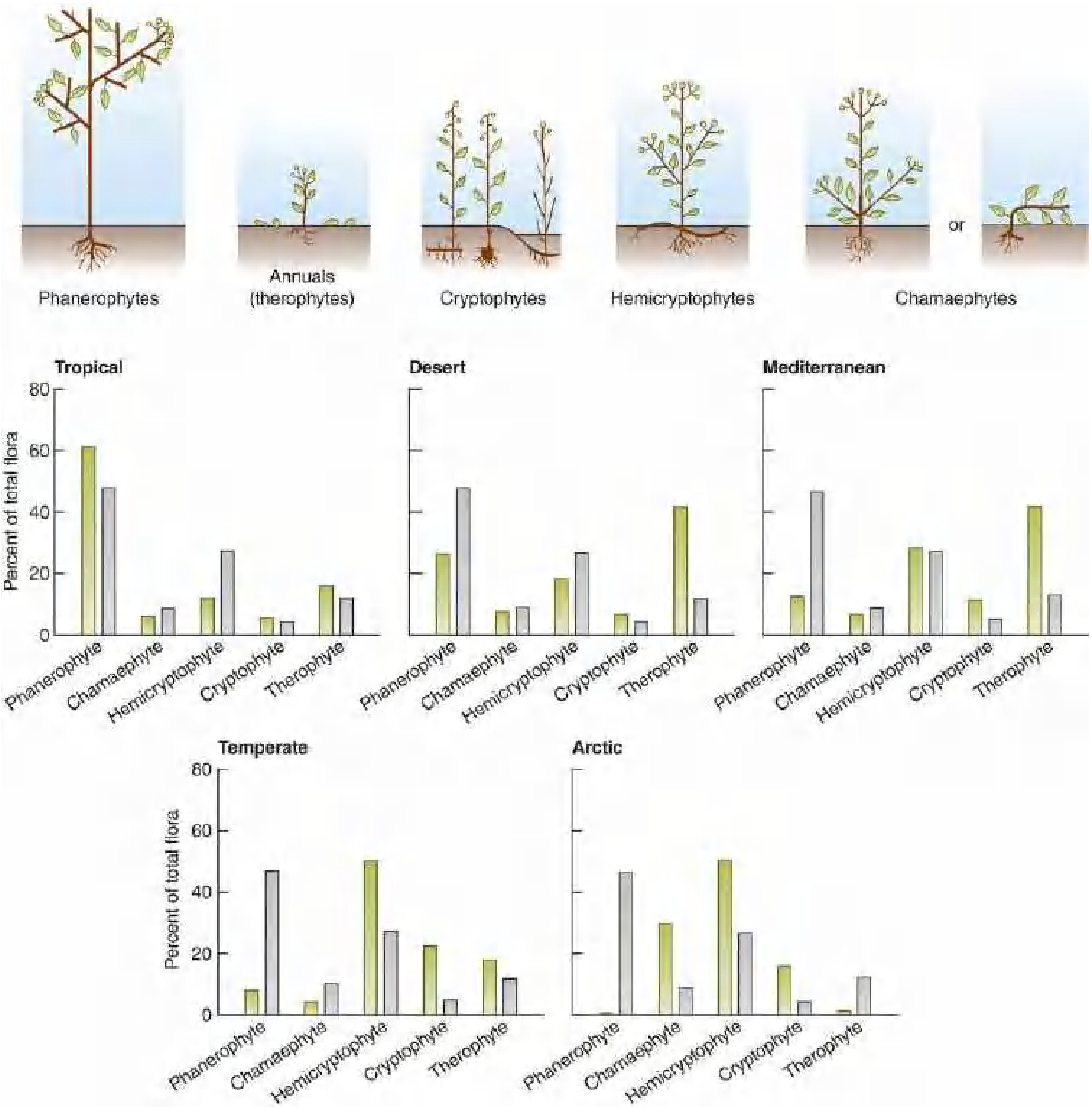


Figure 1.24 **Raunkiaer's life forms.** The drawings above depict the variety of plant forms distinguished by Raunkiaer on the basis of where they bear their buds (shown as green circles). Below are life form spectrums for five different biomes. The green bars show the percentage of the total flora that is composed of species with each of the five different life forms. The grey bars are the proportions of the various life forms in the world flora for comparison.

Source: From Crawley (1986).

Raunkiaer's plant classification

Plants grow by developing new shoots from the buds that lie at the apices (tips) of existing shoots and in the leaf axils. Within the buds, the meristematic cells are the most sensitive part of the whole shoot – the 'Achilles' heel' of plants. Raunkiaer argued that the ways in which these buds are protected in different plants are powerful indicators of the hazards in their environments and may be used to define the different plant forms (Figure 1.24). Thus, trees expose their buds high in the air, fully exposed to the wind, cold and drought; Raunkiaer called them *phanerophytes* (Greek *phanero*, 'visible'; *phyte*, 'plant'). By contrast, many perennial herbs form cushions or tussocks in which buds are borne above ground but are protected from drought and cold in the dense mass of old leaves and shoots (*chamaephytes*: 'on the ground plants'). Buds are even better protected when they are formed at or in the soil surface (*hemicryptophytes*: 'half hidden plants') or on buried dormant storage organs (bulbs, corms and rhizomes – *cryptophytes*: 'hidden plants'; or *geophytes*: 'earth plants'). These allow the plants to make rapid growth and to flower before they die back to a dormant state. A final major category consists of annual plants that depend wholly on dormant seeds to carry their populations through seasons of drought and cold (*therophytes*: 'summer plants'). Therophytes are the plants of deserts (they make up nearly 50% of the flora of Death Valley, USA), sand dunes and repeatedly disturbed habitats. They also include the annual weeds of arable lands, gardens and urban wastelands.

But there is, of course, no vegetation that consists entirely of one growth form. All vegetation contains a mixture, a spectrum, of Raunkiaer's life forms. The composition of the spectrum in any particular habitat is as good a shorthand description of its vegetation as ecologists have yet managed to devise. Raunkiaer compared these with a 'global spectrum' obtained by sampling from a compendium of all species known and described in his time (the *Index Kewensis*), biased by the fact that the tropics were, and still are, relatively unexplored. Thus, for example, we recognise a chaparral type of vegetation when we see it in Chile, Australia, California or Crete because the life form spectrums are similar. Their detailed taxonomies would only emphasise how different they are.

APPLICATION 1.6 Stream invertebrate species traits and agricultural pollution

While plant scientists have tended to be keener on classifying floras than animal scientists on classifying faunas, an analogous approach can be of use to resource managers. For example, agriculture in the catchment area of streams can have profound impacts on the invertebrate communities they contain, and just as with Raunkiaer's classification of plants, the patterns may be more closely related to species traits (paralleling Raunkiaer's life form spectra) than to taxonomic composition. Doledec *et al.* (2006) found that traits associated with population resilience (the ability to bounce back following perturbations), including short generation time and hermaphroditic reproduction, became more prevalent with increasing agricultural intensity in the catchment, reflecting more frequent and intense variations in stream nutrient concentrations (Figure 1.25). There was also a shift away from laying eggs at the water surface and a decrease in gill respiration, reflecting the increasing likelihood of smothering by sediment introduced as a result of ploughing or disturbance of soil and stream banks by grazing animals. The representation of these and other species traits can be used to devise indexes and thresholds of stream health that managers can aspire to attain or restore (Serra *et al.*, 2017).

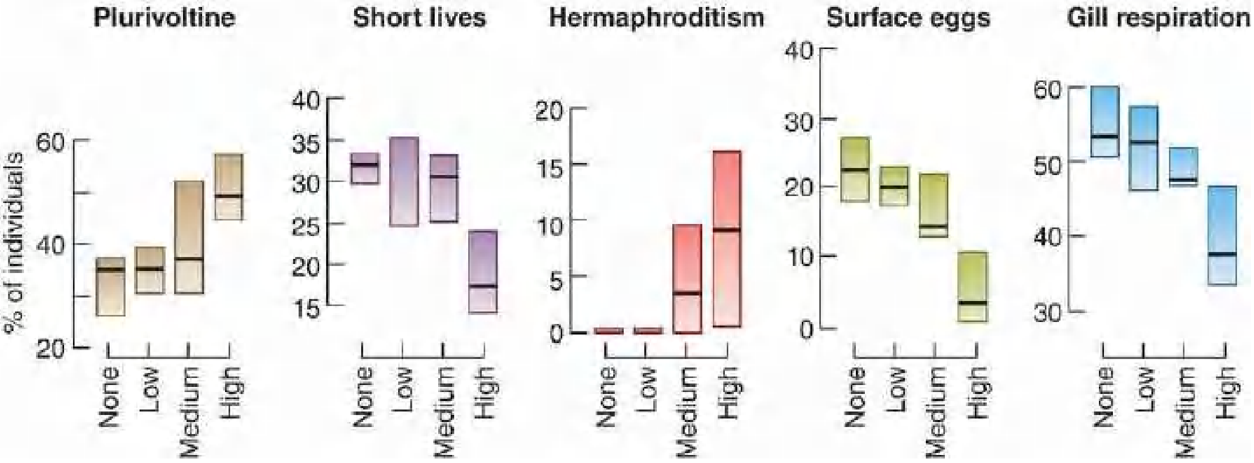


Figure 1.25 **Species traits in streams.** Relationships between the representation of species traits of stream invertebrates (% of individuals possessing the trait in the multispecies community as a whole) and the intensity of agriculture in the catchment area of the stream. None, ungrazed native tussock grassland; low, grazed native tussock grassland; medium, extensively grazed pasture; high, dairy or deer farming. Plurivoltine, more than one generation per year; short lives, 10–30 days; hermaphroditism, one individual possessing both sexes; surface eggs, laying unattached eggs at the stream surface; gill respiration, having external gills.

Source: From Doledec *et al.* (2006).

1.6 The diversity of matches within communities

Although a particular type of organism is often characteristic of a particular ecological situation, it will almost inevitably be only part of a diverse community of species. A satisfactory account, therefore, must do more than identify the similarities between organisms that allow them to live in the same environment – it must also try to explain why species that live in the same environment are often profoundly different. To some extent, this ‘explanation’ of diversity is a trivial exercise. It comes as no surprise that a plant utilising sunlight, a fungus living on the plant, a herbivore eating the plant and a parasitic worm living in the herbivore should all coexist in the same community (food webs will be discussed in Chapter 17 and the flow of energy and nutrients through ecosystems in Chapters 20 and 21). On the other hand, most communities also contain a variety of different species that are all constructed in a fairly similar way and all living (at least superficially) a fairly similar life. We have seen excellent examples among the finches of the Galápagos (Figure 1.9), the cichlid fish of Lake Apoyo (Figure 1.11), the Howea palms of Howe Island (Figure 1.12), and the picture-winged fruit-flies of Hawaii (Figure 1.16). There are several elements in an explanation of this diversity.

environments are heterogeneous

A completely homogeneous environment might well become dominated by one or a very few species that are well adapted to the conditions and resources there. But there are no homogeneous environments in nature. Even a continuously stirred culture of microorganisms is heterogeneous because it has a boundary – the walls of the culture vessel – and cultured microorganisms often subdivide into two forms: one that sticks to the walls and the other that remains free in the medium.

The extent to which an environment is heterogeneous depends on the scale of the organism that senses it. To a mustard seed, a grain of soil is a mountain; and to a caterpillar, a single leaf may represent a lifetime’s diet. A seed lying in the shadow of a leaf may be inhibited in its germination while a seed lying outside that shadow germinates freely. What appears to the human observer as a homogeneous environment may, to members of species within it, be a mosaic of the intolerable and the adequate.

There may also be gradients in space (e.g. altitude) or gradients in time, and the latter, in their turn, may be rhythmic (like daily and seasonal cycles), directional (like the accumulation of a pollutant in a lake) or erratic (like fires, hailstorms and typhoons).

Heterogeneity crops up again and again in later chapters – in part because of the challenges it poses to organisms in moving from patch to patch (Chapter 6), in part because of the variety of opportunities it provides for different species (Chapters 2 and 3), and in part because heterogeneity can alter communities by interrupting what would otherwise be a steady march to an equilibrium state of a few species (Chapters 8 and 18).

pairs of species

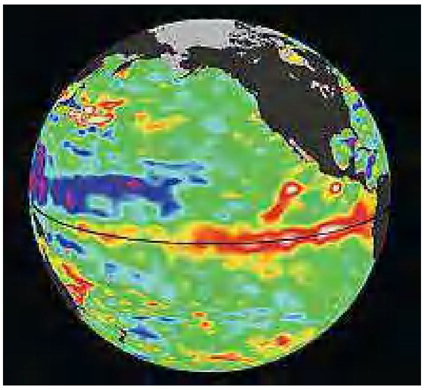
It is important to note that the existence of one type of organism in an area immediately diversifies it for others. Over its lifetime, an organism may increase the diversity of its environment by contributing dung, urine, dead parts (e.g. skin or leaves) and ultimately its dead body. During its life, its body may serve as a place in which other species find homes. Indeed, some of the most strongly developed matches between organisms and their environment are those in which one species has developed a dependence upon another. This is the case in many relationships between consumers and their foods. Whole syndromes of form, behaviour and metabolism constrain the animal within its narrow food niche, and deny it access to what might otherwise appear suitable alternative foods. Similar tight matches are characteristic of the relationships between parasites and their hosts. The various interactions in which one species is consumed by another are the subject matter of Chapters 8–10 and 12.

Where two species have evolved a mutual dependence, the fit may be even tighter. We examine such ‘mutualisms’ in detail in Chapter 13. The association of nitrogen-fixing bacteria with the roots of leguminous plants, and the often extremely precise relationships between insect pollinators and their flowers, are two good examples. When a population has been exposed to variations in the physical factors of the environment, for example a short growing season or a high risk of frost or drought, a once-and-for-all tolerance may ultimately evolve. The physical factor cannot itself change or evolve as a result of the evolution of the organisms. By contrast, when members of two species interact, the change in each produces alterations in the life of the other, and each may generate selective forces that direct the evolution of the other. In such a coevolutionary process the interaction between two species may continually escalate. What we then see in nature may be pairs of species that have driven each other into ever-narrowing ruts of specialisation – an ever closer match.

coexistence of similar species

While it is no surprise that species with rather different roles coexist within the same community, it is also generally the case that communities support a variety of species performing apparently rather similar roles. Do these species compete with one another? Do competing species need to be different if they are to coexist? If so, how different do they need to be: is there some limit to their similarity? Do species like Darwin's finches interact with one another at the present time, or has evolution in the past led to the absence of such interactions in contemporary communities? We return to these questions about coexisting, similar species in Chapter 8, and take them up again in Chapters.

Even at this stage, though, we may note that coexisting species, even when apparently very similar, commonly differ in subtle ways – not simply in their morphology or physiology but also in their responses to their environment and the role they play within the community of which they are part. The ‘ecological niches’ of such species are said to be differentiated from one another. The concept of the ecological niche is itself explained in the next two chapters.



Chapter 2
Conditions

2.1 Introduction

In order to understand the distribution and abundance of a species we need to know its history (Chapter 1), the resources it requires (Chapter 3), the individuals' rates of birth, death and migration (Chapters 4 and 6), their interactions with their own and other species (Chapters 5 and 8–13) and the effects of environmental conditions. This chapter deals with the limits placed on organisms by environmental conditions.

conditions may be altered – but not consumed

A condition is an abiotic environmental factor that influences the functioning of living organisms. Examples include temperature, relative humidity, pH, salinity and the concentration of pollutants. A condition may be modified by the presence of other organisms. For example, temperature, humidity and soil pH may be altered under a forest canopy. But unlike resources, conditions are not consumed or used up by organisms.

For some conditions we can recognise an optimum concentration or level at which an organism performs best, with its activity tailing off at both lower and higher levels (Figure 2.1a). But what is meant by 'performs best'? From an evolutionary point of view, 'optimal' conditions are those under which individuals leave most descendants (are fittest), but these are often impossible to determine in practice because measures of fitness should be made over several generations. Instead, we more often measure the effect of conditions on some key property like the activity of an enzyme, the respiration rate of a tissue, the growth rate of individuals or their rate of reproduction. However, the effect of variation in conditions on these various properties will often not be the same; organisms can usually survive over a wider range of conditions than permit them to grow or reproduce (Figure 2.1a).

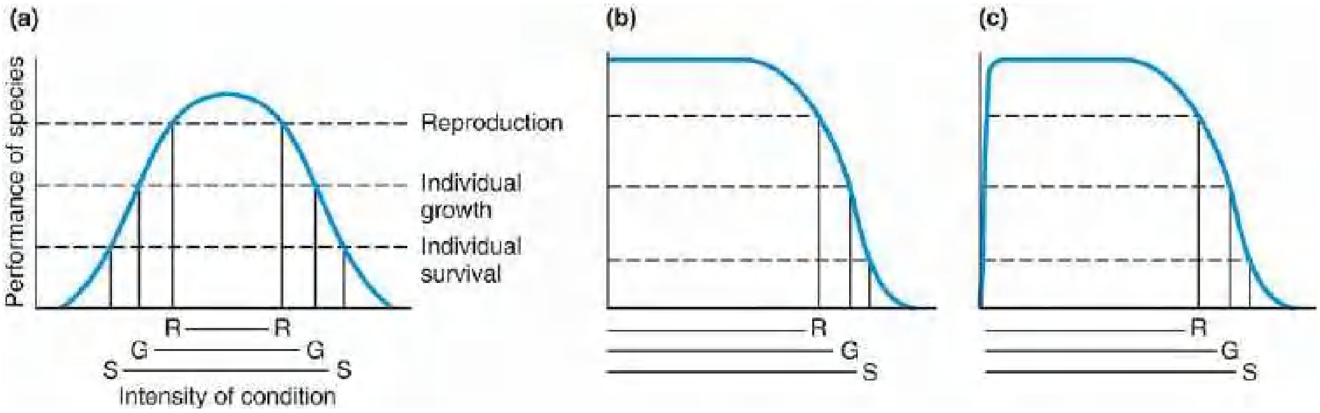


Figure 2.1 Response curves illustrating the effects of a range of environmental conditions on individual survival (S), growth (G) and reproduction (R). (a) Extreme conditions are lethal, less extreme conditions prevent growth, and only optimal conditions allow reproduction. (b) The condition is lethal only at high intensities; the reproduction–growth–survival sequence still applies. (c) Similar to (b), but the condition is required by organisms, as a resource, at low concentrations.

The precise shape of a species' response will vary from condition to condition. The generalised form of response, shown in Figure 2.1a, is appropriate for conditions like temperature and pH in which there is a continuum from an adverse or lethal level (e.g. freezing or very acid conditions), through favourable levels of the condition to a further adverse or lethal level (heat damage or very alkaline conditions). There are, though, many environmental conditions for which Figure 2.1b is a more appropriate response curve: for most toxins, including radioactive emissions and chemical pollutants, a low-level intensity or concentration of the condition has no detectable effect, but an increase begins to cause damage and a further increase may be lethal. There is also a different form of response to conditions that are toxic at high levels but essential for growth at low levels (Figure 2.1c). This is the case for sodium chloride – an essential resource for animals but lethal at high concentrations – and for the many elements that are essential micronutrient resources in the growth of plants and animals (e.g. copper, zinc and manganese), but that can become lethal at the higher concentrations sometimes caused by industrial pollution.

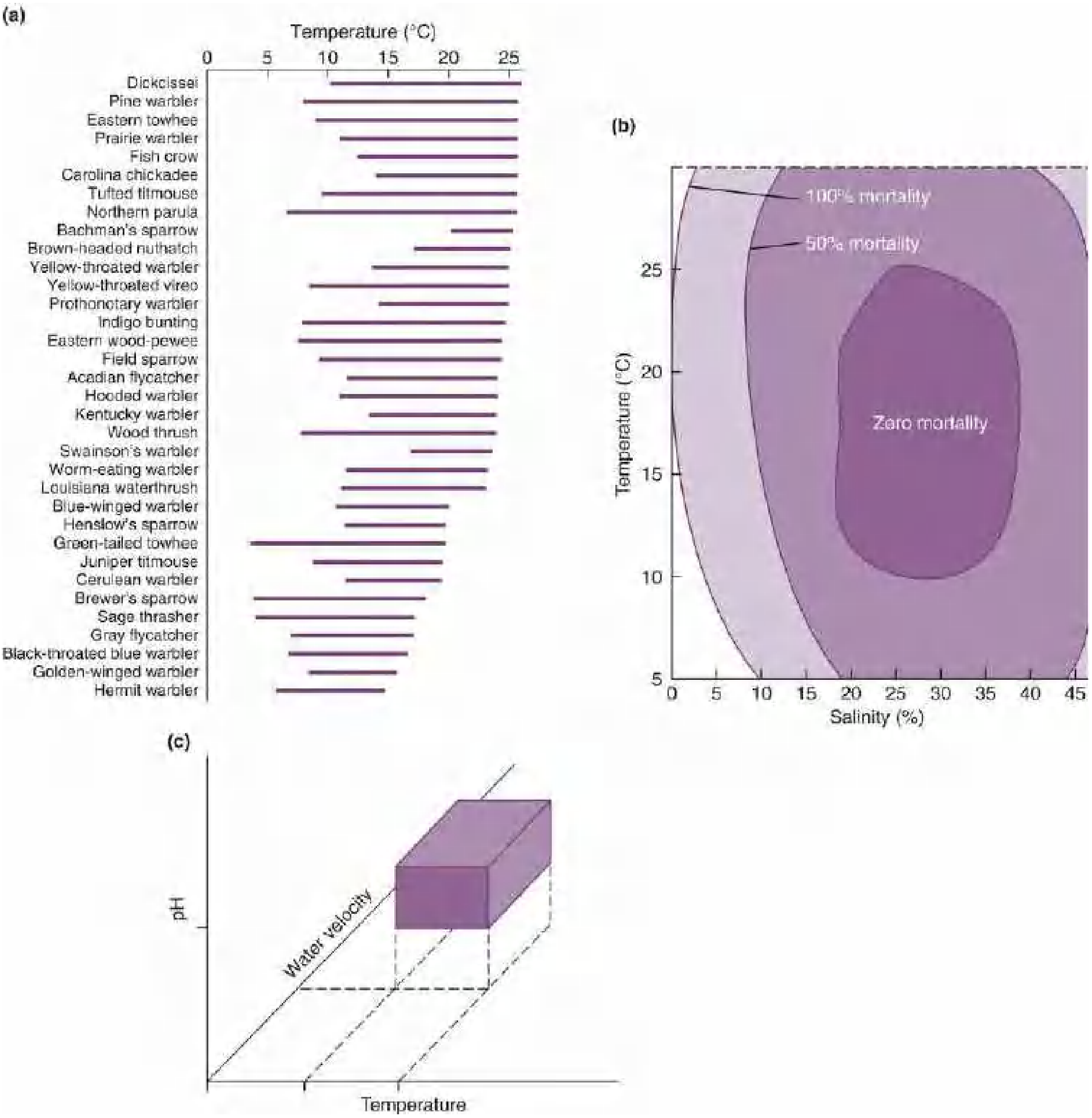
In this chapter, we consider responses to temperature in much more detail than other conditions, because it is the single most important condition that affects the lives of organisms, and many of the generalisations that we make have widespread relevance. We move on to consider a range of other conditions, before returning, full circle, to temperature because of the way that other conditions interact with it. We begin, though, by explaining the framework within which each of these conditions should be understood: the ecological niche.

2.2 Ecological niches

The term *ecological niche* is frequently misunderstood. It is often misused to describe the sort of place in which an organism lives, as in the sentence: 'Woodlands are the niche of woodpeckers'. Strictly, however, where an organism lives is its *habitat*. A niche is not a place but an idea: a summary of the organism's tolerances and requirements. The habitat of a gut microorganism would be an animal's alimentary canal; the habitat of an aphid might be a garden; and the habitat of a fish could be a whole lake. Each habitat, however, provides many different niches: many other organisms also live in the gut, the garden or the lake – and with quite different lifestyles. The word *niche* began to gain its present scientific meaning when Elton wrote in 1933 that the niche of an organism is its mode of life 'in the sense that we speak of trades or jobs or professions in a human community'. The niche of an organism started to be used to describe how, rather than just where, an organism lives.

niche dimensions

The modern concept of the niche was proposed by Hutchinson in 1957 to address the ways in which tolerances and requirements interact to define the conditions (this chapter) and resources (Chapter 3) needed by an individual of a species in order to practice its way of life. Temperature, for instance, limits the growth and reproduction of all organisms, but different organisms tolerate different ranges of temperature. This range is one *dimension* of an organism's ecological niche. Figure 2.2a shows how species of passerine birds in North America vary in this dimension of their niche. But there are many such dimensions of a species' niche – its tolerance of various other conditions (relative humidity, pH, wind speed, water flow and so on) and its need for various resources. Clearly the real niche of a species must be *multi*dimensional.



[**Figure 2.2**](#) **The ecological niche in one, two and three dimensions.** (a) A niche in one dimension showing the thermal range of passerine birds in southern Canada and the contiguous USA recorded during the North American Breeding Bird Survey 2002–06 in relation to minimum and maximum thermal limits of an average of 10 occurrence locations for each species (measured in each case as mean springtime breeding season temperature). (b) A niche in two dimensions for the sand shrimp (*Crangon septemspinosa*) showing the fate of egg-bearing females in aerated water at a range of temperatures and salinities. (c) A diagrammatic niche in three dimensions for a stream-dwelling alga showing a volume defined by temperature, pH and water velocity; in reality, the niche would not appear as a neat cuboid defined by the three tolerance ranges because, for example, temperature tolerance may be reduced when pH is low.

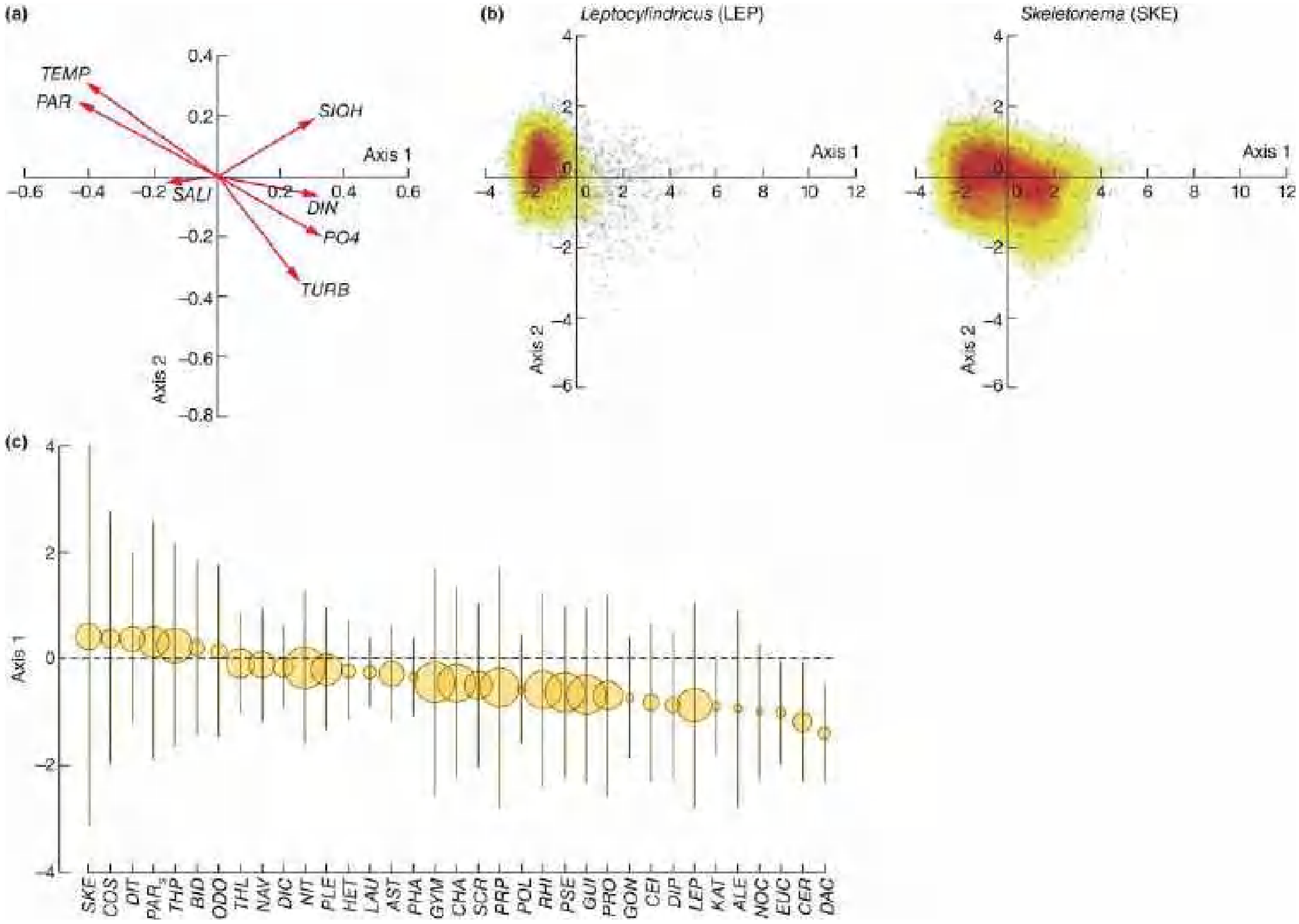
Source: (a) Data from Coristine & Kerr (2015). (b) After Haefner (1970).

the *n*-dimensional hypervolume

It is easy to visualise the early stages of building such a multidimensional niche. Figure 2.2b illustrates the way in which two niche dimensions (temperature and salinity) together define a two-dimensional area that is part of the niche of a sand shrimp. Three dimensions, such as temperature, pH and current velocity in a stream, may define a three-dimensional niche volume of a stream alga (Figure 2.2c). In fact, we consider a niche to be an *n-dimensional hypervolume*, where *n* is the number of dimensions that make up the niche. It is hard to imagine (and impossible to draw) this more realistic picture. Nonetheless, the simplified three-dimensional version captures the idea of the ecological niche of a species. It is defined by the boundaries that limit where it can live, grow and reproduce, and it is very clearly a concept rather than a place. The concept has become a cornerstone of ecological thought.

ordination as an aid to conceiving the *n*-dimensional niche

The difficulties of interpreting the multiplicity of relevant niche dimensions can be reduced by a mathematical technique called *ordination*. This is one of several methods used by ecologists to condense information from many dimensions into a much smaller, more manageable number, in this case allowing us to simultaneously display species and several influential environmental variables along one or more ‘ordination axes’. Species with the most similar niches appear closest together, and the direction of increase or decrease of environmental variables along each axis reveals how the species’ niches are arranged in relation to these variables. In their study of marine phytoplankton along the French coast, Farinas *et al.* (2015) related abundance data for 35 taxa to seven environmental factors: temperature, salinity and turbidity (conditions), and photosynthetically active radiation and the concentrations of three inorganic nutrient concentrations (resources for phytoplankton). The relationships of these variables along two ordination axes derived by the method are displayed in Figure 2.3a. Note, for example, that nutrient concentrations are positively related to the first axis, while salinity, solar radiation and temperature are negatively related to this axis. Figure 2.3b illustrates for two taxa, *Leptocylindrus* spp. and *Skeletonema* spp., the space occupied along the first and second axes of the ordination. *Leptocylindrus* has a narrower niche than *Skeletonema*, and *Leptocylindrus* is displaced towards the negative end of axis 1 and the positive end of axis 2, being more strongly related than *Skeletonema* to temperature and photosynthetically active radiation. Figure 2.3c shows the space occupied by all 35 taxa along axis 1: those with more negative positions are associated with higher temperatures, salinity and photosynthetically active radiation levels; those with more positive positions are associated with higher nutrient concentrations. While a hypervolume with more than three dimensions cannot be visualised, ordination allows it to be more readily comprehended, and hence allows us to see how much species’ niches overlap, which of them are quite distinct, and so on.



[Figure 2.3](#) **The use of ordination to facilitate understanding of the multidimensional niche.** (a) Weights along two ordination axes of seven environmental factors (*TEMP*, water temperature; *PAR*, photosynthetically active radiation; *SALI*, salinity; *TURB*, turbidity; *PO4*, phosphates; *DIN*, dissolved inorganic nitrogen; *SIOH*, silicates) used to characterise the ecological niche of 35 phytoplankton taxa in French coastal seas. (b) Space occupied by two of the taxa, *Leptocylindrus* (LEP) and *Skeletonema* (SKE) spp., along the first and second axes of the ordination analysis. The yellow to red colour gradient represents phytoplankton density from low to high. (c) Space occupied by each taxon along axis 1 of the ordination. The diameter of the circle is proportional to the total occurrence frequency of each taxon. Axis 1 is positively related to nutrient concentrations and negatively related to temperature, salinity and photosynthetically active radiation.

Source: From Farinas *et al.* (2015).

ecological niche modelling approach to the multidimensional niche

Another approach to characterising a multidimensional niche makes use of *ecological niche models* (also known as *climate matching* or *climate envelope models*) (Jeschke & Strayer, 2008). A species' niche characteristics, defined largely by its physiology, are fairly constant, so that it is not unreasonable to expect that the details of a species' niche in one location may be broadly transferable to another. This is the basis for *ecological niche modelling* (Peterson, 2003), where occurrence patterns in a species' native range are used to build a model that can be projected to identify other areas that are potentially habitable, using one of several available software packages: BIOCLIM, GARP, MAXENT and others (Elith & Graham, 2009). The basic process of niche modelling is outlined in [Figure 2.4](#). As much environmental information as possible is taken from all of the locations where a species is currently found and from a range of locations where the species has not been recorded, allowing those locations to be identified that meet the species' requirements even though the species is currently absent. The ability to project into geographic space can be used to predict species distributions in previously unexplored parts of the native range (checking how good the model is) or in new, often quite distant locations of interest (e.g. predicting places where a potentially invasive species may prove problematic; Figure 2.5).

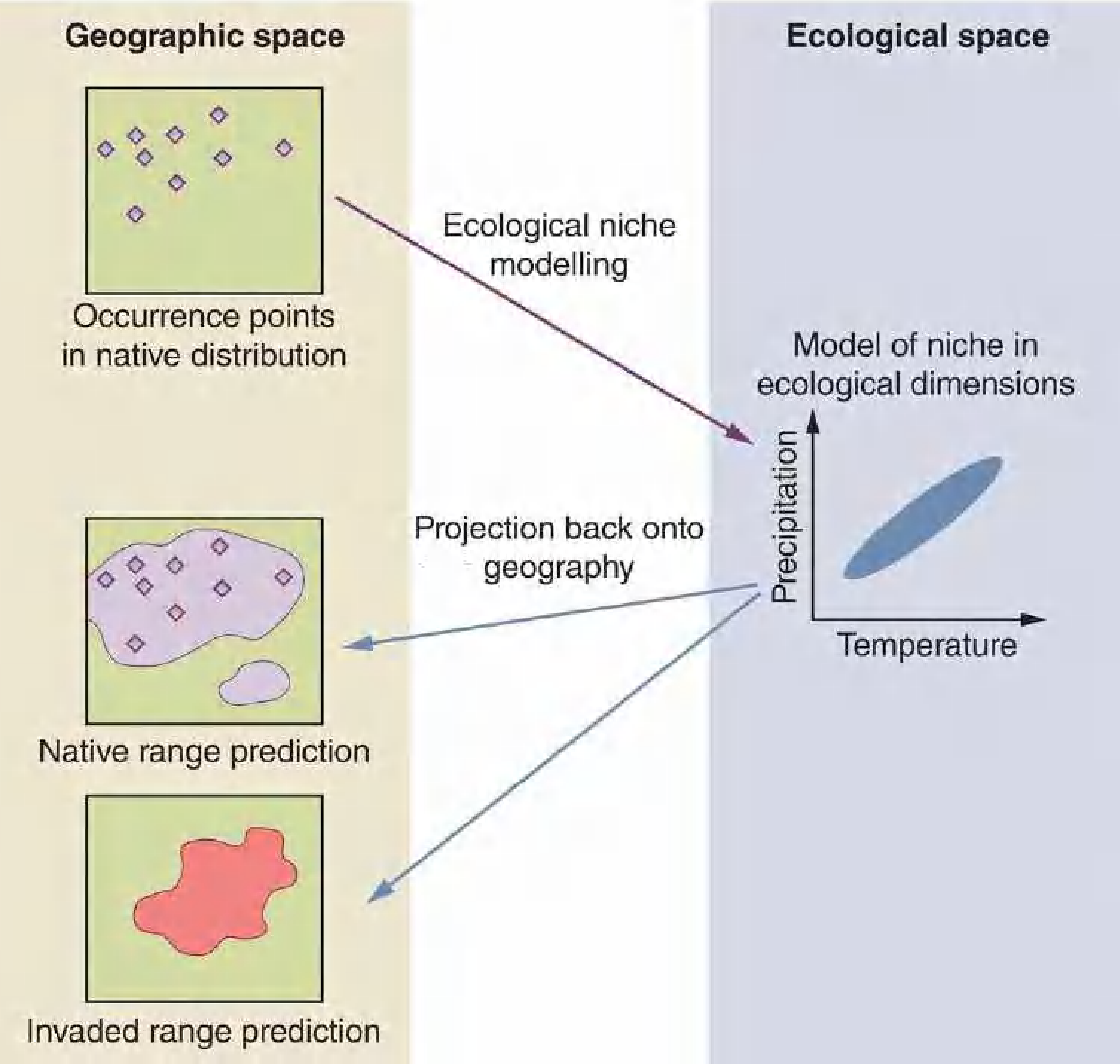


Figure 2.4 **Ecological niche modelling.** The first step is to characterise a species' distribution in two-dimensional geographic space. Then the niche is modelled in ecological space, in terms of a number of influential dimensions of the *n*-dimensional hypervolume (such as temperature, precipitation, humidity, soil pH, etc.). Finally, the occupation of ecological space is projected back into geographic space.

Source: After Peterson (2003).

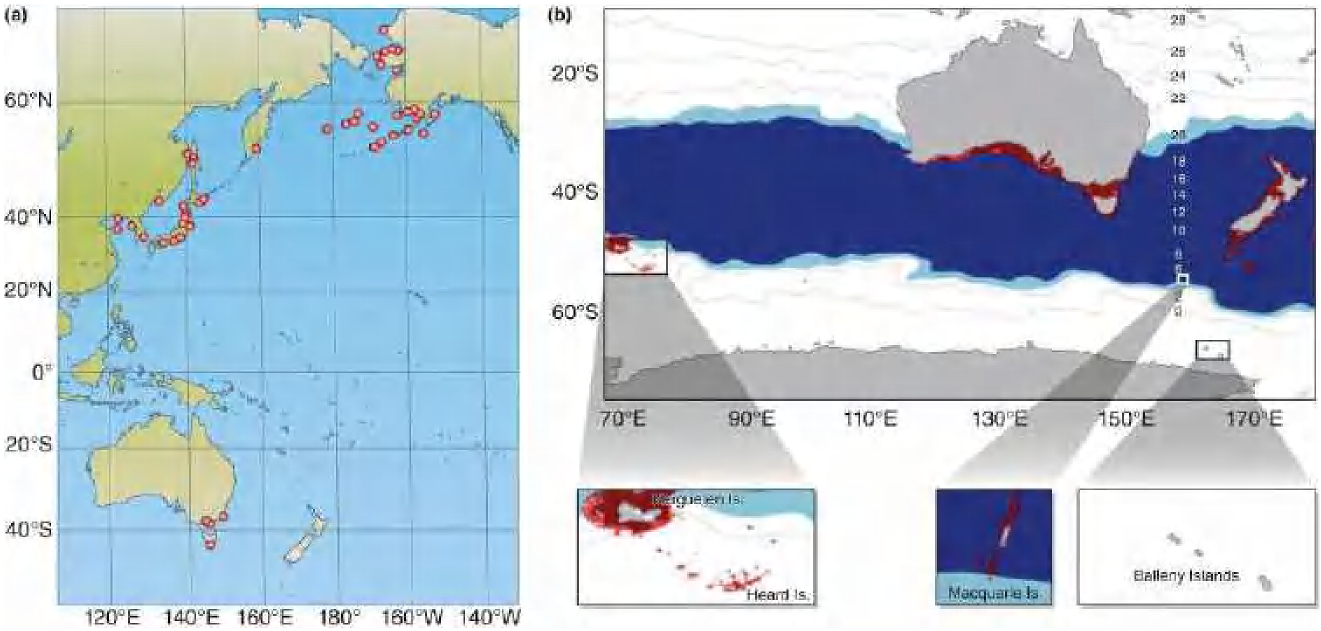


Figure 2.5 Modelling the potential range of an invasive starfish. (a) Current distribution records for the sea star *Asterias amurensi* in its native (northern hemisphere) and invasive (southern hemisphere) range. (b) Modelled distribution in its invasive range. Red regions represent areas with suitable mean winter and summer seafloor temperatures for the benthic adult stage (light red suitable, dark red highly suitable). Blue regions represent areas where the sea surface temperature is suitable for the pelagic larval stages (dark blue optimal). Isotherms represent mean sea surface temperature (°C) during winter. Boxes show islands that might provide a stepping-stone habitat for invasion of *A. amurensis* into Antarctica, especially the Macquarie, Heard and Kerguelen Islands, which are ice-free year-round. Currently the Balleny Islands are only ice-free in summer but this may change with global warming.

Source: From Byrne *et al.* (2016).

APPLICATION 2.1 Ecological niche modelling and ordination as management tools

Managers are frequently confronted by problems associated with invasive species and make use of *climate envelope models* or *ordination* to develop solutions.

The Arctic sea star, *Asterias amurensis*, is among the most ecologically influential of marine invertebrates, being a voracious predator with a particular affinity for bivalves (frequently putting it in conflict with bivalve fishers) and capable of dramatically affecting local biodiversity. Its native range extends in the North Pacific from the Arctic to southern Japan (Figure 2.5a). Accidentally introduced in the early 1980s to Tasmania (probably through the release of pelagic larvae in ship's ballast water), adults became established on the seabed where they caused the extinction of many species. *A. amurensis* has since spread to Victoria along the coast of mainland Australia (Figure 2.5a) but so far it has not invaded New Zealand or the sub-Antarctic Islands. One critical dimension of its multidimensional niche is water depth: the species cannot survive below a depth of 200 m. Both summer and winter temperature ranges are also fundamentally important to the success of the sea stars, and so to assess the potential for range expansion, Byrne *et al.* (2016) used the climate envelope model MaxEnt to characterise the thermal niche of both adults and the dispersive larval stages. Figure 2.5b shows the predicted invasive range, which includes much of New Zealand, together with the sub-Antarctic Macquarie, Heard and Kerguelen Islands. The red areas are considered suitable for adult sea stars (dark red highly suitable), while the blue zones are suitable for the development of dispersing larval stages (dark blue optimal). That the species may spread to many new locations is alarming enough, but there is also a strong possibility that global warming will put much of the Antarctic coastline in peril. Results of such analyses highlight the importance of vigilance and border biosecurity.

Marchetti and Moyle (2001) used an *ordination* technique to define how a suite of fish species, 11 native and 14 invaders, are related to environmental factors in a Californian river (Figure 2.6). The native and invasive species clearly occupy different parts of the multidimensional niche space. Most of the natives were associated with higher mean discharge ($m^3 s^{-1}$), good canopy cover (higher levels of % shade), lower concentrations of plant nutrients (lower conductivity, µS), lower temperatures and a greater percentage of fast-flowing, riffle habitat (less pool habitat). These are all features of the natural, undisturbed state of streams. The invaders, on the other hand, are favoured by the present combination of conditions where water regulation and damming have reduced discharge and riffle habitat, shady riparian vegetation has been removed leading to higher stream temperatures, and nutrient concentrations have increased because of agricultural and domestic runoff. Restoration of more natural flow regimes and riparian planting will be needed to halt the continued decline of native fish, and it is heartening to note that hundreds of dams across the USA, whether originally built for public or private benefit, have been removed in river restoration projects in recent years.

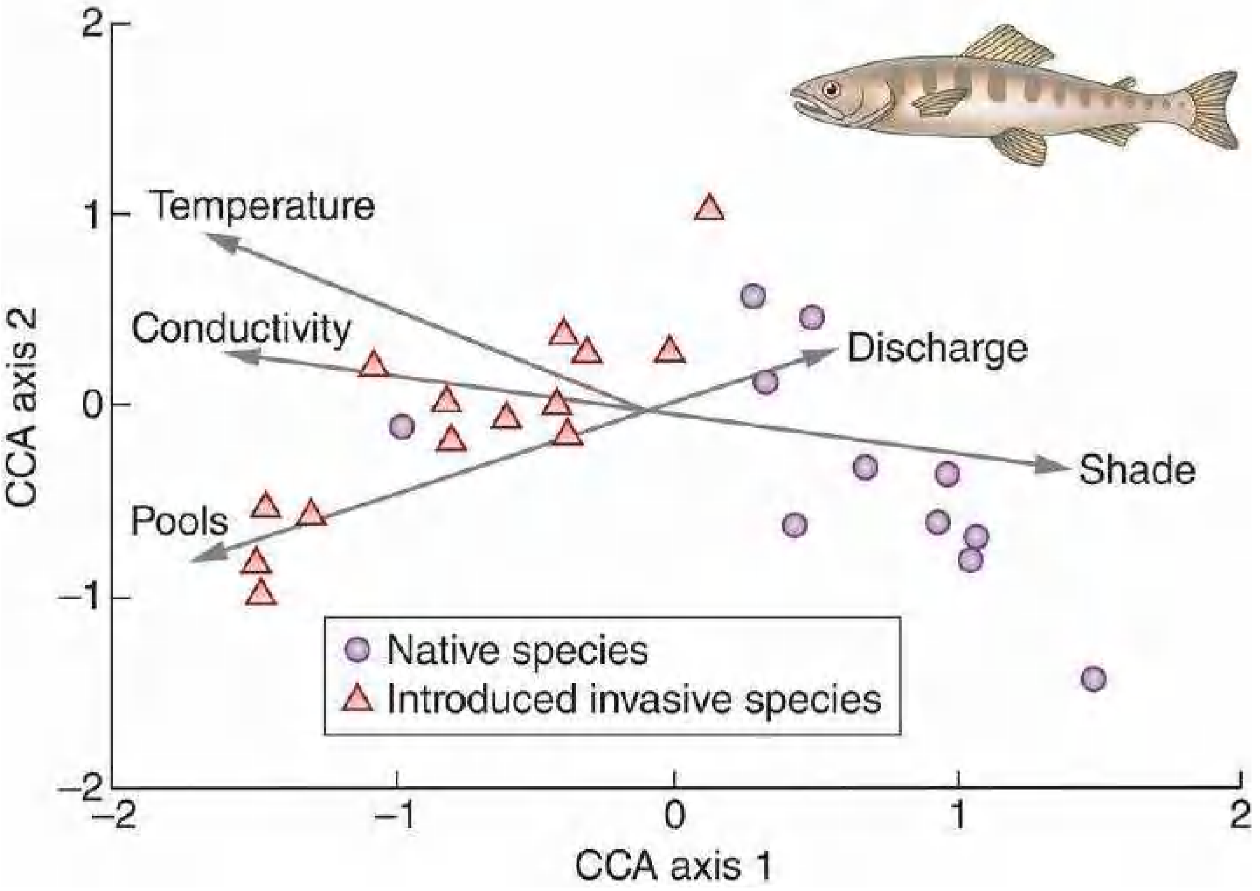


Figure 2.6 Ordination contrasts the multidimensional niches of native and invasive fish. Plot of results of an ordination technique called canonical correspondence analysis (CCA) showing native species of fish (purple circles), introduced invasive species (red triangles) and five influential environmental variables. Note how the native and invasive species occupy different parts of multidimensional niche space.

Source: After Marchetti & Moyle (2001).

fundamental and realised niches

Provided that a location is characterised by conditions within acceptable limits for a given species, and provided also that it contains all the necessary resources, then the species can, potentially, occur and persist there. Whether or not it does so depends on two further factors. First, as we have just seen, it must be able to reach the location, and this depends in turn on its powers of colonisation and the remoteness of the site, or on human agency in spreading invasive species from one area to another. Second, its occurrence may be precluded by the action of individuals of other species that compete with, prey upon or parasitise it.

Usually, a species has a larger ecological niche in the absence of enemies than it has in their presence. In other words, there are certain combinations of conditions and resources that can allow a species to maintain a viable population, but only if it is not being adversely affected by enemies. This led Hutchinson to distinguish between the *fundamental* and the *realised* niche. The former describes the overall potentialities of a species; the latter describes the more limited spectrum of conditions and resources that allow it to persist, even in the presence of competitors, predators and parasites. One of the acknowledged shortcomings of the modelling of niches based on distributions in species' native ranges, described earlier, is that it is the realised niche that is under consideration (on the assumption that competitors, predators and parasites are present and exert an effect). When a species invades a new area, there is every possibility that some or all of its native enemies will be absent, so that it may be able to occupy an expanded niche, closer to its fundamental niche. Modellers need to beware this possibility (Jeschke & Strayer, 2008).

Just as negative interactions can play a role in determining species' distributions (leading to a realised niche smaller than the fundamental niche), so can the positive effects of mutualists that we discuss in more detail in Chapter 13 (potentially producing a realised niche larger than the fundamental niche). Take, for example, the tropical anemone fish *Amphiprion chrysopterus*, which retreats between the stinging tentacles of the sea anemone *Heteractis magnifica* when predators threaten, but protects the anemone against its grazers, increasing anemone survivorship, growth and reproduction (Holbrook & Schmitt, 2005). Either species may tolerate the conditions at a location, but their success also depends on the presence of the other. In similar vein, most higher plants have intimate mutualistic associations between their roots and fungi (mycorrhiza; Section 13.9) that capture nutrients from the soil and transfer them to the plants, as well as improving water uptake and disease resistance, while receiving photosynthetic products from the plant (Delavaux *et al.*, 2017). Many plants can live without their mycorrhizal associates in soils when water and nutrients are in good supply, but in the highly competitive world of plant communities the presence of the fungi is often necessary if the plant is to prosper.

APPLICATION 2.2 Judging the fundamental niche of a species driven to extreme rarity

The takahe (*Porphyrio hochstetteri*), one of only two remaining species of large, herbivorous, flightless birds that dominated the pre-human New Zealand landscape, was itself believed extinct until rediscovery in 1948 of a small population in the remote and climatically extreme Murchison Mountains in the south-west of the South Island (Figure 2.7). Intense conservation efforts have involved captive breeding, habitat management, predator control, wild releases into the Murchison Mountains and nearby ranges as well as translocations to offshore islands that lack the introduced mammals that are now widespread on the mainland (Lee & Jamieson, 2001). From just a handful of individuals, there are now more than 300 in existence. Some ecologists believed that because the takahe is a grassland specialist, feeding mainly on tussock grasses in the genus *Chionocloa*, and adapted to the alpine zone, they would not fare well elsewhere. Others noted that fossil evidence indicated that takahe were once widespread in New Zealand and occurred at altitudes below 300 m, including coastal areas that were a mosaic of forest, shrubland and grassland (Figure 2.7), and that they might therefore be well suited to life on offshore islands that lack the mammals that have caused their demise. Indeed, they have formed self-sustaining populations after introduction to four offshore islands, although the island habitat may not be optimal (with poorer hatching and fledging success in island than mountain populations) (Jamieson & Ryan, 2001). The fundamental niche of takahe probably encompasses much of the South Island, but it became confined to a much smaller realised niche because of the effects of predators (human hunters and introduced stoats, *Mustela erminea*) and competitors for food (introduced red deer, *Cervus elaphus scoticus*). The removal of these mammalian interlopers would enable takahe to occupy something closer to their fundamental niche, as they did before humans and the other invaders arrived in New Zealand.

Figure 2.7 Location of fossil bones of the takahe in the South Island of New Zealand. The population had become restricted to a single site in the Murchison Mountains, but was this a true reflection of its niche requirements?

Source: After Trewick & Worthy (2001).

The remainder of this chapter looks at some of the most important condition dimensions of species' niches, starting with temperature; the following chapter examines resources, which add further dimensions of their own.

2.3 Responses of individuals to temperature

2.3.1 What do we mean by 'extreme'?

It seems natural to describe certain environmental conditions as 'extreme', 'harsh', 'benign' or 'stressful'. It may seem obvious when conditions are 'extreme': the midday heat of a desert, the cold of an Antarctic winter, the salinity of the Great Salt Lake. But this only means that these conditions are extreme *for us*, given our particular physiological characteristics and tolerances. To a cactus there is nothing extreme about the desert conditions in which cacti have evolved; nor are the icy fastnesses of Antarctica an extreme environment for penguins. It is lazy and dangerous for the ecologist to assume that all other organisms sense the environment in the way we do. Rather, the ecologist should try to gain a worm's-eye or plant's-eye view of the environment: to see the world as others see it. Emotive words like harsh and benign, even relativities such as hot and cold, should be used by ecologists only with care.

2.3.2 Metabolism, growth, development and size

exponential effects of temperature on metabolic reactions

Individuals respond to temperature essentially in the manner shown in Figure 2.1a: impaired function and ultimately death at the lower and upper extremes (discussed in Sections 2.3.4 and 2.3.6), with a functional range between the extremes, within which there is an optimum. This is accounted for, in part, simply by changes in metabolic effectiveness. For each 10°C rise in temperature, for example, the rate of biological enzymatic processes often roughly doubles, and thus appears as an exponential curve on a plot of rate against temperature (Figure 2.8a, b). The increase is brought about because a higher temperature increases the speed of molecular movement and speeds up chemical reactions. The factor by which a reaction changes over a 10°C range is referred to as Q_{10}: a rough doubling means that $Q_{10} \approx 2$, and animals generally conform quite closely to this value (Figure 2.8c) as do microbial organisms (Kirchman, 2012) and plants (Lange *et al.*, 2012).

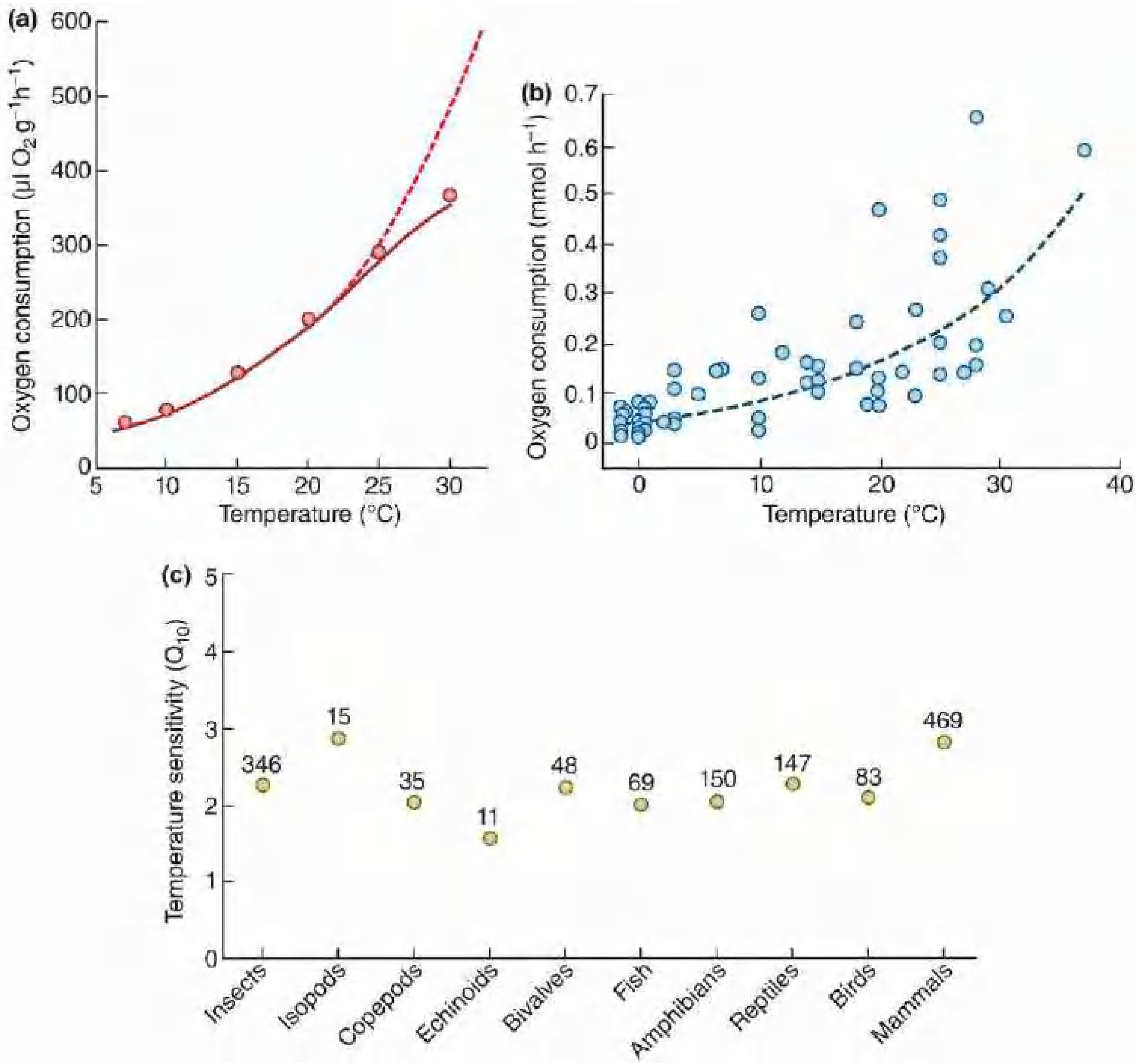
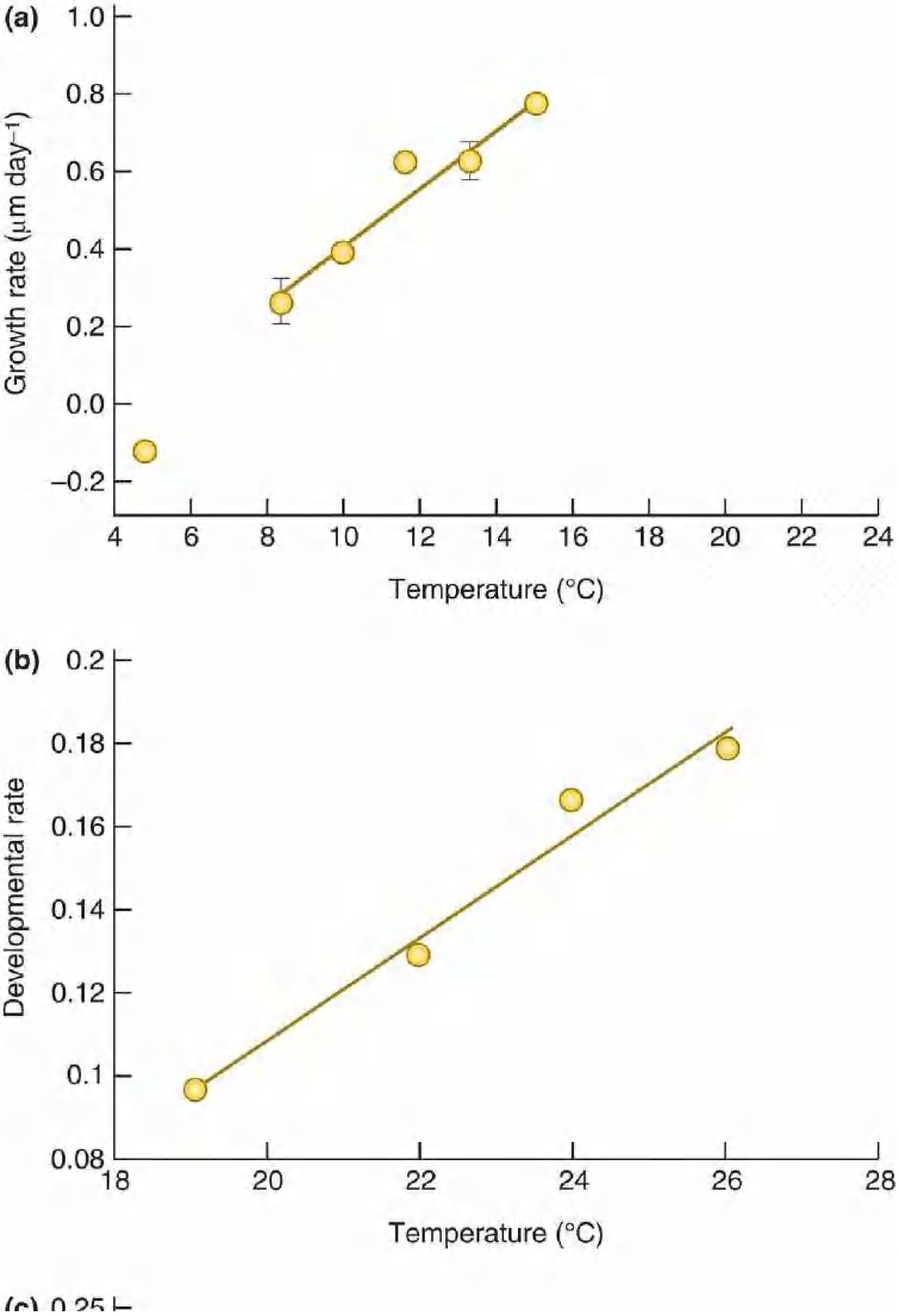


Figure 2.8 **Exponential effects of temperature on metabolic reactions.** (a) The rate of oxygen consumption of the Colorado beetle (*Leptinotarsa decemineata*), which doubles for every 10°C rise in temperature up to 20°C, but increases less fast at higher temperatures. (b) Relationship between resting oxygen consumption of teleost fish, plotted for a variety of species as oxygen consumption at their typical environmental temperature for a standardized 50 g fish. (c) Q_{10} values for oxygen consumption of various invertebrate and vertebrate animal taxa, averaged across multiple published studies (numbers shown).

Source: (a) After Marzusch (1952). (b, c) Data compiled from various sources by Clarke (2017).

effectively linear effects on rates of growth and development

For an ecologist, however, effects on individual chemical reactions are likely to be less important than effects on rates of growth (increases in mass), on rates of development (progression through lifecycle stages) and on final body size, since, as we shall discuss much more fully in Chapter 4, these tend to drive the core ecological activities of survival, reproduction and movement. And when rates of growth and development of whole organisms are plotted against temperature, there is quite commonly an extended range over which there are, at most, only slight deviations from linearity (Figure 2.9).


(a)
Growth rate (μm day⁻¹)
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
4
6
8
10
12
14
16
18
20
22
24
Temperature (°C)
(b)
Developmental rate
0.2
0.18
0.16
0.14
0.12
0.1
0.08
18
20
22
24
26
28
Temperature (°C)
(c) 0.25

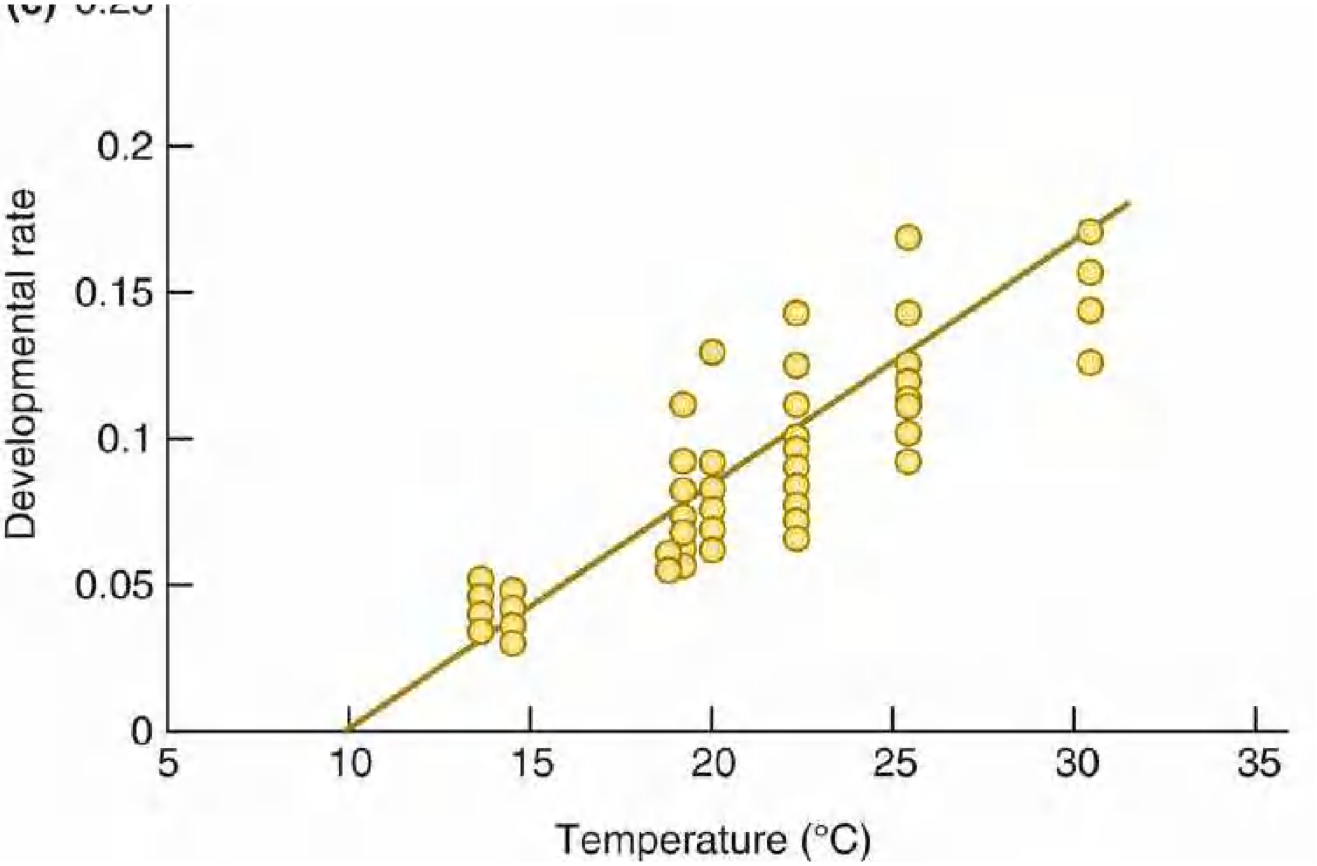


Figure 2.9 Effectively linear relationships between rates of growth and development and temperature. (a) Growth of the protist *Strombidinopsis multiauris*. (b) Egg development in the beetle *Oulema duftschmidi*. (c) Egg to adult development in the mite *Amblyseius californicus*. The vertical scales in (b) and (c) represent the proportion of total development achieved in one day at the temperature concerned.

Source: (a) After Montagnes *et al.* (2003). (b) After Severini *et al.* (2003). (c) After Hart *et al.* (2002).

day-degree concept

When the relationship between growth or development *is* effectively linear, the temperatures experienced by an organism can be summarised in a single very useful value, the number of 'day-degrees'. For instance, Figure 2.9c shows that at 15°C (5.1°C above a development threshold of 9.9°C) the predatory mite, *Amblyseius californicus*, took 24.22 days to develop (i.e. the proportion of its total development achieved each day was 0.041 (= 1/24.22)), but it took only 8.18 days to develop at 25°C (15.1°C above the same threshold). At both temperatures, therefore, development required 123.5 day-degrees (or, more properly, 'day-degrees above threshold'), i.e. 24.22 × 5.1 = 123.5, and 8.18 × 15.1 = 123.5. This is also the requirement for development in the mite at other temperatures within the non-lethal range. Such organisms cannot be said to require a certain length of time for development. What they require is a combination of time and temperature, often referred to as 'physiological time'.

temperature–size rule

Together, the rates of growth and development determine the final size of an organism. For instance, for a given rate of growth, a faster rate of development will lead to smaller final size. Hence, if the responses of growth and development to variations in temperature are not the same, temperature will also affect final size. In fact, development usually increases more rapidly with temperature than does growth, such that, for a very wide range of organisms, final size tends to decrease with increasing temperature: the ‘temperature–size rule’ (see Atkinson *et al.*, 2003). An example for single-celled protists (72 datasets from marine, brackish and freshwater habitats) is shown in Figure 2.10a: for each 1°C increase in temperature, final cell volume decreased by roughly 2.5%.

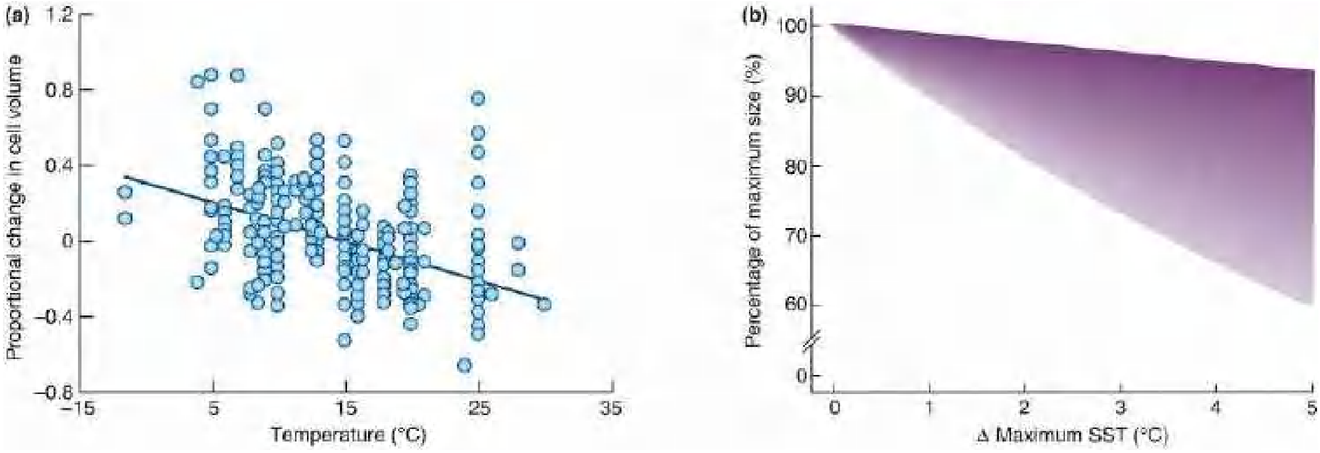


Figure 2.10 The temperature–size rule (final size decreases with increasing temperature). (a) Reduction in protist cell volume (65 datasets combined) with increasing rearing temperature. The vertical scale measures the proportional change from cell volume at 15°C. (b) Maximal size of 74 fish species generally declines across a steep temperature gradient in the Mediterranean Sea. The figure presents the best model based on the whole dataset, plotting percent of maximum size observed with an increase in maximum sea surface temperature (SST) of up to 5°C. The species vary in their activity levels: lighter shading corresponds to more active pelagic species and darker shading to more sedentary benthic species. The active species show a steeper decline in maximum size with increasing temperature.

Source: (a) After Atkinson *et al.* (2003). (b) After van Rijn *et al.* (2017).

These links can equally be seen from the viewpoint of how metabolism co-varies with temperature and size. We pick this up again in Section 3.9 when we look at the so-called metabolic theory of ecology.

APPLICATION 2.3 Getting predictions right in the face of climate change

The effects of temperature on growth, development and size may be of practical rather than simply scientific importance. Increasingly, ecologists are called upon to predict. We may wish to know what the consequences would be, say, of a 2°C rise in temperature resulting from global warming (Section 2.9.2), or to understand the role of temperature in seasonal, interannual and geographic variations in the productivity of, for example, marine ecosystems. We cannot afford to assume exponential relationships with temperature if they are really linear, nor to ignore the effects of changes in organism size on their role in ecological communities. Figure 2.10b shows for 74 fish species how maximum size varies across a steep sea surface temperature gradient in the Mediterranean Sea. If the reason for this pattern is the temperature–size rule (rather than genetic differences between locations) there could be important implications for fishery yields in a warmer climate. Van Rijn *et al.* (2017) suggest that the most pronounced size reductions will occur in large, active, non-migratory species that are often the major source of economic revenue, while elevated temperatures may have smaller effects on benthic, less active, and often less valuable, species. To optimise their catch, fishers may have to adapt their fishing strategies.

2.3.3 Ectotherms and endotherms

Many organisms have a body temperature that differs little, if at all, from their environment. A parasitic worm in the gut of a mammal, a fungal mycelium in the soil and a sponge in the sea acquire the temperature of the medium in which they live. Terrestrial organisms, exposed to the sun and the air, are different because they may acquire heat directly by absorbing solar radiation or be cooled by the latent heat of evaporation of water (typical pathways of heat exchange are shown in Figure 2.11). Various fixed properties may ensure that body temperatures are higher (or lower) than the ambient temperatures. For example, the reflective, shiny or silvery leaves of many desert plants reflect radiation that might otherwise heat the leaves. Organisms that can move have further control over their body temperature because they can seek out warmer or cooler environments, as when a lizard chooses to warm itself by basking on a hot sunlit rock or escapes from the heat by finding shade.

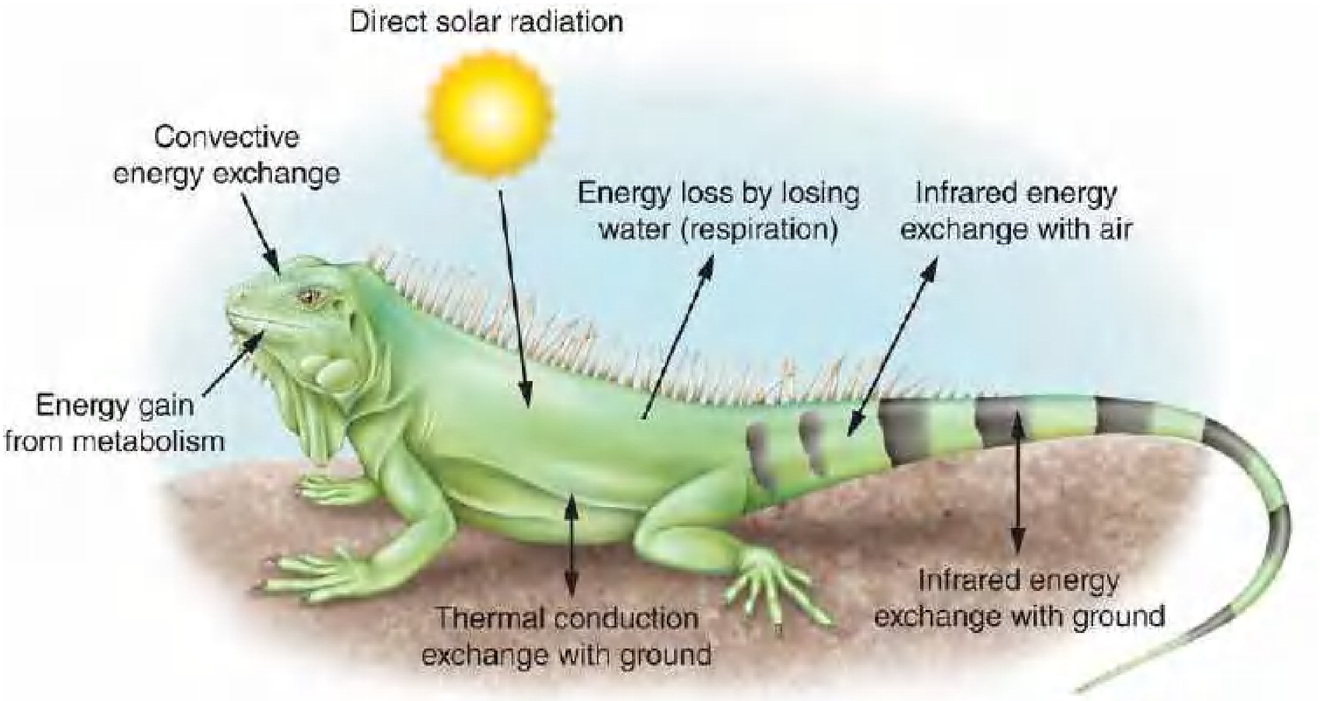


Figure 2.11 The avenues of heat exchange between an ectotherm and its environment.

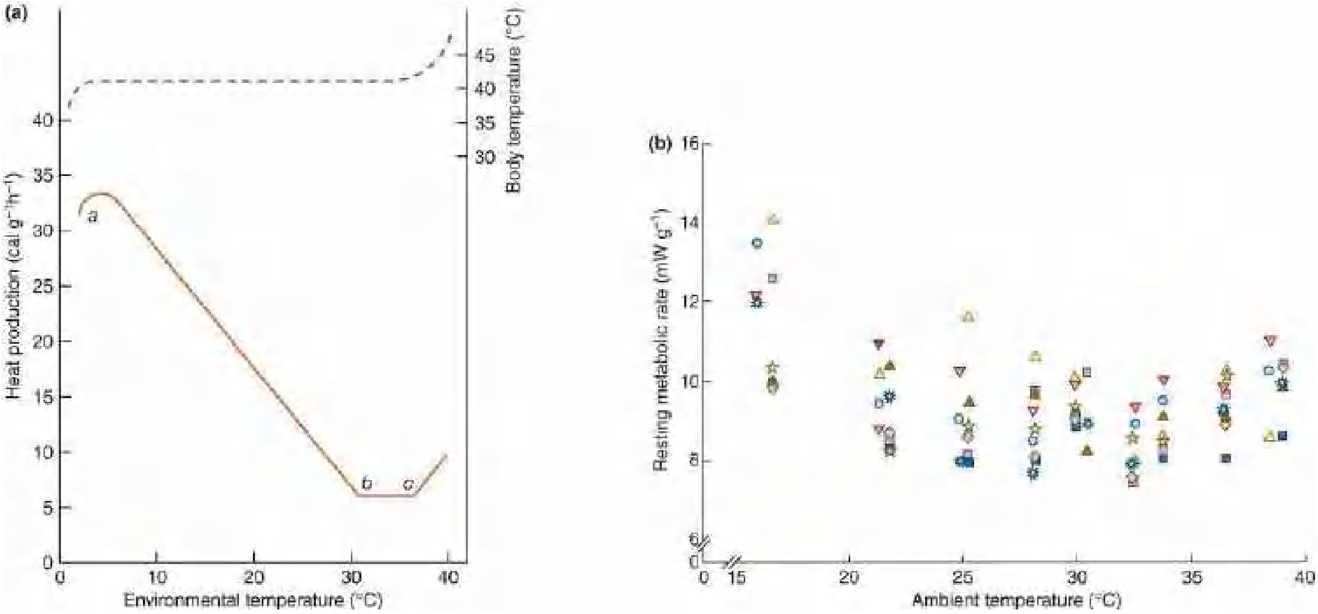
Source: After Fei *et al.* (2012).

Amongst insects there are examples of body temperatures raised by controlled muscular work, as when bumblebees raise their body temperature by shivering their flight muscles. Social insects such as bees and termites may combine to control the temperature of their colonies and regulate them with remarkable thermostatic precision. Even some plants (e.g. *Philodendron*) use metabolic heat to maintain a relatively constant temperature in their flowers; and, of course, birds and mammals use metabolic heat almost all of the time to maintain an almost perfectly constant body temperature.

An important distinction, therefore, is between *endotherms* that regulate their temperature by the production of heat within their own bodies, and *ectotherms* that rely on external sources of heat. But this distinction is not entirely clear-cut. As we have noted, apart from birds and mammals, there are also other taxa that use heat generated in their own bodies to regulate body temperature, but only for limited periods; and there are some birds and mammals that relax or suspend their endothermic abilities at the most extreme temperatures. In particular, many endothermic animals escape from some of the costs of endothermy by hibernating during the coldest seasons: at these times they behave almost like ectotherms.

endotherms: temperature regulation – but at a cost

Birds and mammals usually maintain a constant body temperature between 35 and 42°C, and they therefore tend to lose heat in most environments; but this loss is moderated by insulation in the form of fur, feathers and fat, and by controlling blood flow near the skin surface. When it is necessary to increase the rate of heat loss, this too can be achieved by the control of surface blood flow and by a number of other mechanisms shared with ectotherms like panting and the simple choice of an appropriate habitat. Together, all these mechanisms and properties give endotherms a powerful (but not perfect) capability for regulating their body temperature, and the benefit they obtain from this is a constancy of near-optimal performance. But the price they pay is a large expenditure of energy (Figure 2.12), and thus a correspondingly large requirement for food to provide that energy. Over a certain temperature range (the *thermoneutral zone*) an endotherm consumes energy at a basal rate. But at environmental temperatures further and further above or below that zone, the endotherm consumes more and more energy in maintaining a constant body temperature. Even in the thermoneutral zone, though, an endotherm typically consumes energy many times more rapidly than an ectotherm of comparable size.



[Figure 2.12](#) Examples of the thermoneutral zone. (a) Thermostatic heat production by an endotherm is constant in the thermoneutral zone, between *b*, the lower critical temperature, and *c*, the upper critical temperature. Heat production rises, but body temperature remains constant, as environmental temperature declines below *b*, until heat production reaches a maximum possible rate at a low environmental temperature. Below *a*, heat production and body temperature both fall. Above *c*, metabolic rate, heat production and body temperature all rise. Hence, body temperature is constant at environmental temperatures between *a* and *c*. (b) Mean resting metabolic rate (measured in units of power) versus ambient temperature in nine Japanese quail, *Coturnix japonica* (each bird has a different symbol). The thermoneutral zone extends between 23.2 and 36.0°C and the birds' minimum body temperature within this zone was 40.7°C.

Source: (a) After Hainsworth (1981). (b) After Ben-Hamo *et al.* (2010).

ectotherms and endotherms coexist: both strategies 'work'

The responses of endotherms and ectotherms to changing temperatures, then, are not so different as they may at first appear to be. Both are at risk of being killed by even short exposures to very low temperatures and by more prolonged exposure to moderately low temperatures. Both have an optimal environmental temperature and upper and lower lethal limits. There are also costs to both when they live at temperatures that are not optimal. For the ectotherm these may be slower growth and reproduction, slow movement, failure to escape predators and a sluggish rate of search for food. But for the endotherm, the maintenance of body temperature costs energy that might have been used to catch more prey, produce and nurture more offspring or escape more predators. There are also costs of insulation (e.g. blubber in whales, fur in mammals) and even costs of changing the insulation between seasons. Temperatures only a few degrees higher than the metabolic optimum are liable to be lethal to endotherms as well as ectotherms ([Section 2.3.6](#)).

It is tempting to think of ectotherms as 'primitive' and endotherms as having gained 'advanced' control over their environment, but it is difficult to justify this view. Most environments on earth are inhabited by mixed communities of endothermic and ectothermic animals. This includes some of the hottest – e.g. desert rodents and lizards – and some of the coldest – penguins and whales together with fish and krill at the edge of the Antarctic ice sheet. Rather, the contrast is between the high cost–high benefit strategy of endotherms and the low cost–low benefit strategy of ectotherms. But their coexistence tells us that both strategies, in their own ways, can ‘work’.

2.3.4 Life at low temperatures

The greater part of our planet is below 5°C. More than 70% of the planet is covered with seawater: mostly deep ocean with a remarkably constant temperature of about 2°C. If we include the polar ice caps, more than 80% of earth’s biosphere is permanently cold.

chilling injury

By definition, all temperatures below the optimum have adverse effects, but there is usually a wide range of such temperatures that cause no physical damage and over which any effects are fully reversible. There are, however, two quite distinct types of damage at low temperatures that can be lethal, either to tissues or to whole organisms: chilling and freezing. Many organisms, particularly tropical and subtropical plants, are damaged by exposure to temperatures that are low but above freezing point – so-called ‘chilling injury’. The fruits of the banana blacken and rot after exposure to chilling temperatures and many rainforest species are sensitive to chilling. Many insects also succumb to the effects of chilling at temperatures well above the freezing point of their body fluids. In these chill-sensitive insects, including the majority of temperate, subtropical and tropical species, chilling causes a loss of homeostasis that leads to paralysis, injury and ultimately death (Bale, 2002).

temperatures below 0°C

Temperatures below 0°C can have lethal physical and chemical consequences even though ice may not be formed. Water may ‘supercool’ to temperatures at least as low as −40°C, remaining in an unstable liquid form in which its physical properties change in ways that are certain to be biologically significant: its viscosity increases, its diffusion rate decreases and its degree of ionisation decreases. In fact, ice seldom forms in an organism until the temperature has fallen several degrees below 0°C. Body fluids remain in a supercooled state until ice forms suddenly around particles that act as nuclei. The concentration of solutes in the remaining liquid phase rises as a consequence. It is rare for ice to form within cells and it is then inevitably lethal, but the freezing of extracellular water is one of the factors that prevents ice forming within the cells themselves (Wharton, 2002), since water is withdrawn from the cell, and solutes in the cytoplasm (and vacuoles) become more concentrated. The effects of freezing are therefore mainly osmoregulatory: the water balance of the cells is upset and cell membranes are destabilised. The effects are essentially similar to those of drought and salinity.

frost resistance of plants and their parts

Plants inhabiting high latitudes and altitudes are particularly exposed to frost damage. Woody alpine plants, for example, can experience frost damage at any time of year (Neuner, 2014). In summer, the most frost-susceptible organs are reproductive shoots (−4.6°C), followed by immature leaves (−5.0°C), fully expanded leaves (−6.6°C), vegetative buds (−7.3°C) and xylem tissue (−10.8°C). These levels of resistance can be insufficient to survive frost events in summer and it may be that the most frost-susceptible parts define the upper limit of elevation for the distribution of such woody species. The same plant tissues are much less sensitive to frost damage in winter, the reproductive buds being most susceptible (−23.4°C), but with greater levels of freezing resistance in vegetative buds (−30 to −50°C), leaves (−25 to −58.5°C) and stems (−30 to

−70°C). Mechanisms of frost resistance include the production of cryoprotectant proteins, which appear to protect membranes and proteins against the severe dehydration stress associated with freezing, and the rapid accumulation of soluble carbohydrates, including sucrose, that serve to reduce cellular dehydration during freezing (Wisniewski *et al.*, 2014).

metabolic strategies of cold hardiness

Insects, and other taxa, have two main metabolic strategies that allow survival through the low temperatures of winter. A ‘freeze-avoiding’ strategy uses the synthesis of antifreeze proteins and the accumulation of extremely high levels of carbohydrate cryoprotectants (most often glycerol) so that body fluids can supercool to temperatures well below those normally encountered. A contrasting ‘freeze-tolerant’ strategy involves the regulated freezing of up to about 65% of total body water in extracellular spaces. Ice formation outside the cells is often triggered by the action of specific ice-nucleating agents or proteins, whereas low molecular weight cryoprotectants are used to maintain a liquid intracellular space and to protect membrane structure. Storey and Storey (2012) list two further strategies discovered more recently. The first, described for soil-dwelling invertebrates and particularly polar species, is ‘cryoprotective dehydration’, which combines extreme dehydration, in which virtually all freezable water is lost, together with high cryoprotectant levels that stabilise macromolecules. And more recently still, Sformo *et al.* (2010), in a study of the Alaskan bark beetle (*Cucujus clavipes*), showed that larvae at lower temperatures did not freeze but transitioned into a vitrified state (a non-crystalline amorphous solid) in which they could survive down to −100°C. The ‘vitrification’ strategy is accompanied by extensive dehydration and accumulation of antifreeze proteins and high concentrations of polyols.

acclimation and acclimatisation

The tolerances of organisms to low temperatures are not fixed but are preconditioned by the experience of temperatures in their recent past. This process is called *acclimation* when it occurs in the laboratory and *acclimatisation* when it occurs naturally. Acclimatisation may start as the weather becomes colder in the autumn, stimulating the conversion of almost the entire glycogen reserve of animals into polyols such as glycerol (Figure 2.13), but this can be an energetically costly affair: about 16% of the carbohydrate reserve may be consumed in the conversion of the glycogen reserves.

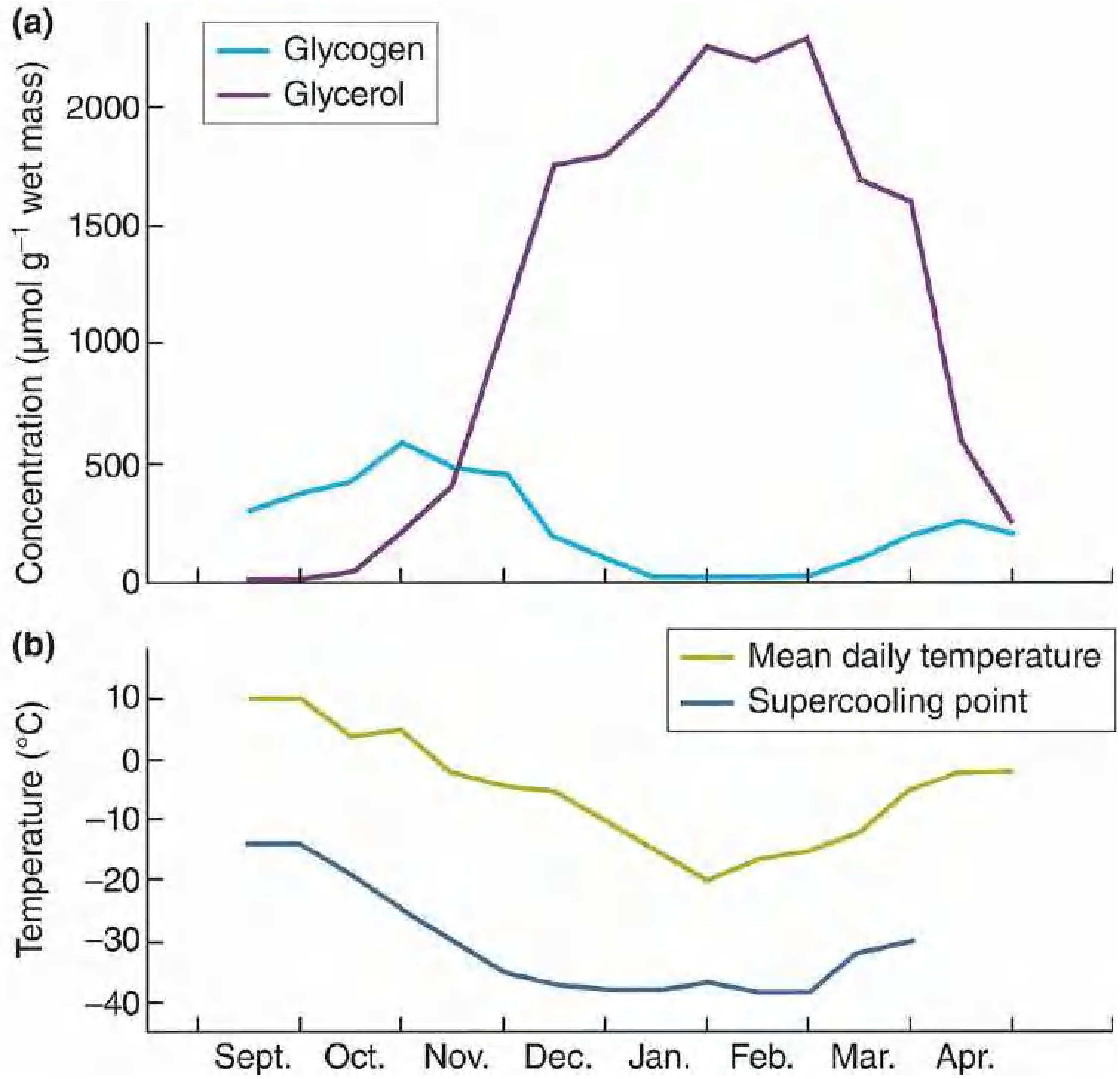


Figure 2.13 Acclimatisation involves conversion of glycogen to glycerol in a caterpillar. (a) As a result of cooling autumn temperatures, larvae of the goldenrod gall moth, *Epiblema scudderiana*, convert their stores of glycogen to glycerol, which eventually constitutes over 19% of the caterpillar's body mass. (b) The high levels of glycerol plus antifreeze proteins suppress the larval supercooling point from −14°C in late summer to −38°C by mid-winter, values well below environmental temperature extremes.

Source: (a, b) From Storey & Storey (2012), after Rickards *et al.* (1987).

Acclimatisation aside, individuals commonly vary in their temperature response depending on the stage of development they have reached. Probably the most extreme form of this is when an organism has a dormant stage in its life cycle. Dormant stages are typically dehydrated, metabolically slow and tolerant of extremes of temperature.

2.3.5 The genetics of cold tolerance

Cold tolerance and acclimatisation have long been recognised as being controlled by many genes, and the advent of technology that allows us to identify these and recognise their function has led to significant advances in the understanding of cold hardiness. In plants, for example, perception of cold temperatures seems to occur at the plasma membrane and is associated with an increase in calcium concentration that sets in train the activation of a wide variety of genes responsible for the biochemical changes already described (Wisniewski *et al.*, 2014). And in insects, Zhang *et al.* (2011) identified cold-responsive genes in the fruit-fly *Drosophila melanogaster* associated with muscle structure and function, immune and stress responses and carbohydrate metabolism.

In a laboratory 'selection' experiment involving plants of alfalfa (*Medicago sativa*, an important animal forage species), Castonguay *et al.* (2011) investigated whether the plant could be selected for improved freezing tolerance. Five weeks after sowing 1500 genotypes of a particular alfalfa cultivar used in eastern Canada, the plants were moved to low-temperature chambers for two week's acclimation at 2°C before being transferred to −2°C for an additional fortnight to simulate 'hardening' conditions in frozen soil. Subsequently, temperature was progressively dropped to the expected lethal temperature for 50% of the plants (their LT_{50} – lethal temperature for 50% of the plants in January), using a stepwise decline of temperature. After five weeks of regrowth at 20°C, genotypes that survived the original freezing cycle were intercrossed and subject to another cycle, and so on for six cycles of recurrent selection. The experiment was repeated with a second cultivar for four cycles of selection. Figure 2.14a shows that for both cultivars, several cycles of selection for freezing tolerance led to a significant decline in LT_{50} between the first cycle and later cycles of selection: in other words, individuals in the populations subject to selection were able to tolerate lower winter temperatures. Associated biochemical (Figure 2.14b, c) and genetic patterns (Figure 2.14d) provide good evidence that recurrent selection for superior freezing tolerance in alfalfa induces marked changes in influential traits. And if deliberate selection can change the tolerance of a domesticated plant we can certainly expect that natural selection has done the same thing for plants, animals and microorganisms in nature.

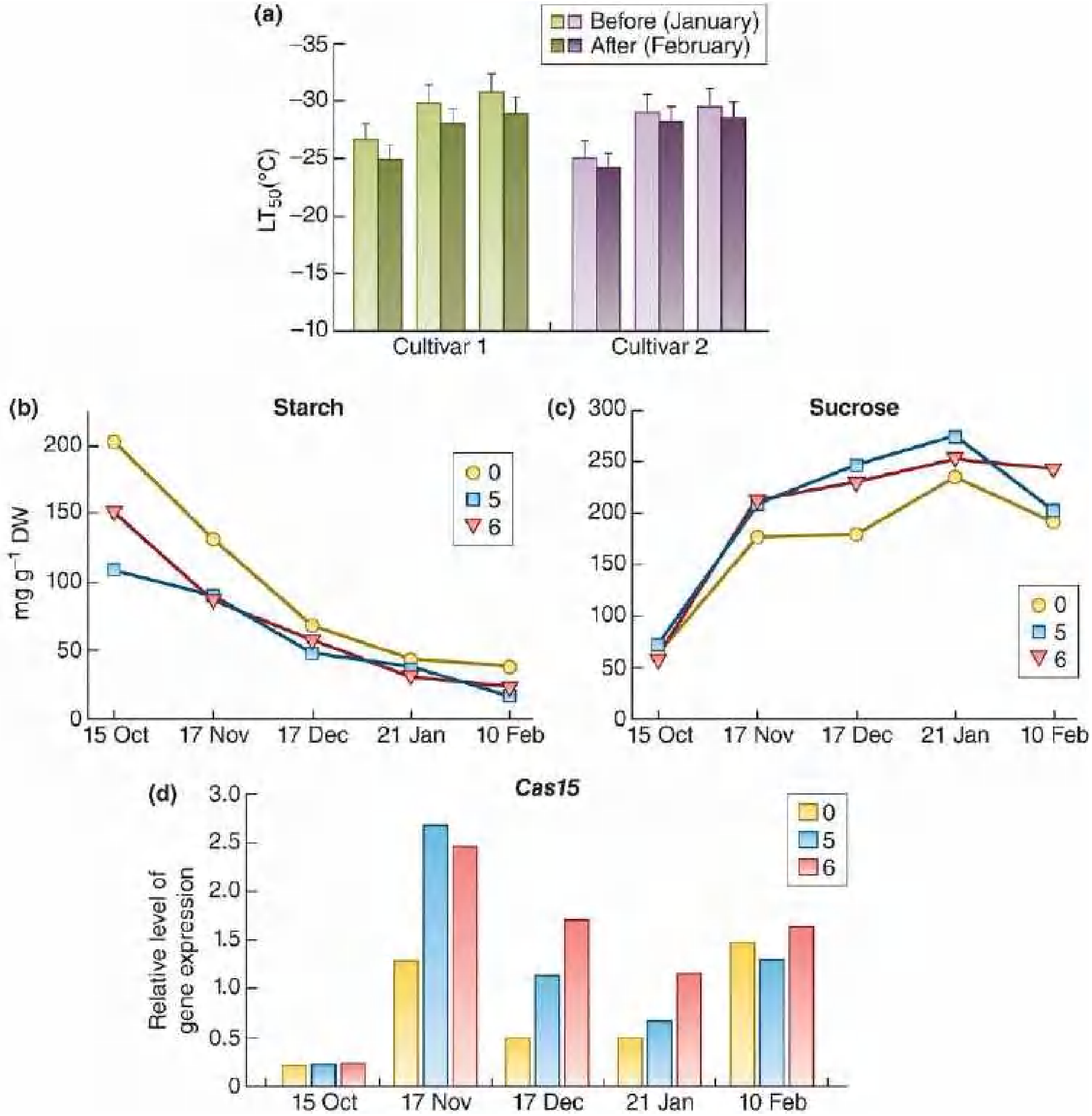


Figure 2.14 **Alfalfa can be selected for improved freezing tolerance.** (a) Tolerance to freezing (LT$_{50}$, 5% confidence levels shown) of populations of two cultivars of alfalfa used for animal forage in eastern Canada, before selection (TF0) or after several cycles of recurrent selection for freezing tolerance (three, four, five or six cycles). (b, c) Starch and sucrose concentrations in crowns of alfalfa plants during autumn and winter (cultivar 1) before (0) and after five or six cycles of selection. (d) Relative expression of the cold-induced gene *cas15* before (0) and after five or six cycles of selection.

Source: From Castonguay *et al.* (2011).

APPLICATION 2.4 Selection for cold tolerance in crops to increase their productivity and geographic range

There have been many striking cases where the geographic range of a crop species has been extended into colder regions of the world by plant breeders. Traditional crop breeding practices have generally used crossing of closely related varieties to produce new crops with desired cold-tolerance traits.

A key challenge for plant breeders is to introgress desirable traits from wild and even quite distantly related species into important domesticated crops but at the same time retain the favourable traits of the crop. Sugar cane (*Saccharum* spp.) is a major crop whose tropical heritage makes it cold sensitive and generally restricted to latitudes between 30°N and 35°S. Another member of the Poaceae family of tall grasses, *Miscanthus* spp., on the other hand, is a temperate-adapted species with marked cold tolerance. Głowacka *et al*. (2016) have shown that the chilling tolerance of *Miscanthus* can be transferred to sugarcane (Figure 2.15) without significant loss of overall sugarcane productivity. The chilling-tolerant hybrid of sugarcane and *Miscanthus* (Miscane US87-1019) has immediate potential for increased stock food and biofuel production, and at the same time provides the basis for extending sugarcane's range as a crop into higher latitudes and altitudes, once we better understand the genes that confer the cold-tolerance advantage.

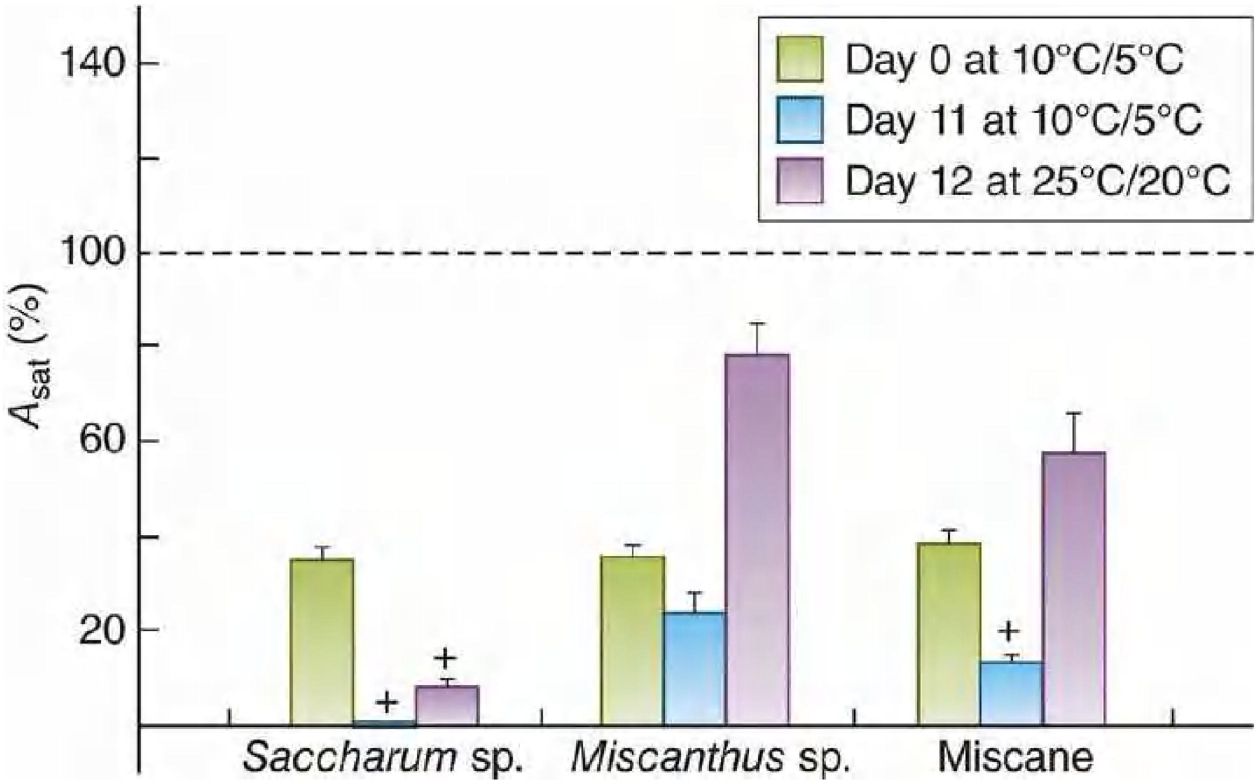


Figure 2.15 The chilling tolerance of *Miscanthus* can be transferred to *Saccharum*. Comparison of cold tolerance in a laboratory experiment involving plants of sugarcane (*Saccharum* sp. L79-1002), *Miscanthus* (Mxg 'Illinois') and a hybrid of *Saccharum* and *Miscanthus*, referred to as 'Miscane' (US87-1019). The light-saturated leaf net CO_2 uptake rate (A_{sat} in μmol m^{-2} s^{-1}) is shown for warm conditions before chilling treatment (25°C day, 20°C night: dashed line), after transfer of plants to chilling (day 0: 10°C day, 5°C night), on day 11 of chilling treatment and one day after transfer of plants back to warm conditions (day 12: recovery) expressed as a percentage of rates observed in warm conditions before chilling (control). A plus sign indicates a significantly lower value than the control. As expected, *Miscanthus* was the most cold tolerant, sugarcane the most cold sensitive, while the hybrid did not differ significantly from *Miscanthus* after recovery.

Source: From Głowacka *et al.* (2016).

potential of genomics, transcriptomics and proteomics in crop breeding

Future crop improvements to increase production and range in colder environments are certain to involve identification of the genes responsible for cold tolerance (both chilling and sub-zero tolerance) and acclimatisation. Erath *et al.* (2017) have, for example, identified genomic regions involved in frost tolerance of winter rye (*Secale cereale*) by mapping of quantitative trait loci (QTLs). A QTL is a section of DNA that correlates with variation in the quantitative trait of the phenotype (cold tolerance in this case); the QTL can be expected to contain the genes that control the trait. In winter rye, a QTL on chromosome 5R harbours the Frost resistance locus 2 (Fr-R2) and the 'Puma' allele at this locus was found to significantly increase frost tolerance. Discoveries of this kind can be expected to increase selection intensity for frost tolerance by preselecting plant breeding lines based on markers from the Fr-R2 locus.

2.3.6 Life at high temperatures

Perhaps the most important thing about dangerously high temperatures is that, for a given organism, they usually lie only a few degrees above the metabolic optimum. This is largely an unavoidable consequence of the physicochemical properties of most enzymes (Wharton, 2002). High temperatures may be dangerous because they lead to the inactivation or even the denaturation of enzymes, but they may also have damaging indirect effects by leading to dehydration.

high temperature and water loss in terrestrial environments

All terrestrial organisms need to conserve water, and at high temperatures the rate of water loss by evaporation can be lethal, but they are caught between two stools because evaporation is an important means of reducing body temperature. If surfaces are protected from evaporation (e.g. by closing stomata in plants or spiracles in insects) the organisms may be killed by too high a body temperature, but if their surfaces are not protected they may die of desiccation.

Death Valley, California, in the summer, is probably the hottest place on earth in which higher plants make active growth. Air temperatures during the daytime may approach 50°C and soil surface temperatures may be very much higher. The perennial plant, desert honeysweet (*Tidestromia oblongifolia*), grows vigorously in such an environment despite the fact that its leaves are killed if they reach the same temperature as the air. Very rapid transpiration keeps the temperature of the leaves at 40–45°C, and in this range they are capable of extremely rapid photosynthesis (Berry & Björkman, 1980).

Most of the plant species that live in very hot environments suffer severe shortage of water and are therefore unable to use the latent heat of evaporation of water to keep leaf temperatures down. This is especially the case in desert succulents in which water loss is minimised by a low surface to volume ratio and a low frequency of stomata. In such plants the risk of overheating or of damage to photosynthetic machinery may be reduced by spines (which shade the surface of a cactus) (Loik, 2008) or hairs or waxes (which reflect a high proportion of the incident radiation). Nevertheless, such species experience and tolerate temperatures in their tissues of more than 60°C when the air temperature is above 40°C.

fire

Fires are responsible for the highest temperatures that terrestrial organisms face on earth and, before the fire-raising activities of humans, were caused mainly by lightning strikes. The recurrent risk of fire has shaped the species composition of arid and semiarid woodlands in many parts of the world. All plants are damaged by burning but it is the remarkable powers of regrowth from protected meristems on shoots and seeds that allow a specialised subset of species to recover from damage and form characteristic fire floras (see, for example, Rundel *et al.*, 2016).

Decomposing organic matter in heaps of farmyard manure, compost heaps and damp hay may reach very high temperatures. Stacks of damp hay are heated to temperatures of 50–60°C by the metabolism of fungi such as *Aspergillus fumigatus*, carried further to approximately 65°C by other thermophilic fungi such as *Mucor pusillus* and then a little further by bacteria and archaea. Biological activity stops well short of 100°C but autocombustible products are formed that cause further heating, drive off water and may even result in fire.

high temperature and oxygen supply in aquatic environments

In aquatic environments there is a situation analogous to the interplay between temperature and water supply discussed earlier for terrestrial environments. But in the aquatic case the interplay is between temperature and oxygen supply. At high temperatures, oxygen supply may not be able to keep up with the organism's metabolic demand for oxygen, such that upper thermal limits occur at temperatures lower that those that denature proteins (the oxygen and capacity-limited thermal tolerance theory of Pörtner (2001)).

thermal vents

An ecologically very remarkable hot aquatic environment was first described only towards the end of the last century. In 1979, a deep oceanic site was discovered in the eastern Pacific at which fluids at high temperatures ('smokers') were vented from the sea floor forming thin-walled 'chimneys' of mineral materials. Since that time many more vent sites have been discovered at mid-ocean crests in both the Atlantic and Pacific Oceans. They lie 2000–4000 m below sea level at pressures of 200–400 bars (20–40 MPa). The boiling point of water is raised to 370°C at 200 bars and to 404°C at 400 bars. The superheated fluid emerges from the chimneys at temperatures as high as 350°C, and as it cools to the temperature of seawater at about 2°C it provides a continuum of environments at intermediate temperatures.

Environments at such extreme pressures and temperatures are obviously extraordinarily difficult to study *in situ* and in most respects impossible to maintain in the laboratory. Some thermophilic bacteria collected from vents, such as *Pyrococcus furiosus*, exhibit optimal growth at 100°C (Zeldes *et al.*, 2017), but there is circumstantial evidence that some microbial activity occurs at even higher temperatures and may form the energy resource for the warm water communities outside the vents.

There is a rich eukaryotic fauna in the local neighbourhood of vents that is quite atypical of the deep oceans in general. At one vent in Middle Valley, north-east Pacific, surveyed photographically and by video, at least 55 taxa were documented of which 15 were new or probably new species (Juniper *et al.*, 1992). There can be few environments in which so complex and specialised a community depends on so localised a special condition. The closest known vents with similar conditions were 2500 km distant. Such communities add a further list to the planet's record of species richness, and more than 500 new animal species from the world's thermal vents have been described (Desbruyères *et al.*, 2006).

2.3.7 Temperature as a stimulus

We have seen that temperature as a condition affects the rate at which organisms develop. It may also act as a stimulus, determining whether or not the organism starts its development at all. For instance, for many species of temperate, arctic and alpine herbs, a period of chilling or freezing (or even of alternating high and low temperatures) is necessary before germination will occur. A cold experience (physiological evidence that winter has passed) is required before the plant can start on its cycle of growth and development. Temperature may also interact with other stimuli (e.g. photoperiod) to break dormancy and so time the onset of growth. The seeds of the birch (*Betula pubescens*) require a photoperiodic stimulus (i.e. experience of a particular regime of day length) before they will germinate, but if the seed has been chilled it starts growth without a light stimulus.

2.4 Correlations between temperature and the distribution of plants and animals

2.4.1 Spatial and temporal variations in temperature

Variations in temperature on and within the surface of the earth have a variety of causes: latitudinal, altitudinal, continental, seasonal, diurnal and microclimatic effects and, in soil and water, the effects of depth.

Latitudinal and seasonal variations cannot really be separated. The angle at which the earth is tilted relative to the sun changes with the seasons, and this drives some of the main temperature differentials on the earth's surface. Superimposed on these broad geographic trends are the influences of altitude and 'continentality'. There is a drop of 1°C for every 100 m increase in altitude in dry air, and a drop of 0.6°C in moist air. This is the result of the 'adiabatic' expansion of air as atmospheric pressure falls with increasing altitude. The effects of continentality are largely attributable to the different rates of heating and cooling of the land and the sea. The land surface reflects less heat than the water, so the surface warms more quickly, but it also loses heat more quickly. The sea therefore has a moderating, 'maritime' effect on the temperatures of coastal regions and especially islands; both daily and seasonal variations in temperature are far less marked than at more inland, continental locations at the same latitude. Moreover, there are comparable effects within landmasses: dry, bare areas like deserts suffer greater daily and seasonal extremes of temperature than do wetter areas like forests. Thus, global maps of temperature zones hide a great deal of local variation.

microclimatic variation

On a smaller scale still there can be a great deal of microclimatic variation. For example, the sinking of dense, cold air into the bottom of a valley at night can make it as much as 30°C colder than the side of the valley only 100 m higher; the winter sun, shining on a cold day, can heat the south-facing side of a tree (and the habitable cracks and crevices within it) to as high as 30°C; and the air temperature in a patch of vegetation can vary by 10°C over a vertical distance of 2.6 m from the soil surface to the top of the canopy. Hence, we need not confine our attention to global or geographic patterns when seeking evidence for the influence of temperature on the distribution and abundance of organisms.

ENSO and NAO

Long-term temporal variations in temperature, such as those associated with the ice ages, were discussed in the previous chapter (Section 1.4.3). Between these, however, and the very obvious daily and seasonal changes that we are all aware of, a number of medium-term patterns have become increasingly apparent. Notable amongst these are the El Niño–Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO). The ENSO is an alternation between a warm (El Niño) and a cold (La Niña) state of the waters of the tropical Pacific Ocean off the coast of South America (Figure 2.16a), although it affects temperature and the climate generally in terrestrial and marine environments throughout the whole Pacific basin and beyond (Figure 2.16b). The NAO refers to a north–south alternation in atmospheric mass between the subtropical Atlantic and the Arctic (Figure 2.16c) and again affects climate in general rather than just temperature (Figure 2.16d). Positive index values (Figure 2.16c) are associated, for example, with relatively warm conditions in North America and Europe and relatively cool conditions in North Africa and the Middle East. An example of the effect of NAO variation on species abundance, that of cod, *Gadus morhua*, in the Barents Sea, is shown in Figure 2.17.

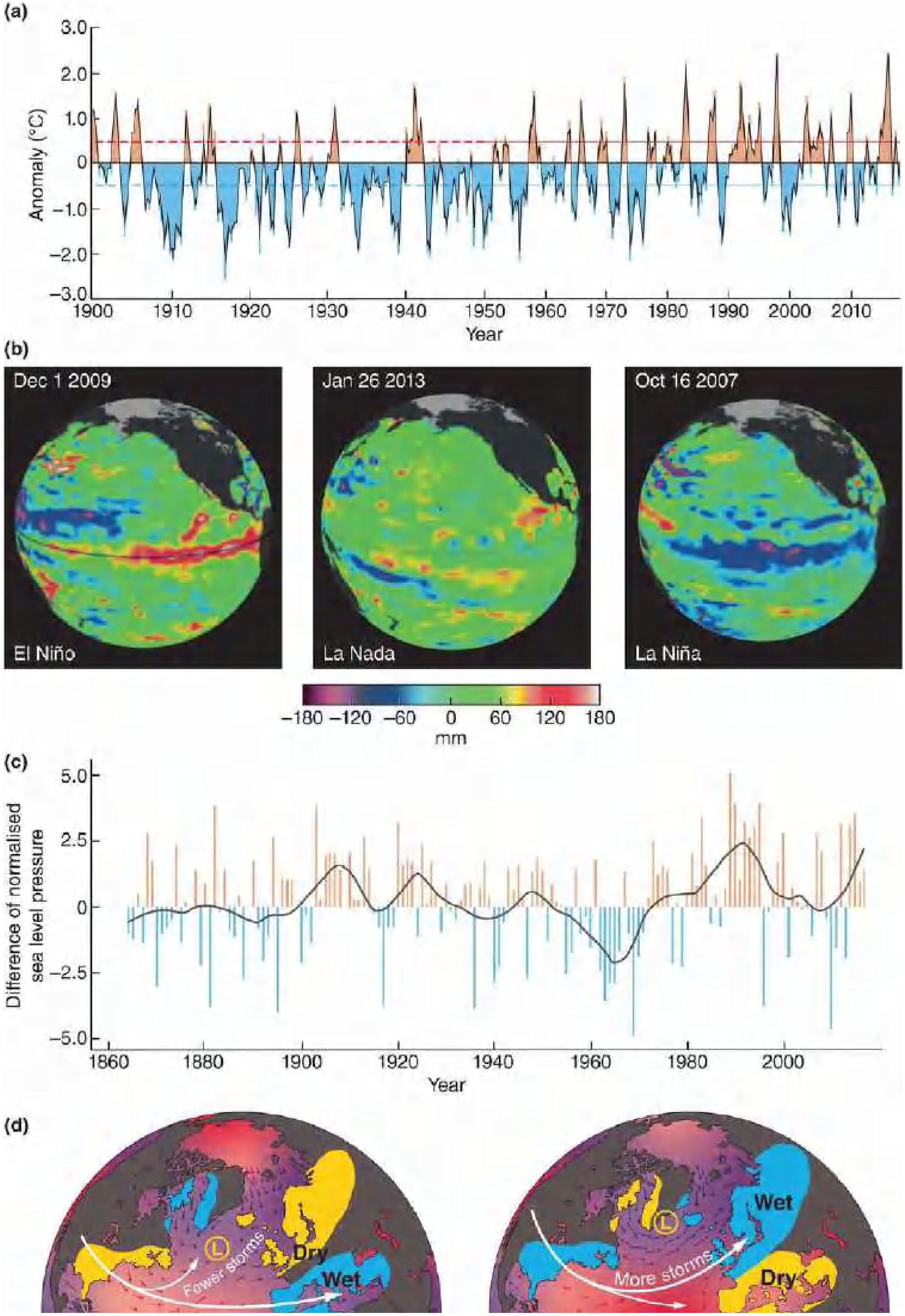
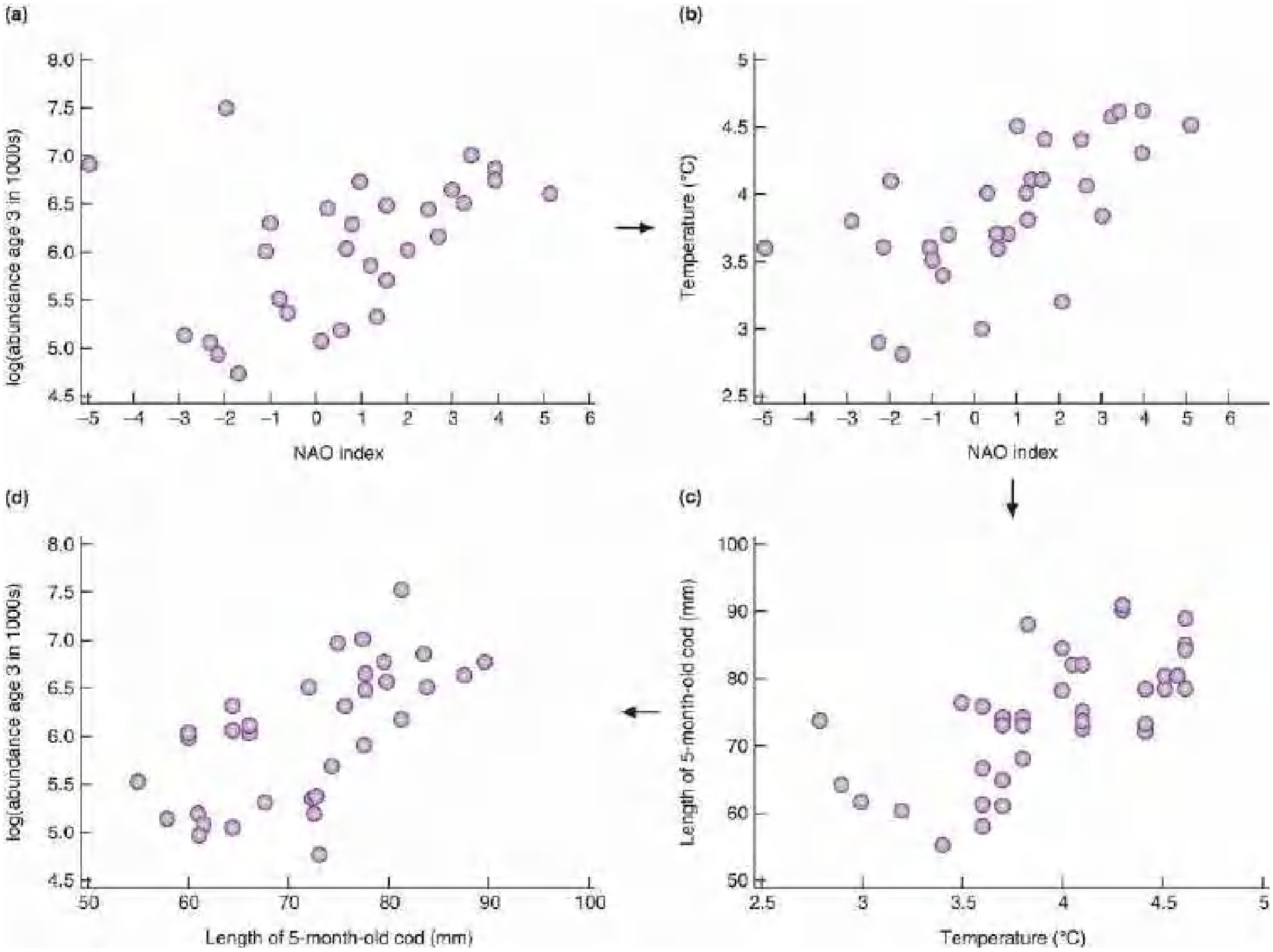

(a)
3.0
2.0
1.0
0
−1.0
−2.0
−3.0
Anomaly (°C)
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
Year
(b)
Dec 1 2009
El Niño
Jan 26 2013
La Nada
Oct 16 2007
La Niña
−180
−120
−60
0
60
120
180
mm
(c)
5.0
2.5
0
−2.5
−5.0
Difference of normalised sea level pressure
1860
1880
1900
1920
1940
1960
1980
2000
Year
(d)
L
Fewer storms
Dry
Wet
L
Wet
More storms
Dry

Figure 2.16 **Features of the El Niño–Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO).** (a) ENSO from 1900 to 2017 as measured by sea surface temperature (SST) anomalies (differences from the mean) in the equatorial mid-Pacific. El Niño events are defined as occurring when the SST is more than 0.4°C above the mean (red dashed line) and La Niña events when the SST is more than 0.4°C below the mean (blue dashed line). (b) Examples of El Niño (December 2009) and La Niña events (October 2007) as well as a neutral state (La Nada; January 2013) in terms of sea height above average levels. Warmer seas are higher; for example, a sea height 150–200 mm below average equates to a temperature anomaly of approximately 2–3°C. (c) NAO from 1864 to 2017 as measured by the normalised sea-level pressure difference between Lisbon in Portugal and Reykjavik in Iceland. (d) Typical winter conditions when the NAO index is positive or negative. Conditions that are more than usually warm, cold, dry or wet are indicated. The positions of the Icelandic low pressure (L) and the Azores high pressure (H) zones are shown.

Source: (a) Compiled from the US National Ocean and Atmospheric Administration (NOAA), https://www.ncdc.noaa.gov/teleconnections/enso/indicators/sst.php). (b) From the US National Aeronautics and Space Administration (NASA), https://sealevel.jpl.nasa.gov/science/elninopdo/elnino/). (c) From https://climatedataguide.ucar.edu/climate-data/hurrell-north-atlantic-oscillation-nao-index-station-based). (d) From http://www.ldeo.columbia.edu/NAO/.



[Figure 2.17](#) The abundance of three-year-old cod, *Gadus morhua*, in the Barents Sea is positively correlated with the value of the North Atlantic Oscillation (NAO) index. The mechanism underlying the correlation (a) is suggested in (b–d). (b) Annual mean temperature increases with the NAO index. (c) The length of five-month-old cod increases with annual mean temperature. (d) The abundance of cod at age three years increases with their length at five months.

Source: After Ottersen *et al.* (2001).

2.4.2 Typical temperatures and distributions

isotherms

There are very many examples of plant and animal distributions that are strikingly correlated with some aspect of environmental temperature (e.g. Figure 2.2a) and this kind of pattern may still hold even at gross taxonomic and systematic levels (Figure 2.18). At a finer scale, the distributions of many species closely match maps of some aspect of temperature. For example, the northern cool range boundary of wild madder plants (*Rubia peregrina*) is closely correlated with the position of the January 4.5°C isotherm (an isotherm is a line on a map joining places that experience the same temperature).

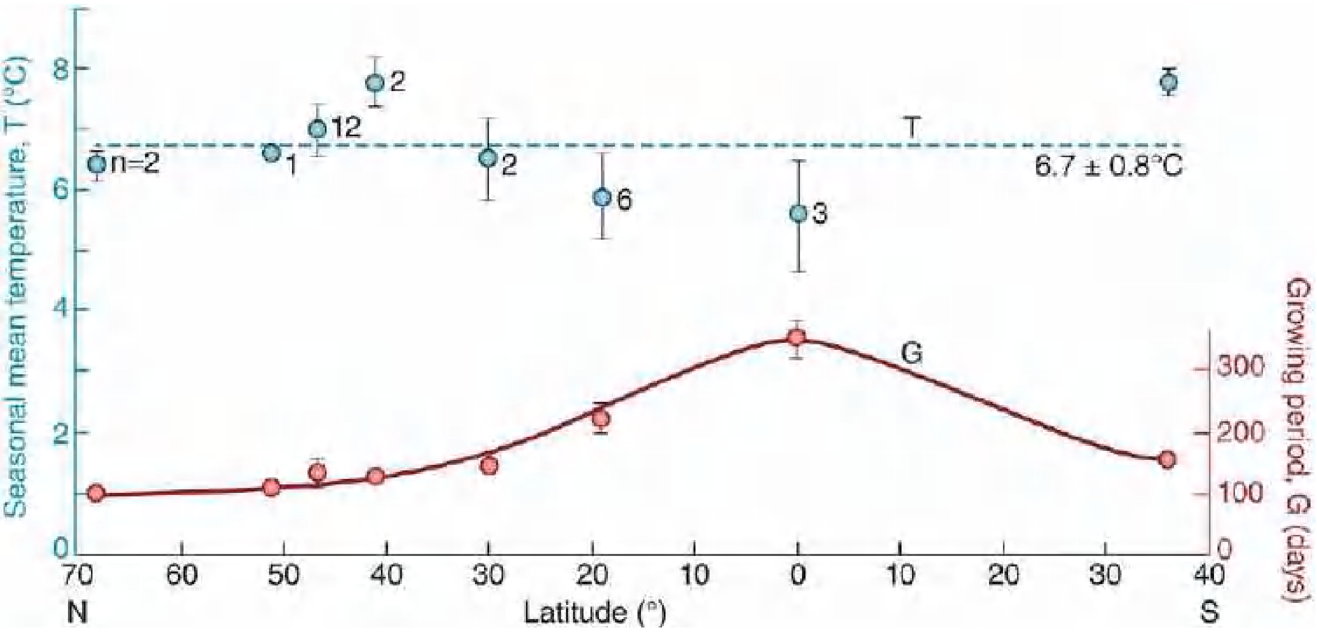


Figure 2.18 The treelines (high-altitude limits of forest cover) of the world's mountains seem to follow a common isotherm. This is 6.7 ± 0.8°C), with very similar mean ground temperatures during the growing season across a wide range of latitudes from subarctic through equatorial areas to temperate southern hemisphere regions (the growing period differs according to latitude). The species composition of the forests is, of course, vastly different in the different regions.

Source: From Körner & Paulsen (2004).

However, such relationships need to be interpreted with some caution: they can be extremely valuable in predicting where we might and might not find a particular species (e.g. Figure 2.5); they may suggest that some feature related to temperature is important in the life of the organisms; but they do not prove that temperature *causes* the limits to a species' distribution. For one thing, the temperatures measured for constructing isotherms for a map are only rarely those that the organisms experience. In nature an organism may choose to lie in the sun or hide in the shade and, even in a single day, may experience a baking midday sun and a freezing night. Moreover, temperature varies from place to place on a far finer scale than will usually concern a geographer, but it is the conditions in these 'microclimates' that will be crucial in determining what is habitable for a particular species. For example, the prostrate shrub *Dryas octopetala* is restricted to altitudes exceeding 650 m in north Wales, UK, where it is close to its southern limit. But to the north, in Sutherland in Scotland, where it is generally colder, it is found right down to sea level.

On the other hand, Payne *et al.* (2016) were able to demonstrate a strong correlation between the warm boundary isotherms of nine well-studied fish species and their optimum temperatures for activity, somatic growth and reproductive growth (Figure 2.19): this is good evidence of a causal link.

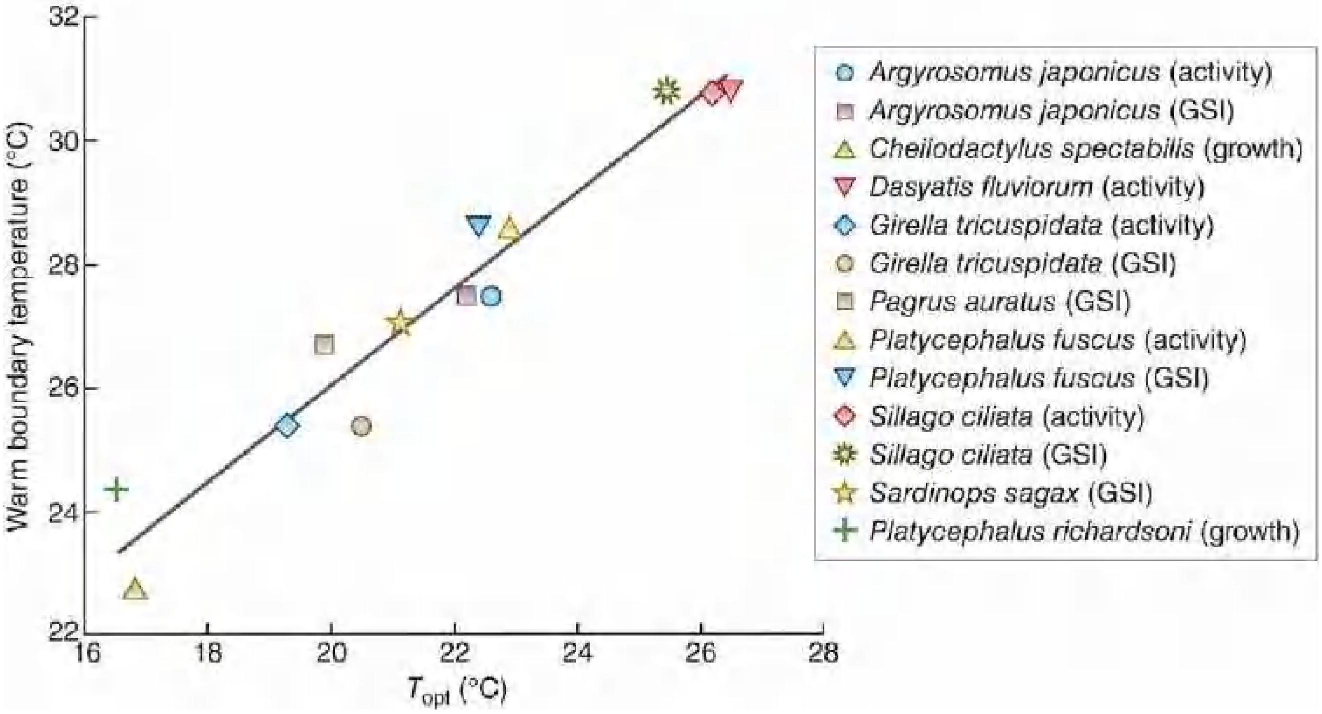


Figure 2.19 Warm boundary limits of nine Australian fish species are correlated with species-specific optimum fish performance. Optimum temperature (T_{opt}) is shown for maximum activity, somatic growth or reproductive growth (gonadosomatic index, GSI, a measure of gonad mass relative to total body mass) measured in the wild (four species provided both activity and reproductive growth data, giving 13 points in total). The species-specific warm equatorward range boundary is the average temperature of the warmest month at the range limit.

Source: From Payne *et al.* (2016).

2.4.3 Distributions and extreme conditions

For many species, distributions are accounted for not so much by average temperatures as by occasional extremes, especially occasional lethal temperatures that preclude its existence. For instance, injury by frost is probably the single most important factor limiting plant distribution. To take one example: the saguaro cactus (*Carnegiea gigantea*) is liable to be killed when temperatures remain below freezing for 36 h, but if there is a daily thaw it is under no threat. In Arizona, the northern and eastern edges of its distribution correspond to a line joining places where on occasional days it fails to thaw. Thus, the saguaro is absent where there are occasionally lethal conditions – an individual need only be killed once.

you only die once

Similarly, there is scarcely any crop that is grown on a large commercial scale in the climatic conditions of its wild ancestors, and it is well known that crop failures are often caused by extreme events, especially frosts and drought. For instance, the climatic limit to the geographic range for the production of coffee (*Coffea arabica* and *C. robusta*) is defined by the 13°C isotherm for the coldest month of the year. Much of the world's crop is produced in the highland microclimates of the São Paulo and Paraná districts of Brazil. Here, the average minimum temperature is 20°C, but occasionally cold winds and just a few hours of temperature close to freezing are sufficient to kill or severely damage the trees (and influence world coffee prices).

global variation in thermal tolerances

Species at higher latitudes experience greater seasonal temperature variation and are expected to be able to withstand greater temperature extremes. This general pattern has been reported for ectotherms (Figure 2.20a) (Sunday *et al.*, 2011) but does it also apply to endotherms, which maintain a high and constant temperature and are thus decoupled to a degree from the influence of ambient conditions? The *climate variability hypothesis* has been tested in a meta-analysis of physiological studies that allowed the thermoneutral zones (see Section 2.3.3) to be estimated for hundreds of bird and mammal species (Khaliq *et al.*, 2014). Figure 2.20b and c plots the thermoneutral zones estimated for individual bird and mammal species in relation to both latitude and climatic variability (the latter estimated as the annual range between the monthly average of daily maximum temperatures of the warmest month and minimum temperatures of the coldest month). Note how temperature variability is generally lower in tropical regions. The climate variability hypothesis was supported in the case of birds but not mammals, a difference that may be related to their different lifestyles, with mammals often able to create their own preferred microclimates in burrows and dens.

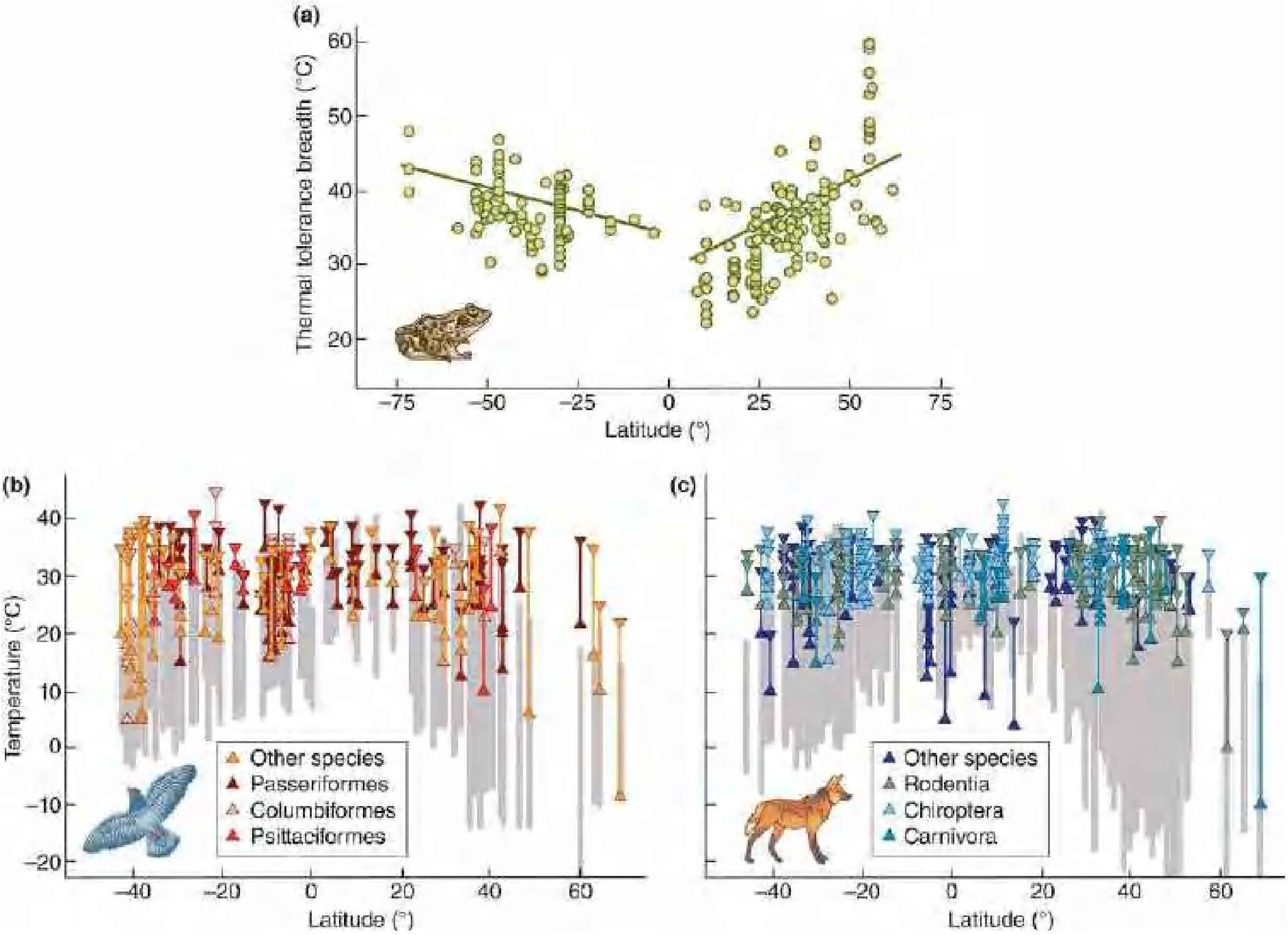


Figure 2.20 **Geographic variation in thermal tolerances.** (a) Terrestrial ectotherms, including arthropods, reptiles and amphibians, as estimated by the range between their upper and lower lethal or critical temperature limits, and (b) birds and (c) mammals as illustrated by the breadths of their thermoneutral zones, plotted in relation to latitude and, in the case of (b) and (c) climatic variability (grey bars) at the capture sites of the individuals used in published physiological experiments. Major bird and mammal orders are highlighted by different colours. Climatic variability is the annual range between maximum and minimum temperatures.

Source: (a) After Sunday *et al.* (2011). (b, c) After Khaliq *et al.* (2014).

APPLICATION 2.5 Tropical species at particular risk from climate change

Khaliq *et al*. (2014) found that most of the endotherm species in their dataset (Figure 2.20b, c) currently experience maximum ambient temperatures that are within their tolerance limits during most months of the year, and note that in a warming climate many should still be able to find suitable temperature conditions within their current range. However, vulnerability to higher future ambient temperatures increases from the poles towards tropical regions, even though increases of temperature predicted for temperate and polar regions exceed those in the tropics. This is because species in tropical regions tend to live closer to their upper temperature limits and even small increases in ambient temperatures may challenge their survival. Moreover, projections of declining precipitation in tropical areas worsen the prospects for tropical species, because water availability is crucial for endotherms to compensate thermal stress. These results highlight the threats from global climate change in tropical zones, which harbour the greatest amount of biodiversity worldwide.

2.4.4 Distributions and the interaction of temperature with other factors

Although organisms respond to each condition in their environment, the effects of conditions may be determined largely by the responses of other community members. Temperature does not act on just one species: it also acts on its competitors, prey, parasites and so on. This, as we saw in Section 2.2, was the difference between a fundamental niche (where an organism *could* live) and a realised niche (where it *actually* lives). For example, an organism will suffer if its food is another species that cannot tolerate an environmental condition. This is illustrated by the distribution of the rush moth (*Coleophora alticolella*) in England. The moth lays its eggs on the flowers of the rush *Juncus squarrosus* and the caterpillars feed on the developing seeds. Above 600 m, the moths and their caterpillars are little affected by the low temperatures, but the rush, although it grows, fails to ripen its seeds. This, in turn, limits the distribution of the moth, because caterpillars that hatch in the colder elevations will starve as a result of insufficient food (Randall, 1982).

disease

The effects of conditions on disease may also be important. Conditions may favour the spread of infection (winds carrying fungal spores), or favour the growth of the parasite, or weaken the defenses of the host. For example, during an epidemic of southern corn leaf blight (*Helminthosporium maydis*) in a corn field in Connecticut, the plants closest to the trees that were shaded for the longest periods were the most heavily diseased (Harper, 1955).

competition

Competition between species can also be profoundly influenced by environmental conditions, especially temperature. Two stream salmonid fishes, *Salvelinus malma* and *S. leucomaenis*, coexist at intermediate altitudes (and therefore intermediate temperatures) on Hokkaido Island, Japan, whereas only the former lives at higher altitudes (lower temperatures) and only the latter at lower altitudes. A reversal, by a change in temperature, of the outcome of competition between the species plays a key role in this pattern (discussed more fully in Section 8.2.3).

temperature and water availability

Many of the interactions between temperature and other physical conditions are so strong that it is not sensible to consider them separately. The relative humidity of the atmosphere, for example, is an important condition in the life of terrestrial organisms because it plays a major part in determining the rate at which they lose water. In practice, it is rarely possible to make a clean distinction between the effects of relative humidity and of temperature. This is simply because a rise in temperature leads to an increased rate of evaporation. A relative humidity that is acceptable to an organism at a low temperature may therefore be unacceptable at a higher temperature. Microclimatic variations in relative humidity can be even more marked than those involving temperature. For instance, it is not unusual for the relative humidity to be almost 100% at ground level amongst dense vegetation and within the soil, whilst the air immediately above, perhaps 40 cm away, has a relative humidity of only 50%. The organisms most obviously affected by humidity in their distribution are those ‘terrestrial’ animals that are actually, in terms of the way they control their water balance, ‘aquatic’. Amphibians, terrestrial isopods, nematodes, earthworms and molluscs are all, at least in their active stages, confined to microenvironments where the relative humidity is at or very close to 100%. The major group of animals to escape such confinement are the terrestrial arthropods, especially insects. Even here though, the evaporative loss of water often confines their activities to habitats (e.g. woodlands) or times of day (e.g. dusk) when relative humidity is relatively high.

APPLICATION 2.6 Farmers’ choice of cover crops in relation to temperature and soil water potential

Farmers have a wide choice of species that can be sown as cover crops during fallow periods to improve soil quality and reduce soil erosion and runoff. But which should they choose? Seed germination is a key stage in plant establishment, particularly when sowing occurs in summer, when temperatures are high and water availability low, and germination for 34 species of potential cover crops in four families was monitored in the laboratory at temperatures ranging from 4.5 to 43°C and at four water potentials (Figure 2.21). Optimal temperatures for germination of seeds varied from 21.3 to 37.2°C; maximum temperatures at which the species could germinate varied from 27.7 to 43.0°C; and base water potentials, the lowest water potential at which a seed can germinate, varied from –0.1 to –2.6 MPa. (Note that at a potential of 0 MPa soil is in a state of saturation, while at –1.5 MPa soil is at its permanent wilting point (see Section 3.3.2).)

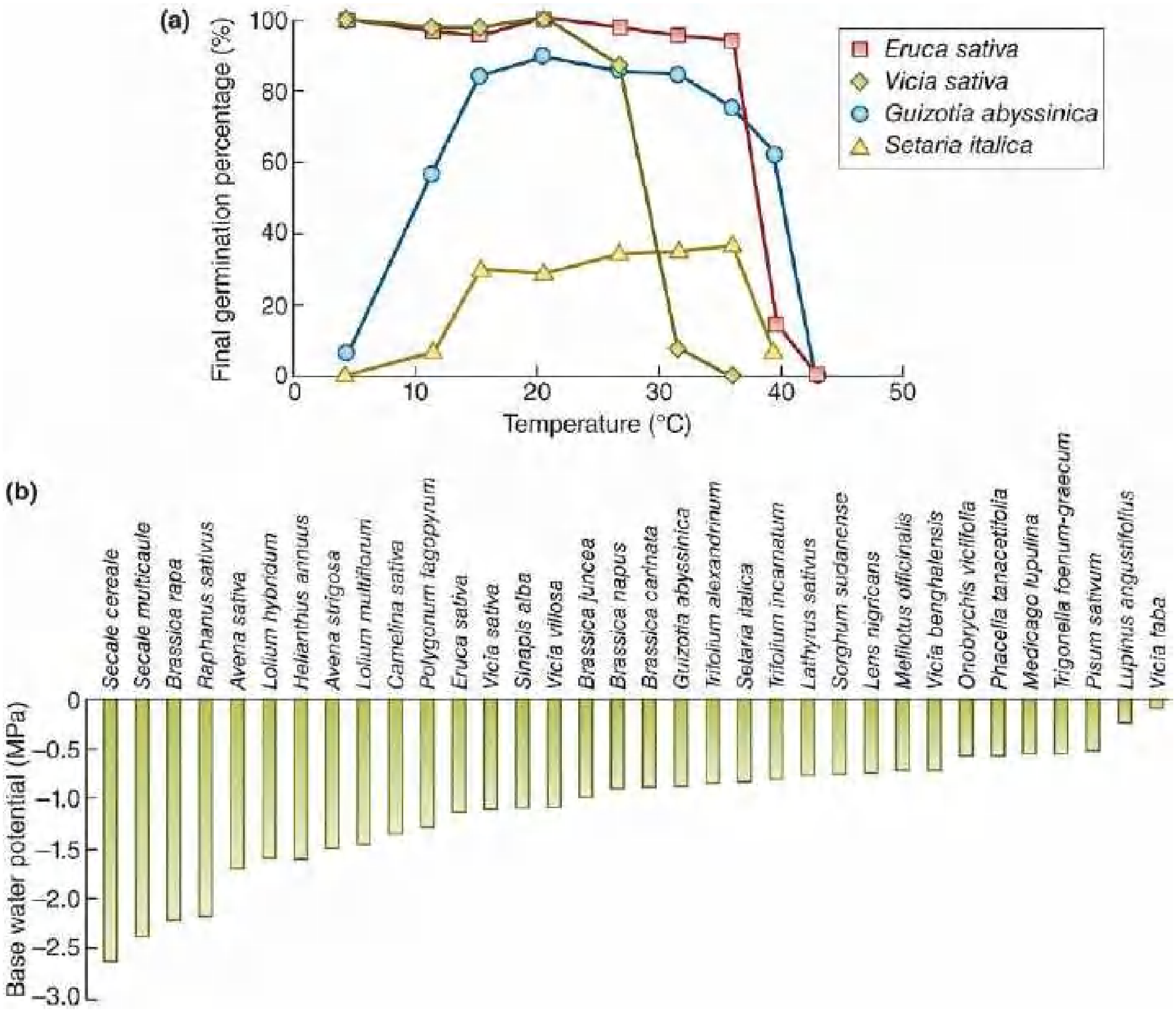


Figure 2.21 Niches of cover crops in terms of temperature and base water potential. (a) Response curves to temperature for selected species of cover crops in terms of percentage of seeds that germinate. (b) Niches in one dimension for various species of cover crop. Base water potential is the lowest water potential at which a seed can germinate. At one extreme, *Vicia faba* is very sensitive to water potential for germination while, at the other, *Secale cereale* can germinate at very low water potentials.

Source: From Tribouillois *et al.* (2016).

Most of the cover crops were adapted to summer sowing with a high mean optimal temperature for germination, but some, such as *Vicia sativa* (Figure 2.21a), were more sensitive to high temperatures. Others, such as *Secale cereale* (Figure 2.21b), were more resistant to water deficit and germinated even when water potential was very low. Tribouillois *et al.* (2016) classified the cover crops into functional groups that are of value to farmers when choosing species appropriate for their particular conditions. Thus, functional group 1, which includes *Guizotia abyssinica* and *Setaria italica*, has a minimal temperature of 10°C, a maximal temperature of 41.2°C and a base water potential of –0.9 KPa. Functional group 4, on the other hand, which includes *Brassica rapa* and *Secale cereale*, has a minimal temperature of 0.4°C, a maximal temperature of 38.6°C and a base water potential of –2.4 KPa.

2.5 pH of soil and water

The pH of soil in terrestrial environments or of water in aquatic ones is a condition that can exert a powerful influence on the distribution and abundance of organisms. The protoplasm of the root cells of most vascular plants is damaged as a direct result of toxic concentrations of H^+ or OH^- ions in soils below pH 3 or above pH 9, respectively. Further, indirect effects occur because soil pH influences the availability of nutrients and/or the concentration of toxins.

Increased acidity (low pH) may act in three ways: (i) directly, by upsetting osmoregulation, enzyme activity or gaseous exchange across respiratory surfaces; (ii) indirectly, by increasing the concentration of toxic heavy metals at higher pHs, particularly aluminium (Al^{3+}) but also manganese (Mn^{2+}) and iron (Fe^{3+}), which are essential plant nutrients; and (iii) indirectly, by reducing the quality and range of food sources available to animals. Tolerance limits for pH vary amongst plant species, but only a minority are able to grow and reproduce at a pH below about 4.5.

In alkaline soils, iron (Fe^{3+}) and phosphate (PO_4^{3+}), and certain trace elements such as manganese (Mn^{2+}), are fixed in relatively insoluble compounds, and plants may then suffer because there is too little rather than too much of them. For example, calcifuge plants (those characteristic of acid soils) commonly show symptoms of iron deficiency when they are transplanted to more alkaline soils. In general, however, soils and waters with a pH above 7 tend to be hospitable to many more species than those that are more acid. Chalk and limestone grasslands carry a much richer flora (and associated fauna) than acid grasslands and the situation is similar for animals inhabiting streams, ponds and lakes.

Some Archaea can tolerate and even grow best in environments with a pH far outside the range tolerated by eukaryotes. Such environments are rare, but occur in volcanic lakes and geothermal springs where they are dominated by sulphur-oxidising bacteria whose pH optima lie between 2 and 4 and which cannot grow at neutrality (Stolp, 1988). *Thiobacillus ferroxidans* occurs in the waste from industrial metal-leaching processes and tolerates pH 1; *T. thiooxidans* cannot only tolerate but can grow at pH 0. Towards the other end of the pH range are the alkaline environments of soda lakes with pH values of 9–11, which are inhabited by cyanobacteria such as *Anabaenopsis arnoldii* and *Spirulina platensis*.

2.6 Salinity

For terrestrial plants, the concentration of salts in the soil water offers osmotic resistance to water uptake. The most extreme saline conditions occur in arid zones where the predominant movement of soil water is towards the surface and crystalline salt accumulates. This occurs especially when crops have been grown in arid regions under irrigation; salt pans then develop and the land is lost to agriculture. The main effect of salinity is to create the same kind of osmoregulatory problems as drought and freezing and the problems are countered in much the same ways. For example, many of the higher plants that live in saline environments (halophytes) accumulate electrolytes in their vacuoles, but maintain a low concentration in the cytoplasm and organelles. Such plants maintain high osmotic pressures and so remain turgid, and are protected from the damaging action of the accumulated electrolytes by polyols and membrane protectants.

Freshwater environments present a set of specialised environmental conditions because water tends to move into organisms from the environment and this needs to be resisted. In marine habitats, the majority of organisms are isotonic to their environment so that there is no net flow of water, but there are many that are hypotonic so that water flows out from the organism to the environment, putting them in a similar position to terrestrial organisms. Thus, for many aquatic organisms the regulation of body fluid concentration is a vital and sometimes an energetically expensive process. The salinity of an aquatic environment can have an important influence on distribution and abundance, especially in places like estuaries where there is a particularly sharp gradient between truly marine and freshwater habitats.

The freshwater shrimps *Palaemonetes pugio* and *P. vulgaris*, for example, co-occur in estuaries on the eastern coast of the USA at a wide range of salinities, but the former seems to be more tolerant of lower salinities than the latter, occupying some habitats from which the latter is absent. Figure 2.22 shows the mechanism likely to be underlying this (Rowe, 2002). Over the low salinity range (though not at the effectively lethal lowest salinity) metabolic expenditure was significantly lower in *P. pugio*. *P. vulgaris* requires far more energy simply to maintain itself, putting it at a severe disadvantage in competition with *P. pugio* even when it is able to sustain such expenditure.

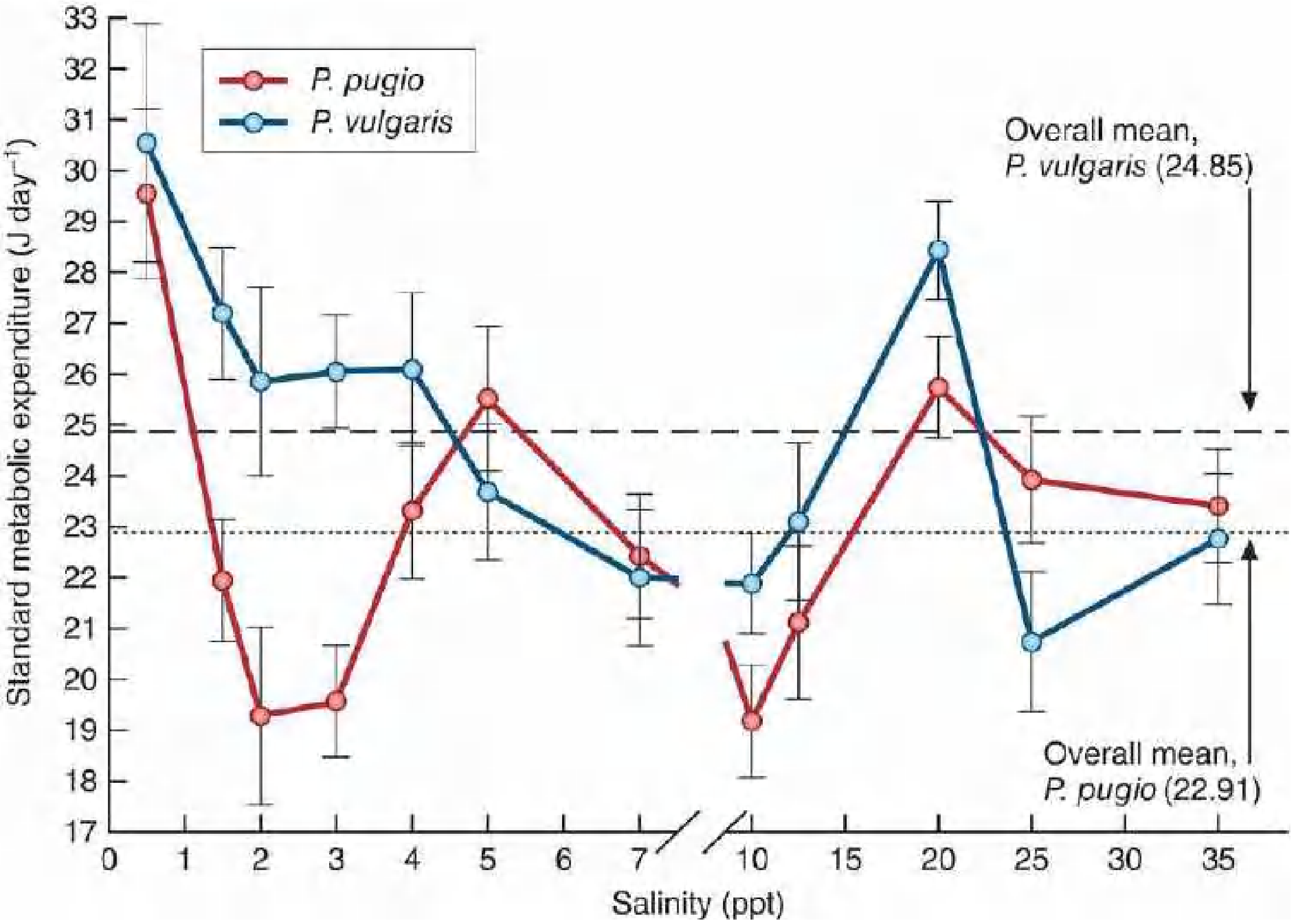


Figure 2.22 Metabolic expenditure in relation to salinity for two shrimp species. Standard metabolic expenditure (estimated through minimum oxygen consumption) in *Palaemonetes pugio* and *P. vulgaris* at a range of salinities. Note that there was significant mortality of both species over the experimental period at 0.5 ppt (parts per thousand), especially in *P. vulgaris* (75% compared with 25%).

Source: After Rowe (2002).

2.6.1 Conditions at the boundary between the sea and land

Salinity has important effects on the distribution of organisms in intertidal areas but it does so through interactions with other conditions – notably exposure to the air and the nature of the substrate.

algae and higher plants

Algae of all types have found suitable habitats permanently immersed in the sea, but permanently submerged higher plants are almost completely absent. This is a striking contrast with submerged freshwater habitats where a variety of flowering plants have a conspicuous role. The main reason seems to be that higher plants require a substrate in which their roots can find anchorage. Large marine algae, which are continuously submerged except at extremely low tides, largely take their place in marine communities. These do not have roots but attach themselves to rocks by specialised ‘holdfasts’. They are excluded from regions where the substrates are soft and holdfasts cannot ‘hold fast’. It is in such regions that the few truly marine flowering plants, for example sea grasses such as *Zostera* and *Posidonia*, form submerged communities that support complex animal communities.

Most species of higher plants that root in seawater have leaves and shoots that are exposed to the atmosphere for a large part of the tidal cycle, such as mangroves, species of the grass genus *Spartina* and extreme halophytes such as species of *Salicornia* that have aerial shoots but whose roots are exposed to the full salinity of seawater. Where there is a stable substrate in which plants can root, communities of flowering plants may extend right through the intertidal zone in a continuum extending from those continuously immersed in full-strength seawater (like the sea grasses) through to totally non-saline conditions. Salt marshes, in particular, encompass a range of salt concentrations running from full-strength seawater down to totally non-saline conditions.

Higher plants are absent from intertidal rocky seashores except where pockets of soft substrate may have formed in crevices. Instead, such habitats are dominated by the algae, which give way to lichens at and above the high tide level where the exposure to desiccation is highest. The plants and animals that live on rocky seashores are influenced by environmental conditions in a very profound and often particularly obvious way by the extent to which they tolerate exposure to the aerial environment and the forces of waves and storms. This expresses itself in the *zonation* of the organisms, with different species at different heights up the shore (Figure 2.23).

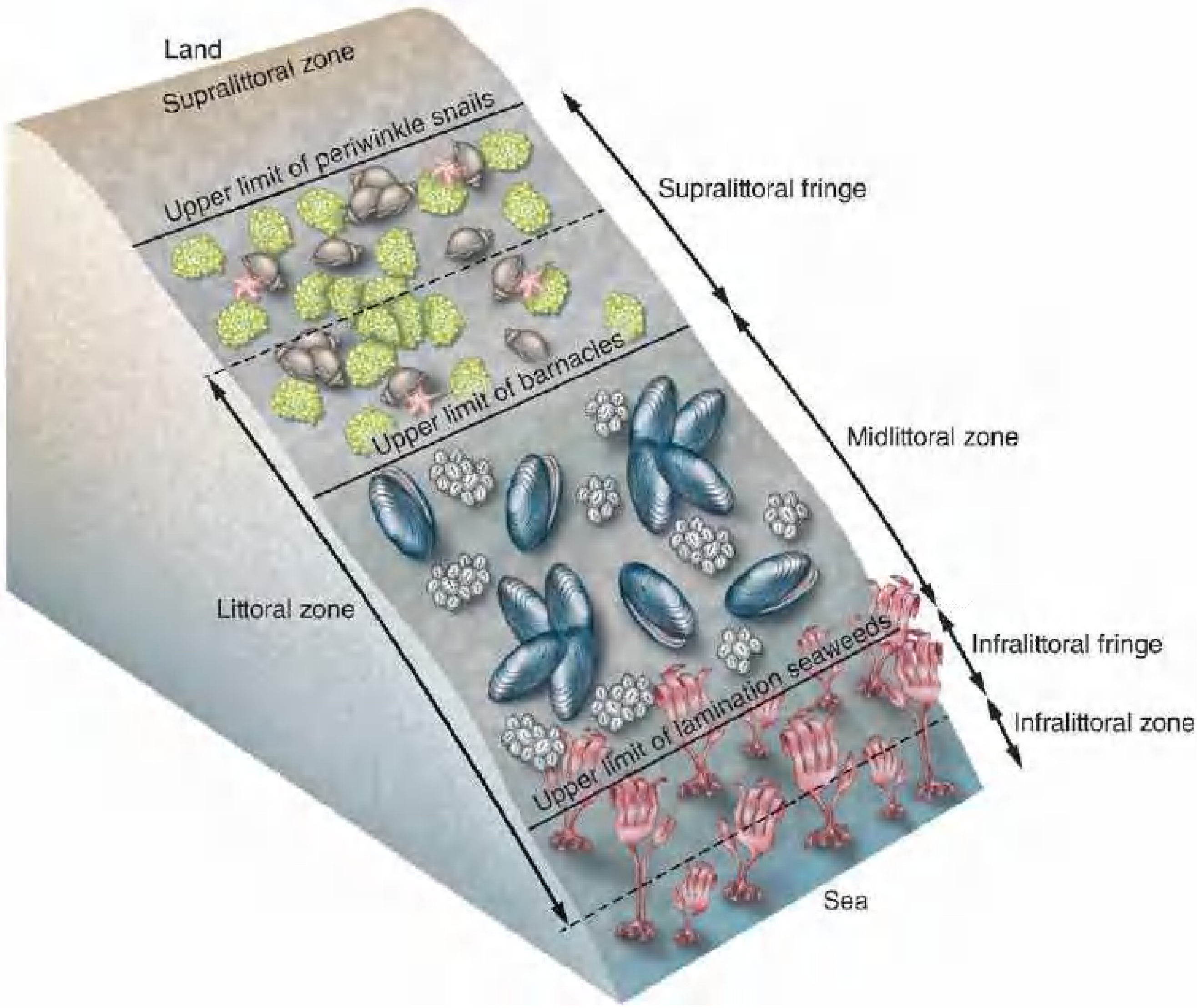


Figure 2.23 General zonation scheme for the seashore determined by relative lengths of exposure to the air and to the action of waves. The littoral zone extends between the extreme high water and extreme low water of spring tides (upper and lower dashed lines).

Source: After Raffaelli & Hawkins (1999).

zonation

The extent of the intertidal zone depends on the height of tides and the slope of the shore. Away from the shore, the tidal rise and fall are rarely greater than 1 m, but closer to shore, the shape of the landmass can funnel the ebb and flow of the water to produce extraordinary spring tidal ranges of, for example, nearly 20 m in the Bay of Fundy (between Nova Scotia and New Brunswick, Canada). In contrast, the shores of the Mediterranean Sea experience scarcely any tidal range. On steep shores and rocky cliffs the intertidal zone is very short and zonation is compressed.

To talk of 'zonation as a result of exposure', however, is to oversimplify the matter greatly (Raffaelli & Hawkins, 1999). In the first place, 'exposure' can mean a variety, or a combination of, many different things: desiccation, extremes of temperature, changes in salinity, excessive illumination and the sheer physical forces of pounding waves and storms (to which we turn in Section 2.7). Furthermore, 'exposure' only really explains the *upper* limits of these essentially marine species, and yet zonation depends on them having lower limits too. For some species there can be *too little exposure* in the lower zones. For instance, green algae would be starved of blue and especially red light if they were submerged for long periods too low down the shore. For many other species though, a lower limit to distribution is set by competition and predation. The seaweed *Fucus spiralis* will readily extend lower down the shore than usual in Great Britain whenever other competing midshore fucoid seaweeds are scarce.

2.7 Hazards, disasters and catastrophes: the ecology of extreme events

The wind and the tides are normal daily 'hazards' in the life of many organisms. The structure and behaviour of these organisms bear some witness to the frequency and intensity of such hazards in the evolutionary history of their species. Thus, most trees withstand the force of most storms without falling over or losing their living branches. Most limpets, barnacles and kelps hold fast to the rocks through the normal day-to-day forces of the waves and tides. We can also recognise a scale of more severely damaging forces (we might call them 'disasters') that occur occasionally, but with sufficient frequency to have contributed repeatedly to the forces of natural selection. When such a force recurs it will meet a population that still has a genetic memory of the selection that acted on its ancestors – and may therefore suffer less than they did. In the woodlands and shrub communities of arid zones, fire has this quality, and tolerance of fire damage is a clearly evolved response (see Section 2.3.6).

When disasters strike natural communities it is only rarely that they have been carefully studied before the event. One exception is cyclone 'Hugo' which struck the Caribbean island of Guadeloupe in 1994. Detailed accounts of the dense humid forests of the island had been published only recently before (Ducrey & Labbé, 1985, 1986). The cyclone devastated the forests with mean maximum wind velocities of 270 km h^{-1} and gusts of 320 km h^{-1}. Up to 300 mm of rain fell in 40 h. The early stages of regeneration after the cyclone (Labbé, 1994) typify the responses of long-established communities on both land and sea to massive forces of destruction. Even in 'undisturbed' communities there is a continual creation of gaps as individuals (e.g. trees in a forest, kelps on a seashore) die and the space they occupied is recolonised (see Section 18.6.1).

In contrast to conditions that we have called 'hazards' and 'disasters' there are natural occurrences that are enormously damaging, yet occur so rarely that they may have no lasting selective effect on the evolution of the species. We might call such events 'catastrophes', for example the devastating Japanese tsunami (tidal wave) of 2011, or the volcanic eruptions of Mt St Helens in 1980 or of the island of Krakatau in 1883. The next time that Krakatau erupts there are unlikely to be any genes persisting that were selected for volcano tolerance!

APPLICATION 2.7 Coral reefs and mangrove forests may ameliorate the impact of tsunamis

ecosystem services

Ecosystems often provide valuable *ecosystem services* (see Section 15.4.1) that people use and enjoy. *Provisioning services* include wild meat and berries, medicinal herbs, fibre products, fuel and drinking water; *cultural services* include aesthetic fulfillment, education and recreation; *regulating services* include the ecosystem's capacity to ameliorate the effects of pollutants or to moderate disasters (such as tsunamis); finally, *supporting services*, such as primary production and nutrient cycling, underlie all the others (Townsend, 2008).

The devastating tsunamis of 2004 and 2011, caused by earthquakes off Sumatra (9.3 on the Richter scale) and north-eastern Japan (9.0), took huge tolls in human lives and livelihoods and, hardly surprisingly, also greatly changed near-shore and coastal ecosystems, both aquatic and terrestrial (e.g. Urabe *et al.*, 2013). More surprising, perhaps, has been the finding that intact coral reefs can absorb some of the wave's power (Kunkel *et al.*, 2006). According to the American Geophysical Union, illegal coral mining off the south-west coast of Sri Lanka allowed far more destruction from the 2004 Pacific-wide tsunami than occurred in nearby areas where coral reefs were intact. It seems that exploitation of a *provisioning service* (coral crushed to create road surface) resulted in loss of a *regulating service*. Moreover, muddy shores with intact mangrove forest also seem to have moderated the devastation caused by the 2004 tsunami, both by reducing human mortality inland and by preventing the inland intrusion of saltwater that, where mangroves had been removed, ruined rice and groundnut crops (Kathiresan & Rajendran, 2005). The conservation and restoration of coral reefs and mangrove forests should help protect against these natural catastrophes.

2.8 Environmental pollution

A number of environmental conditions that are, regrettably, becoming increasingly important are due to the accumulation of toxic byproducts of human activities. Sulphur dioxide emitted from power stations, and metals like copper, zinc and lead, dumped around mines or deposited around refineries, are just some of the pollutants that limit distributions, especially of plants. Many such pollutants are present naturally but at low concentrations, and some are indeed essential nutrients for plants. But in polluted areas their concentrations can rise to lethal levels. The loss of species is often the first indication that pollution has occurred, and changes in the species richness of a river, lake or area of land provide bioassays of the extent of their pollution.

rare tolerators

Yet it is rare to find even the most inhospitable polluted areas entirely devoid of species; there are usually at least a few individuals of a few species that can tolerate the conditions. Even natural populations from unpolluted areas often contain a low frequency of individuals that tolerate the pollutant; this is part of the genetic variability present in natural populations. Such individuals may be the only ones to survive or colonise as pollutant levels rise. They may then become the founders of a tolerant population to which they have passed on their 'tolerance' genes, and, because they are the descendants of just a few founders, such populations may exhibit notably low genetic diversity overall (Figure 2.24). Thus, in very simple terms, a pollutant has a two-fold effect. When it is newly arisen or is at extremely high concentrations, there will be few individuals of any species present (the exceptions being naturally tolerant variants or their immediate descendants). Subsequently, however, the polluted area is likely to support a much higher density of individuals, but these will be representatives of a much smaller range of species than would be present in the absence of the pollutant. Such novel, species-poor communities are now an established part of human environments (Bradshaw, 1987).

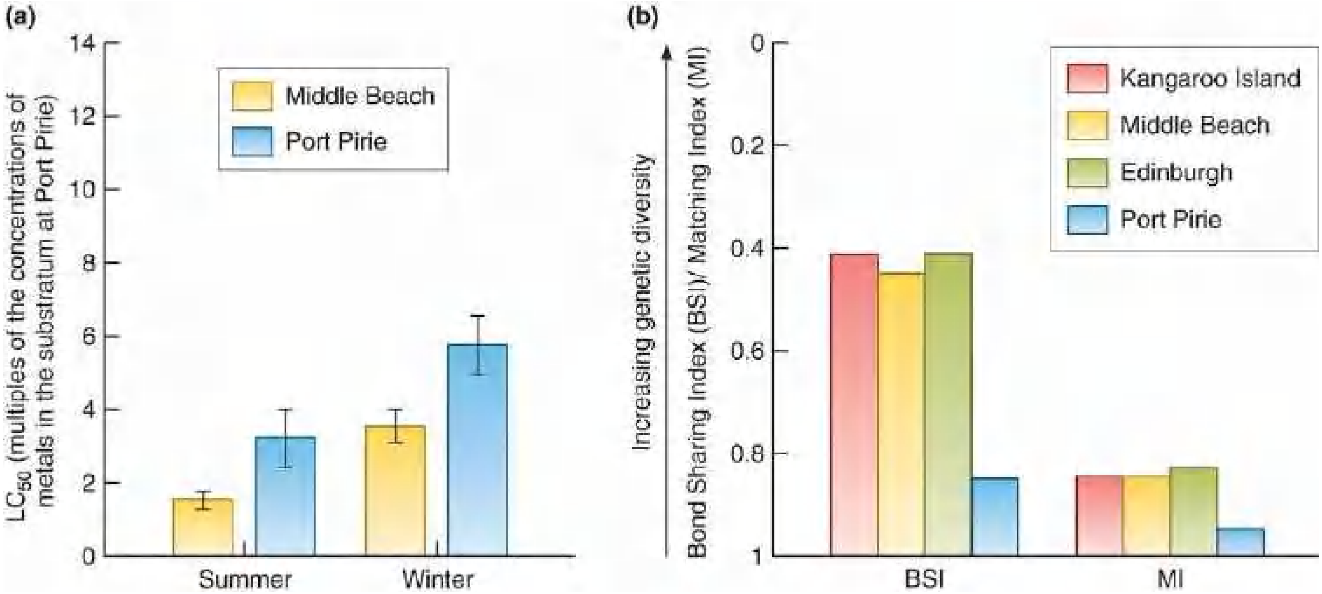


Figure 2.24 Individuals of *Platynympha longicaudata* in a polluted site are more tolerant of pollution and have lower genetic diversity. (a) Tolerance of this marine isopod around Port Pirie, South Australia (the largest lead smelting operation in the world), was significantly higher ($P < 0.05$) than for animals from a control (unpolluted) site, as measured by the concentration in food of a combination of metals (lead, copper, cadmium, zinc and manganese) required to kill 50% of the population (LC_{50}). (b) Genetic diversity at Port Pirie was significantly lower than at three unpolluted sites, as measured by two indices of diversity based on RAPD (random amplified polymorphic DNA).

Source: After Ross *et al.* (2002).

APPLICATION 2.8 Bioremediation and phytomining

Species may differ greatly in their ability to tolerate pollutants. Some plants (often assisted by microbial symbionts in their rhizosphere) are *hyperaccumulators* of heavy metals – lead, cadmium and so on – with an ability not only to tolerate but also to accumulate much higher concentrations than the norm. As a result, species such as *Solanum nigrum* have an important role to play in bioremediation (Sun *et al*., 2017), removing pollutants from the soil so that eventually other, less tolerant plants can grow there too.

Some may even be used for *phytomining*, where hyperaccumulator plants are used to accumulate a metal of interest from metal-rich soils and transport them to the shoots, followed by harvesting of the shoots as a bio-ore (Thijs *et al*., 2017). Thus, *Alyssum bertolonii* can accumulate in its aerial parts 7000–12 000 μg g^{-1} dry weight of nickel, while *Arabidopsis halleri* and *S. nigrum* can accumulate and tolerate similarly high concentrations of zinc and cadmium, respectively.

Organisms with bioremediation potential also include fungi and bacteria (de Alencar *et al*., 2017), and remediation can be directed not only at heavy metals but also at many other pollutants, including petroleum- and explosives-contaminated soil and polycyclic aromatic hydrocarbons.

Pollution can, of course, have its effects far from the original source. Toxic effluents from a mine or a factory may enter a watercourse and affect its flora and fauna for its whole length downstream. Effluents from large industrial complexes can pollute and change the flora and fauna of many rivers and lakes in a region and cause international disputes.

acid rain

A striking example of pollution at a distance is the creation of ‘acid rain’ – atmospheric deposition of acidic constituents (particularly sulphuric and nitric acid) that reach the ground as rain, snow, particulates, gases and vapour. Acid rain results predominantly from emissions of sulphur dioxide and oxides of nitrogen (Figure 2.25a, b) from the burning of fossil fuels to generate electricity, transport and industry, and increased dramatically after the Industrial Revolution in Europe and North America. Profound ecological effects, often across national boundaries at considerable distances from the polluting source, have included damage to forests and soil communities and acidification of rivers and lakes, with associated loss of biodiversity and recreational activities such as fishing.

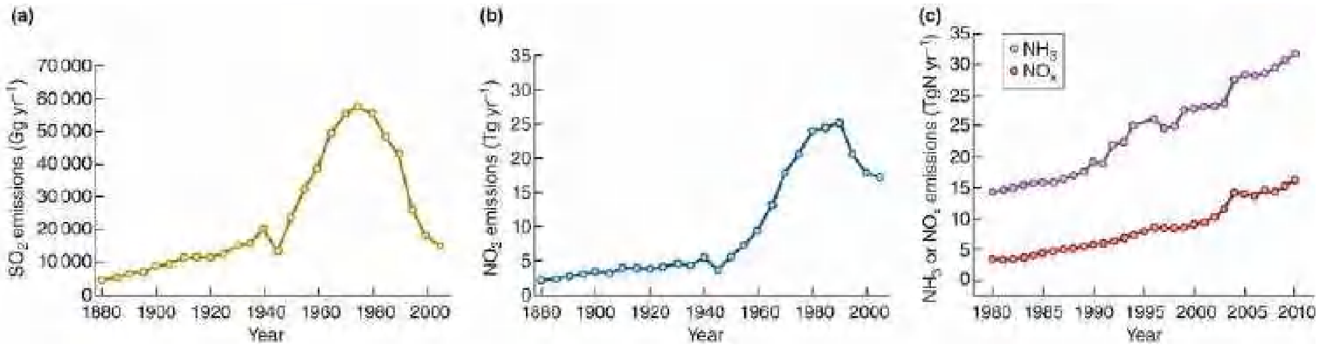


Figure 2.25 Acid emissions have been decreasing in Europe since 1970 while they continued to increase in China. Annual emissions of (a) sulphur dioxide and (b) nitrogen dioxide in Europe from 1880 to 2005 and of (c) ammonia and oxides of nitrogen (NO_x) in China from 1980 to 2010.

Source: (a, b) After Hildrew (2018). (c) From Liu *et al.* (2013).

The only option to treat the causes of acidic deposition is to reduce emissions and the introduction of stringent air pollution regulations in Europe and North America, aimed at sulphur dioxide, oxides of nitrogen and ammonia, produced impressive results. In the UK alone, emissions of sulphur dioxide fell by 94% and nitrogen by 58% between 1970 and 2010. Emission reductions in Europe as a whole have been almost as good, while reductions in North America have been somewhat smaller. It should be stressed that the reductions are not entirely explained by government anti-pollution initiatives, but are partly due to the 'export' of emissions to China and elsewhere where many goods destined for import to the northern hemisphere are now made. Indeed, acid rain is much less of an issue now in the north while the highest rates of deposition currently occur in parts of Asia (Figure 2.25c).

As a consequence of emission reductions, chemical recovery of northern hemisphere waterways has been evident but biological recovery has generally been rather muted. This may be partly because chemical recovery is not yet complete or there may be biological constraints, such as a lack of colonists for previously impacted habitats or biotic resistance associated with changes to food webs, such that a simple reversal of acidification does not occur and the end point might not be the same community that existed before acidification (Hildrew, 2018). More encouraging has been the recent recovery of north-eastern US fish populations in lakes that were previously incapable of sustaining wild fish populations because of acid conditions (Warren *et al.*, 2017).

2.9 Global change

In Chapter 1 we discussed some of the ways in which global environments have changed over the long timescales involved in continental drift and the shorter timescales of the repeated ice ages. Over these timescales some organisms have failed to accommodate to the changes and have become extinct, others have migrated so that they continue to experience the same conditions but in a different place, and others have changed their nature (evolved) and tolerated some of the changes. We now turn to consider global changes that are occurring in our own lifetimes – consequences of our own activities – and that are predicted to bring about profound changes in the ecology of the planet. Although part of the wider syndrome now called 'global change', the acid rain just discussed is not truly global but rather regional because of the restricted mean residence time of the acidic pollutants in the atmosphere (a few days) compared with carbon dioxide, whose residence time is very much longer (Hildrew, 2018). We discuss this next.

2.9.1 Industrial gases and the greenhouse effect

A major element of the Industrial Revolution was the switch from the use of sustainable fuels to the use of coal (and later, oil) as a source of power. Between the middle of the 19th and the middle of the 20th century the burning of fossil fuels, together with extensive deforestation, added about 90 gigatonnes (Gt) of carbon dioxide (CO_2) to the atmosphere and more has been added since. The concentration of CO_2 in the atmosphere before the Industrial Revolution (measured in gas trapped in ice cores) was about 280 ppm, a fairly typical interglacial 'peak' (Figure 2.26a), but this had risen to around 370 ppm by the turn of the millennium (Figure 2.26b) and in May 2013 reached 400 ppm for the first time in at least the last 800 000 years.

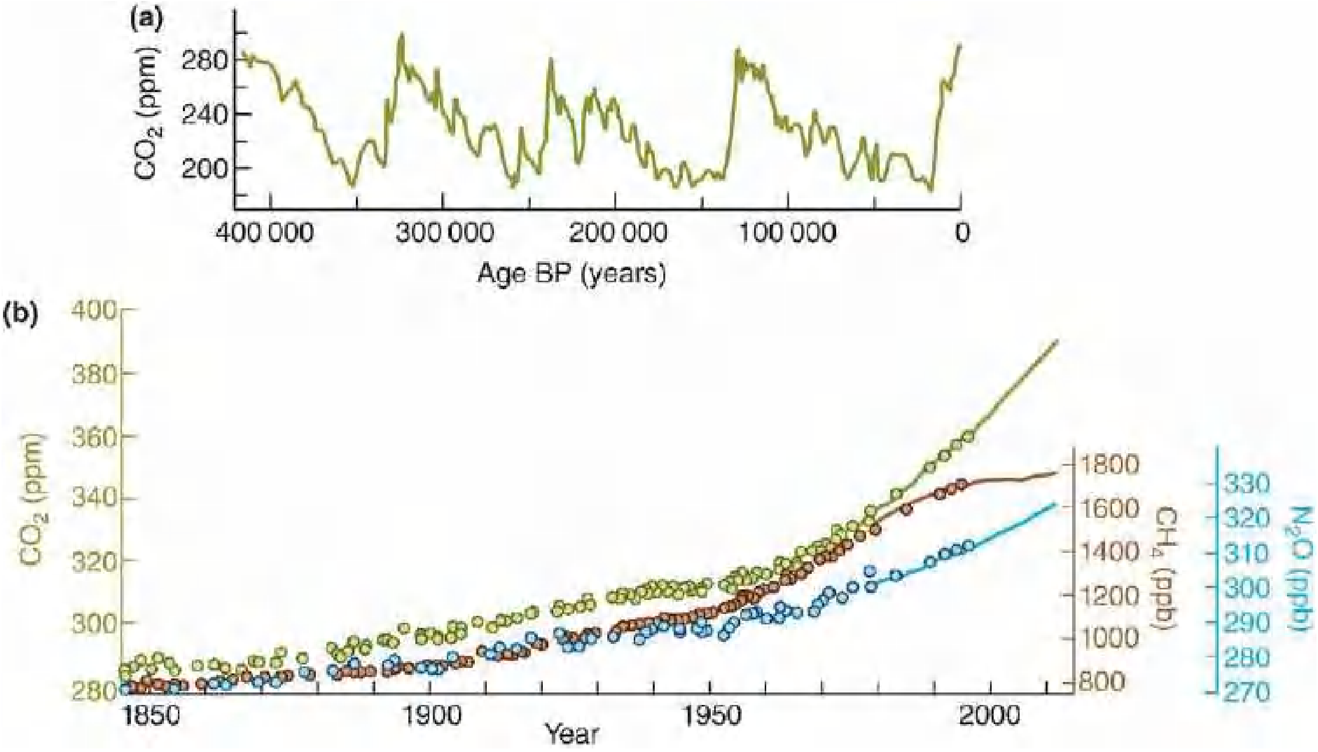


Figure 2.26 Atmospheric concentrations of CO_2during the past 420 000 years and since 1850. (a) Concentrations of CO_2 in gas trapped in ice cores from Vostok, Antarctica. Transitions between glacial and warm epochs, and peaks in CO_2, occurred around 335 000, 245 000, 135 000 and 18 000 years ago. (b) Atmospheric concentrations of the greenhouse gases CO_2 (green), methane (CH_4, brown) and nitrous oxide (N_2O, blue) determined from ice core data (dots) and from direct atmospheric measurements (lines) since the mid-18th century. BP, before present; ppb, parts per billion; ppm, parts per million.

Source: (a) After Petit *et al.* (1999) and Stauffer (2000). (b) After IPCC (2014).

Solar radiation incident on the earth's atmosphere is in part reflected, in part absorbed, and part is transmitted through to the earth's surface, which absorbs and is warmed by it. Some of this absorbed energy is radiated back to the atmosphere where atmospheric gases, mainly water vapour and CO_2, absorb about 70% of it. It is this trapped reradiated energy that heats the atmosphere in what is called the 'greenhouse effect'. The greenhouse effect was of course part of the normal environment before the Industrial Revolution and was responsible for some of the environmental warmth before industrial activity started to enhance it. At that time, the greater proportion of the greenhouse effect was due to atmospheric water vapour.

CO_2 – but not only CO_2

In addition to the enhancement of greenhouse effects by CO_2 emissions, other trace gases have increased markedly in the atmosphere, particularly methane (CH_4) and nitrous oxide (N_2O) (Figure 2.27) and to a smaller extent the chlorofluorocarbons (CFCs, e.g. trichlorofluoromethane

(CCl_3F) and dichlorodifluoromethane (CCl_2F_2)) and some other minor contributors. Each greenhouse gas has a global warming potential (usually expressed as 'equivalents of CO_2') that depends on how long it stays in the atmosphere and how strongly it absorbs energy. Thus, CH_4 and N_2O have global warming potentials some 30 and 300 times that of CO_2 over a 100-year period (they persist in the atmosphere for around 10 or 100 years, respectively, compared with thousands of years for CO_2, but absorb energy much more efficiently). Together, these gases contribute about 35% to enhancing the greenhouse effect, compared with 65% by CO_2 (Figure 2.26). The increase in CH_4 is mainly of microbial origin in intensive agriculture on anaerobic soils (especially increased rice production) and in the digestive process of ruminants (a cow produces approximately 40 litres of CH_4 each day). N_2O is emitted during agricultural and industrial production and the combustion of fossil fuels and solid waste. The effect of the CFCs from refrigerants, aerosol propellants and so on was potentially great (their global warming potentials are thousands or tens of thousands greater than CO_2), but international agreements, mainly to counteract damage to the ozone layer, have strongly moderated increases in their concentrations. However, the rate of increase in annual greenhouse gas emissions has accelerated since the turn of the millennium (Figure 2.27).

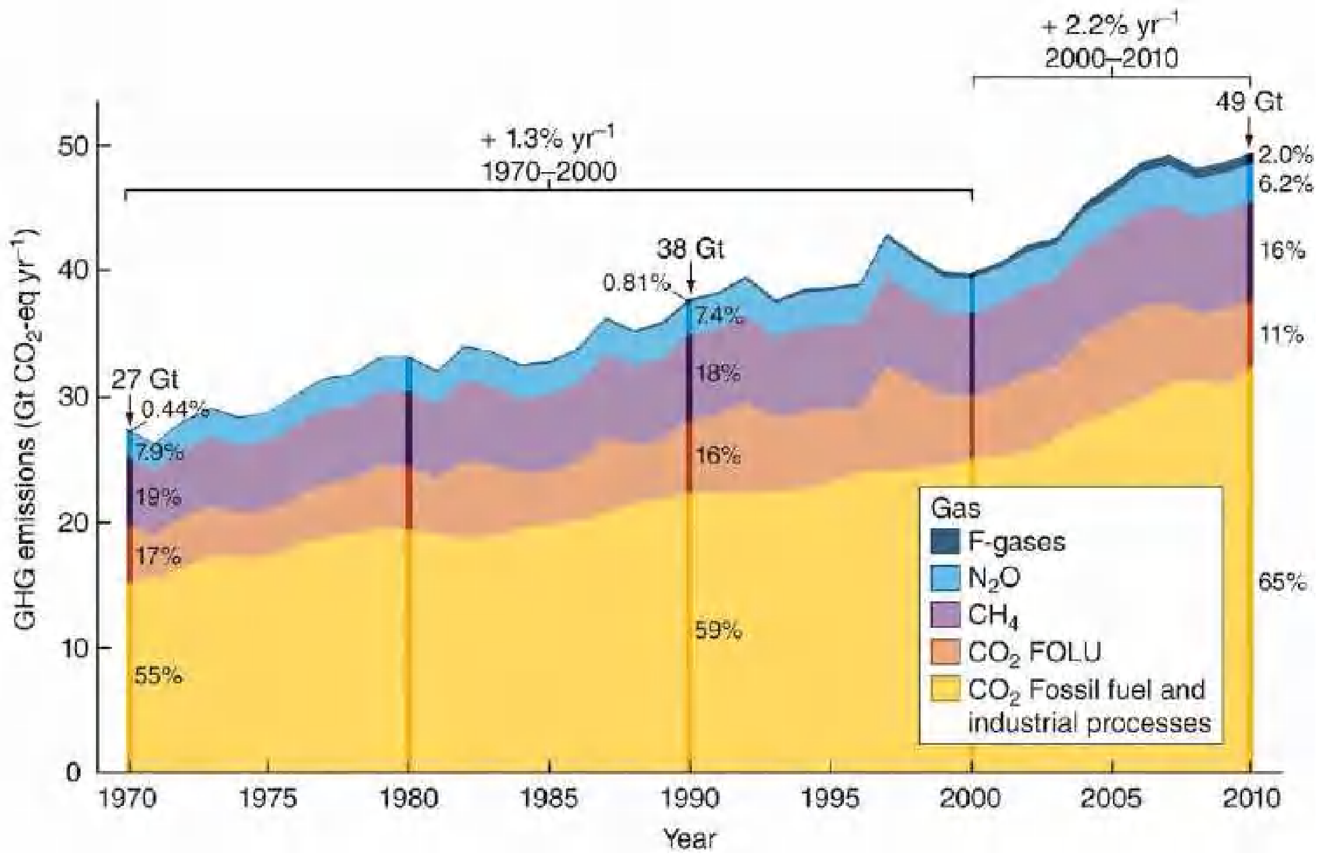


Figure 2.27 Total annual anthropogenic greenhouse gas (GHG) emissions from 1970 to 2010 converted to gigatonne equivalents of CO_2 per year. FOLU, forestry and other land use change.

Source: IPCC (2014).

It is possible to draw up a balance sheet of how the CO_2 produced by human activities translates into changes in concentration in the atmosphere. Human activities have released more than 2000 Gt CO_2 since 1750, but the increase in atmospheric CO_2 accounts for only 40% of this (IPCC, 2014). The oceans absorb an estimated 30% of CO_2 released by human activities. Furthermore, recent analyses indicate that terrestrial vegetation has been 'fertilised' by the increased atmospheric CO_2, so that a considerable amount of extra carbon has been locked up in vegetation biomass. And more is to be found as soil carbon. This softening of the blow by the oceans and terrestrial vegetation notwithstanding, however, atmospheric CO_2 and the greenhouse effect are increasing.

The most profound effect of anthropogenic CO_2 emissions, global warming, is dealt with in the next section. In addition, ocean acidification is another worrying consequence.

ocean acidification

A large proportion of anthropogenic CO_2 is absorbed by the oceans, thus far reducing seawater pH by 0.1 units since the Industrial Revolution (equivalent to a 30% increase in acidity) as well as reducing carbonate ion concentrations. We have already seen that pH is a condition with significant influences on the success of organisms, but the fact that many parts of the ocean are also becoming undersaturated with calcium carbonate minerals is expected to have profound consequences for calcifying species such as corals, molluscs, sea urchins and calcareous plankton. On the other hand, photosynthetic production in the oceans is likely to benefit from higher CO_2 concentrations.

2.9.2 Global warming

We started this chapter discussing temperature, moved through a number of other environmental conditions to pollutants, and now return to temperature because of the effects of those pollutants on global temperatures. The globally averaged combined land and ocean surface temperature has increased by about 0.85°C from 1880 to 2012 (Figure 2.28a). We have already witnessed melting of arctic ice and rises in sea level (Figure 2.28b) (related to thermal expansion and the input of ice meltwater) and can expect further melting of the ice caps, a consequent rising of sea level and significant shifts in the pattern of global climates and changes to the distribution of species.

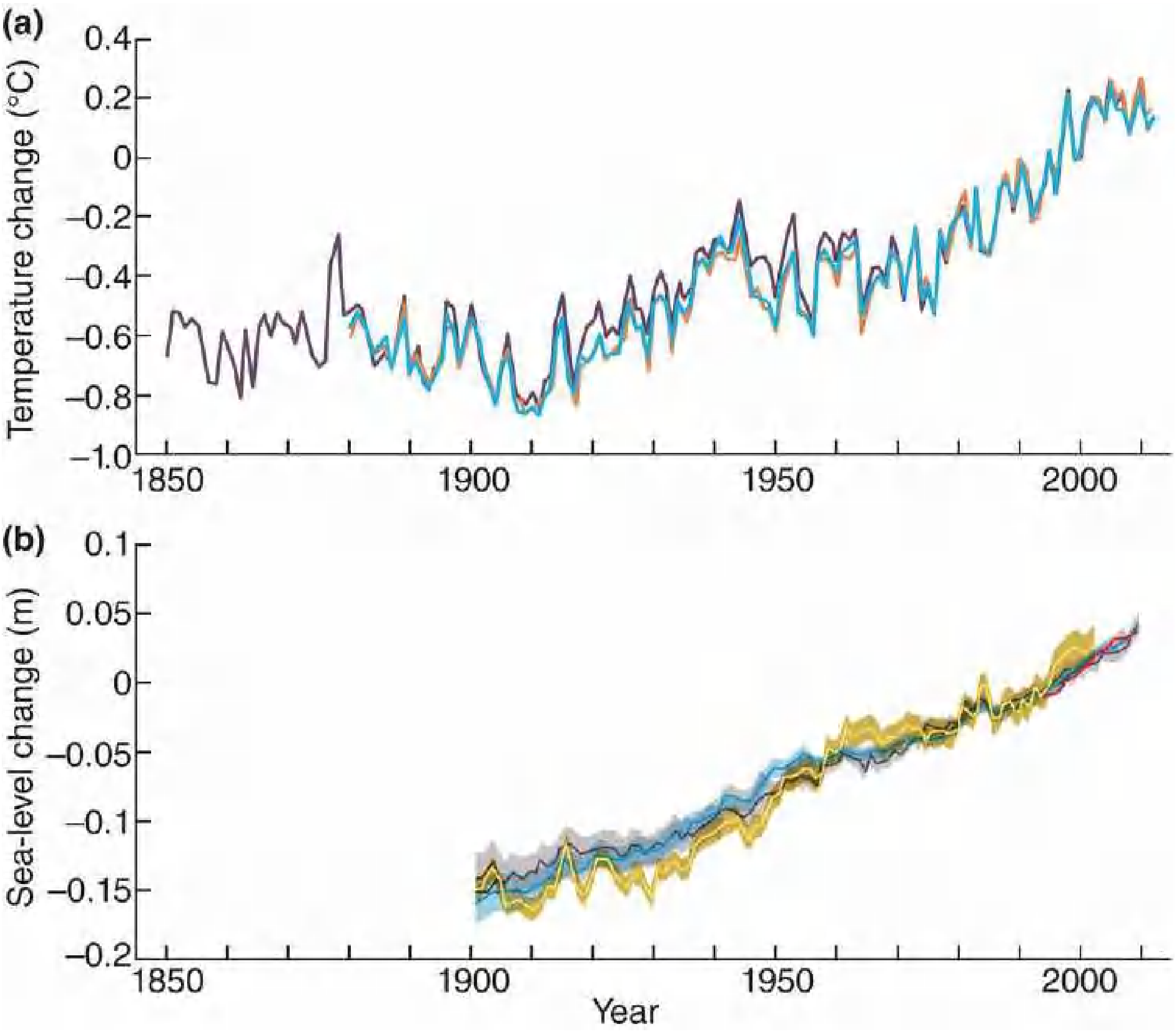


Figure 2.28 **Annual land and ocean surface temperature anomalies and sea-level changes.** (a) Globally averaged combined annual land and ocean surface temperature anomalies from 1850 relative to the average over the period 1986–2005. Colours indicate different datasets. (b) Globally averaged annual sea-level changes from 1900 relative to the average over the period 1986–2005. Colours indicate different datasets that have been aligned to have the same value in 1993. Uncertainties are indicated by shading.

Source: IPCC (2014).

Predictions of the extent of global warming resulting from the enhanced greenhouse effect come from two sources: (i) trends detected in measured datasets, including the width of tree rings, sea-level records and measures of the rate of retreat of glaciers; and (ii) predictions based on sophisticated computer models that simulate the world's climate according to a variety of possible mitigation scenarios. The latter range from the best case, where there is a concerted and international political drive to minimise the temperature rise by the use and development of efficient technologies (e.g. switching to renewable energy and bioenergy, carbon capture and geological storage) to the worst case where very little is done and the expected outcome is close to the no-mitigation baseline scenario.

global distribution of climate change

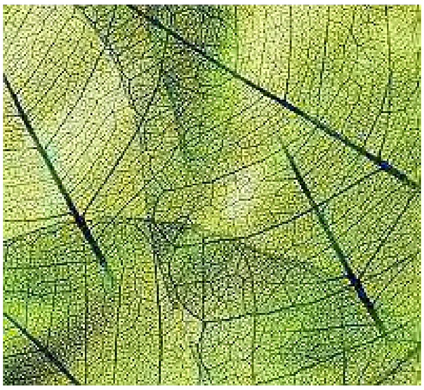
Global warming so far has not been evenly distributed over the surface of the earth, and neither will it be in the future. Northern high latitudes are expected to change more rapidly than the tropics, land areas will change more rapidly than the oceans, and small islands and coastal regions will be particularly prone to associated rises in sea level.

We have emphasised how the distributions of species are strongly influenced by temperature and water availability, and how many organisms are impacted by occasional extremes rather than by average conditions. Computer modelled projections imply that global climatic change will also bring greater variance in temperature, rainfall, hurricanes and so on. Hence, not only the predicted average climate changes, but also the increased frequency and severity of extremes, are certain to be accompanied by marked responses in the distribution of species and biomes.

can the biota keep up with the pace?

Global temperatures have changed naturally in the past, as we have seen. We are currently approaching the end of one of the warming periods that started around 20 000 years ago, during which global temperatures have risen by about 8°C. The greenhouse effect adds to global warming at a time when temperatures are already higher than they have been for 400 000 years. Buried pollen provides evidence that North American forest boundaries have migrated north at rates of 100–500 m year^{-1} since the last ice age. However, this rate of advance has not been fast enough to keep pace with postglacial warming. The rate of warming forecast to result from the greenhouse effect is 50–100 times faster than postglacial warming. Thus, of all the types of environmental pollution caused by human activities, none may have such profound effects as global warming. We must expect latitudinal and altitudinal changes to species' distributions and widespread extinctions as floras and faunas fail to track and keep up with the rate of change in global temperatures. What is more, large tracts of land over which vegetation might advance and retreat have been fragmented in the process of civilisation, putting major barriers in the way of vegetational advance. It will be very surprising if many species do not get lost on the journey.

The ecological implications of greenhouse gas emissions are profound indeed for the spread of pest species, for future conservation and restoration management, and for the production of wild fisheries, agriculture and aquaculture. These topics will crop up throughout the book, but especially in its final chapter.



Chapter 3
Resources

3.1 Introduction

According to Tilman (1982), all things consumed by an organism are resources for it. But consumed does not simply mean ‘eaten’. Bees and squirrels do not eat holes, but a hole that is occupied is no longer available to another bee or squirrel. Similarly, females that have already mated may be unavailable to other mates. All these things have been consumed in the sense that their stock or supply can be reduced by the activities of the organisms concerned.

autotrophs and heterotrophs

There is a fundamental distinction between *autotrophic* and *heterotrophic* organisms. Autotrophs assimilate simple inorganic resources into packages of organic molecules (proteins, carbohydrates, etc.). These become the resources for the heterotrophs (decomposers, parasites, predators and grazers), which take part in a chain of events in which each consumer of a resource becomes, in turn, a resource for another consumer. At each link in this food chain, the most obvious distinction is between *saprotrophs* and *predators*. Saprotrophs – bacteria, fungi and detritivorous animals (see Chapter 11) – use other organisms as food but only after they have died, or they consume another organism’s waste or secretory products. Predators, defined broadly, feed on other living organisms, or parts of other living organisms (see Section 3.7).

photoautotrophs and chemoautotrophs

Autotrophs may themselves be divided into *photoautotrophs* and *chemoautotrophs*. The photoautotrophs – green plants and algae, and photosynthetic protists and bacteria – utilise solar radiation, carbon dioxide (CO_2), water and mineral nutrients as resources. Through photosynthesis, they use the radiation as a source of energy to reduce CO_2 to obtain the organic compounds and energy that they need for growth and reproduction. Directly or indirectly, photosynthesis is the source of all energy in terrestrial and most aquatic ecosystems. Its evolution has led to the current 21% levels of oxygen in the atmosphere, driving down the levels of CO_2. By contrast, chemoautotrophs – certain bacteria and archaea – use chemical energy from the oxidation of inorganic substances such as hydrogen sulphide, elemental sulphur, ferrous iron or ammonia to reduce CO_2 and so obtain the organic compounds and energy that they need. They typically live in ‘extreme’ environments such as hot springs and deep-sea vents.

For both autotrophs and heterotrophs, resources, once consumed, are no longer available to another consumer. This has the important consequence that organisms may *compete* with each other to capture a share of a limited resource – a topic to which we turn in Chapter 5.

In this chapter we start (Sections 3.2–3.6) with the resources that fuel the growth of individual plants, and so, collectively, determine the primary productivity of whole areas of land or volumes of water: the rate, per unit area or volume, at which plants produce biomass. Broad-scale patterns of primary productivity are examined in Chapter 20. Relatively little space in this chapter (Section 3.7) is given to food as a resource for animals, simply because a series of later chapters (Chapters 9–13) is devoted to the ecology of predators, grazers, parasites and saprotrophs (the consumers and decomposers of dead organisms). This chapter then closes with sections on two important topics, drawing on material from the present chapter and the last – one (Section 3.8) on the ecological niche and resource classification, and a second (Section 3.9) on a so-called metabolic theory of ecology.

3.2 Radiation

Solar radiation is the only source of energy that can be used in metabolic activities by green plants and algae. It comes to the plant as a flux of radiation from the sun, either directly, or having been diffused to a greater or lesser extent by the atmosphere, or after being reflected or transmitted by other objects. The direct fraction is highest at tropical latitudes north and south of the equator, since cloud cover is typically high at the equator itself (Figure 3.1). Moreover, for much of the year in temperate climates, and for the whole of the year in arid climates, the leaf canopy in terrestrial communities does not cover the land surface, so that most of the incident radiation falls on bare branches or bare ground.

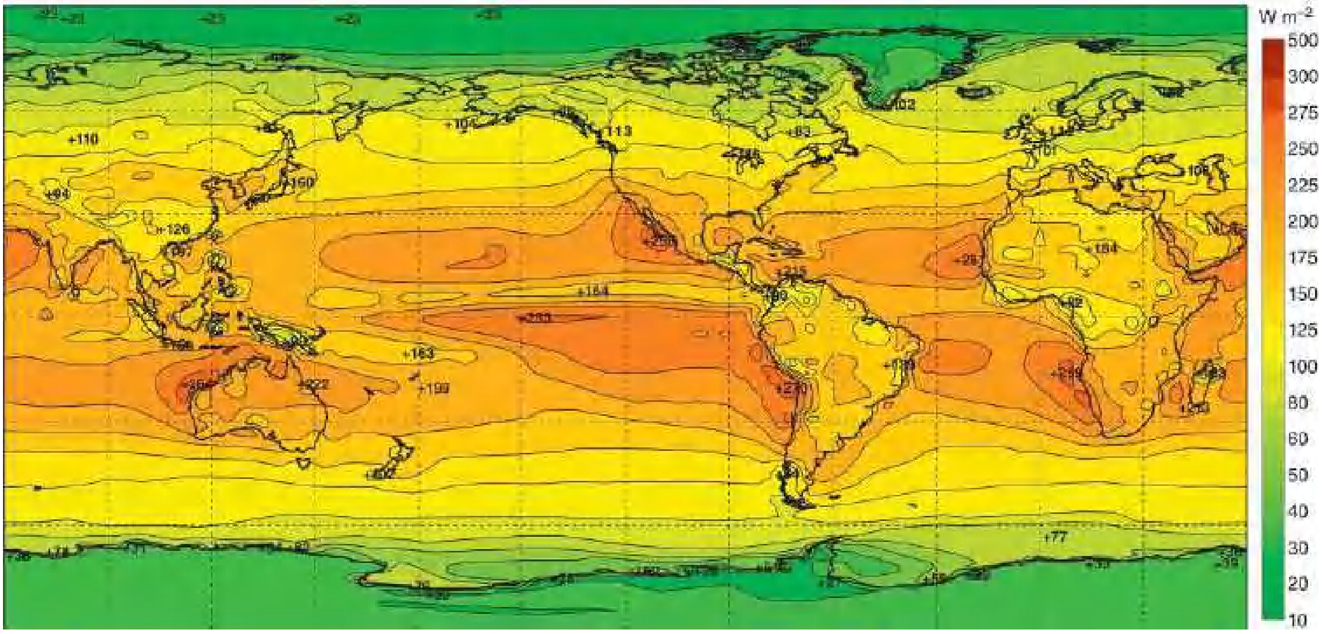


Figure 3.1 Global map of the solar radiation absorbed annually in the earth–atmosphere system: from data obtained with a radiometer on the Nimbus 3 meteorological satellite.

Source: After Laing & Evans (2011).

the fate of radiation

When a plant intercepts radiant energy it may be reflected (with its wavelength unchanged), transmitted (after some wavebands have been filtered out) or absorbed. Part of the fraction that is absorbed may raise the plant's temperature and be reradiated at much longer wavelengths. In terrestrial plants, part may contribute latent heat of evaporation of water and so power the transpiration stream. Something like 80% may reach the chloroplasts and drive the process of photosynthesis, but of this, only a small proportion may end up in the plant's organic molecules, because there is insufficient capacity in carbon metabolism to use all the energy absorbed. Again, the remainder is dissipated as heat.

radiant energy must be captured or is lost forever

Radiant energy is converted during photosynthesis into energy-rich chemical compounds of carbon, which will subsequently be broken down in respiration, either by the plant itself or by organisms that consume it. But unless the radiation is captured and chemically fixed at the instant it falls on the leaf, it is irretrievably lost for photosynthesis. Radiant energy that has been fixed in photosynthesis passes just once through the world. This is in complete contrast to an atom of nitrogen or carbon or a molecule of water that may cycle repeatedly through endless generations of organisms.

photosynthetically active radiation

Solar radiation is a resource continuum: a spectrum of different wavelengths. But the photosynthetic apparatus is able to gain access to energy in only a restricted band of this spectrum. All green plants depend on chlorophyll and other pigments for the photosynthetic fixation of carbon, and these pigments fix radiation in a waveband between roughly 400 and 700 nm. This is the band of *photosynthetically active radiation* (PAR). It corresponds broadly with the range of the spectrum visible to the human eye that we call 'light'. About 56% of the radiation incident on the earth's surface lies outside the PAR range and is thus unavailable as a resource for green plants. In other organisms, though, there are pigments, for example bacteriochlorophyll in bacteria, that operate in photosynthesis outside the PAR range of green plants. Our understanding of the breadth and importance of prokaryotic photosynthesis is increasing rapidly (Bryant & Frigaard, 2006).

Note that it is not the case simply that the rate of photosynthesis increases with the intensity of radiation. At high intensities, excess light can increase the production of potentially damaging intermediates in the photosynthetic process and *photoinhibition* of photosynthesis may occur (Li *et al.*, 2009), though what constitutes excess light varies considerably with the state of the plant. Under conditions of excess light, rapid changes in the photosynthetic membrane result in the excess absorbed light energy being harmlessly dissipated as heat, but the highest intensities of radiation may also lead to dangerous overheating. Radiation is an essential resource for plants, but they can have too much as well as too little.

Nonetheless, the highest efficiency of utilisation of radiation by green plants is 3–4.5%, obtained from cultured microalgae at low intensities of PAR. In tropical forests values fall within the range 1–3%, and in temperate forests 0.6–1.2%. The approximate efficiency of temperate crops is only about 0.6%. These can themselves be viewed in the context of a theoretical maximum efficiency of photosynthesis of 4.5–6% (Zhu *et al.*, 2010). It is on such paltry levels of efficiency that the energetics of all communities depend.

3.2.1 Variations in the intensity and quality of radiation

systematic variations in supply

One important reason why plants seldom achieve their full photosynthetic capacity is that the intensity of radiation varies continually (Figure 3.2), and the plant morphology and physiology that are optimal for photosynthesis at one intensity will be suboptimal at another. As with all resources, this supply of radiation can vary both systematically and unsystematically. Annual and diurnal rhythms are systematic variations in solar radiation (Figure 3.2a, b). The green plant experiences periods of famine and glut in its radiation resource every 24 hours (except near the poles) and seasons of famine and glut every year (except in the tropics). In aquatic habitats, an additional systematic and predictable source of variation in radiation intensity is the reduction in intensity with depth in the water column, though the extent of this may vary greatly. For example, differences in water clarity mean that seagrasses may grow on solid substrates as much as 90 m below the surface in the relatively unproductive open ocean, whereas macrophytes in fresh waters rarely grow at depths below 10 m (Sorrell *et al.*, 2001), and often only at considerably shallower locations, in large part because of differences in concentrations of suspended particles and phytoplankton (Figure 3.2c).

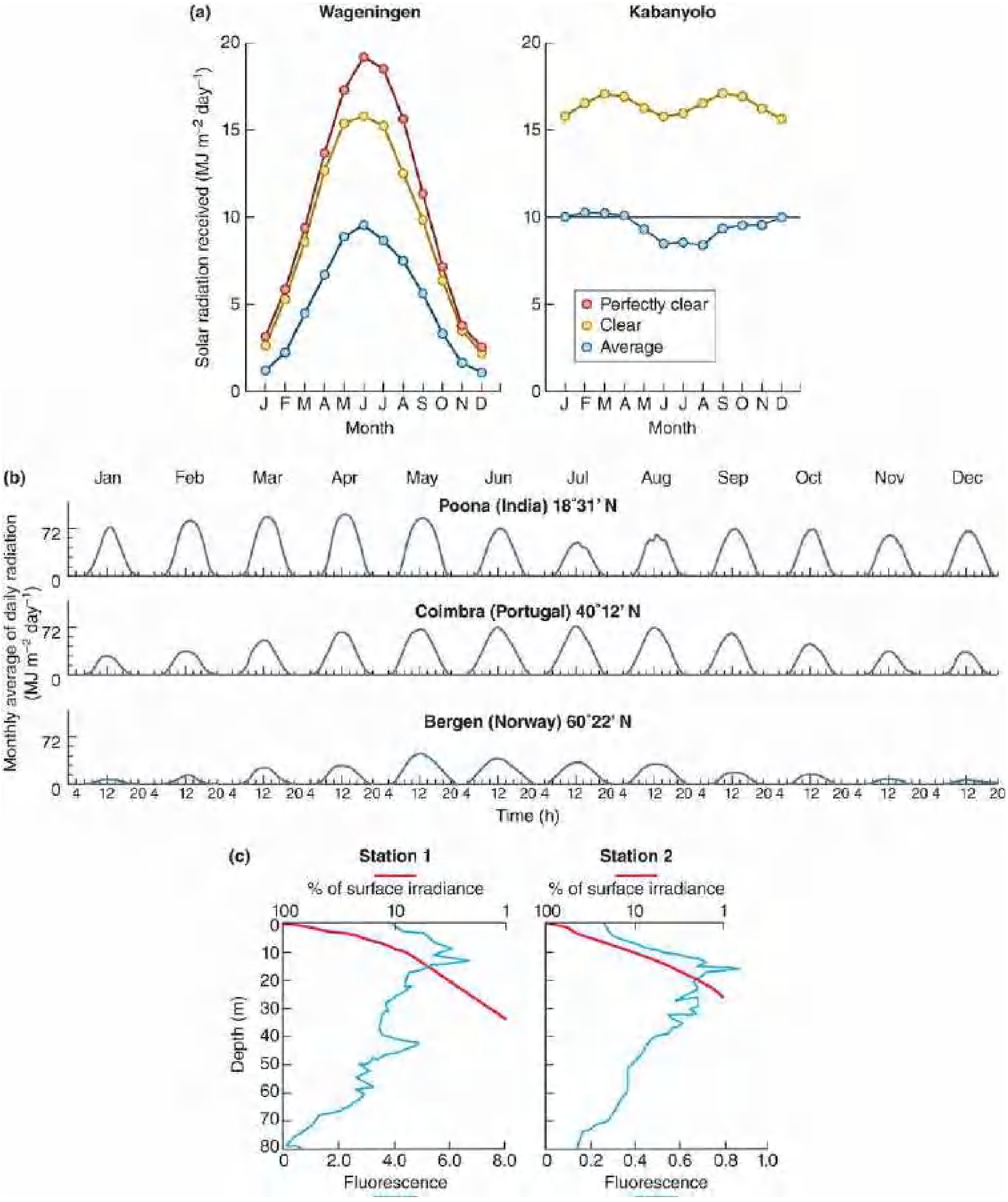

(a)
Wageningen
Kabanyolo
Solar radiation received (MJ m^{-2} day^{-1})
0
5
10
15
20
J F M A M J J A S O N D
Month
Perfectly clear
Clear
Average
(b)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Poona (India) 18°31′ N
Coimbra (Portugal) 40°12′ N
Bergen (Norway) 60°22′ N
Monthly average of daily radiation (MJ m^{-2} day^{-1})
72
0
4 12 20
Time (h)
(c)
Station 1
Station 2
% of surface irradiance
100
10
1
Depth (m)
0
10
20
30
40
50
60
70
80
Fluorescence
0 2.0 4.0 6.0 8.0
0 0.2 0.4 0.6 0.8 1.0

Figure 3.2 Levels of solar radiation vary over time and space and with depth in water. (a) The daily totals of solar radiation received throughout the year at Wageningen (the Netherlands) and Kabanyolo (Uganda). (b) The monthly average of daily radiation recorded at Poona (India), Coimbra (Portugal) and Bergen (Norway). (c) The vertical distribution of algal abundance (measured as fluorescence in units of mg chlorophyll a m^{-3}) and of irradiance as a percentage of that at the surface, for two stations off the Arctic island of Svalbard. The decline in irradiance with water depth is apparent at both stations, but at Station 1, higher algal densities in the surface waters led to that decline being more rapid: 10% of surface irradiance at around 7 m compared with 12 m at Station 2.

Source: (a, b) After de Wit (1965) and other sources. (c) After Meshram *et al*. (2017).

shade: resource-depletion zones and spectral changes

Less systematic variations in the radiation environment of a leaf are caused by the nature and position of neighbouring leaves. Leaves in a canopy, by intercepting radiation, create a *resource-depletion zone* (RDZ) – in this case, a moving band of shadow over other leaves of the same plant, or of others. The composition of radiation that has passed through leaves in a canopy, or through a body of water, is also altered. Typically, it is depleted in the blue and (especially through water) the red parts of the spectrum – the most effective wavelengths for photosynthesis. Figure 3.3 shows an example for the variation with depth in a freshwater habitat.

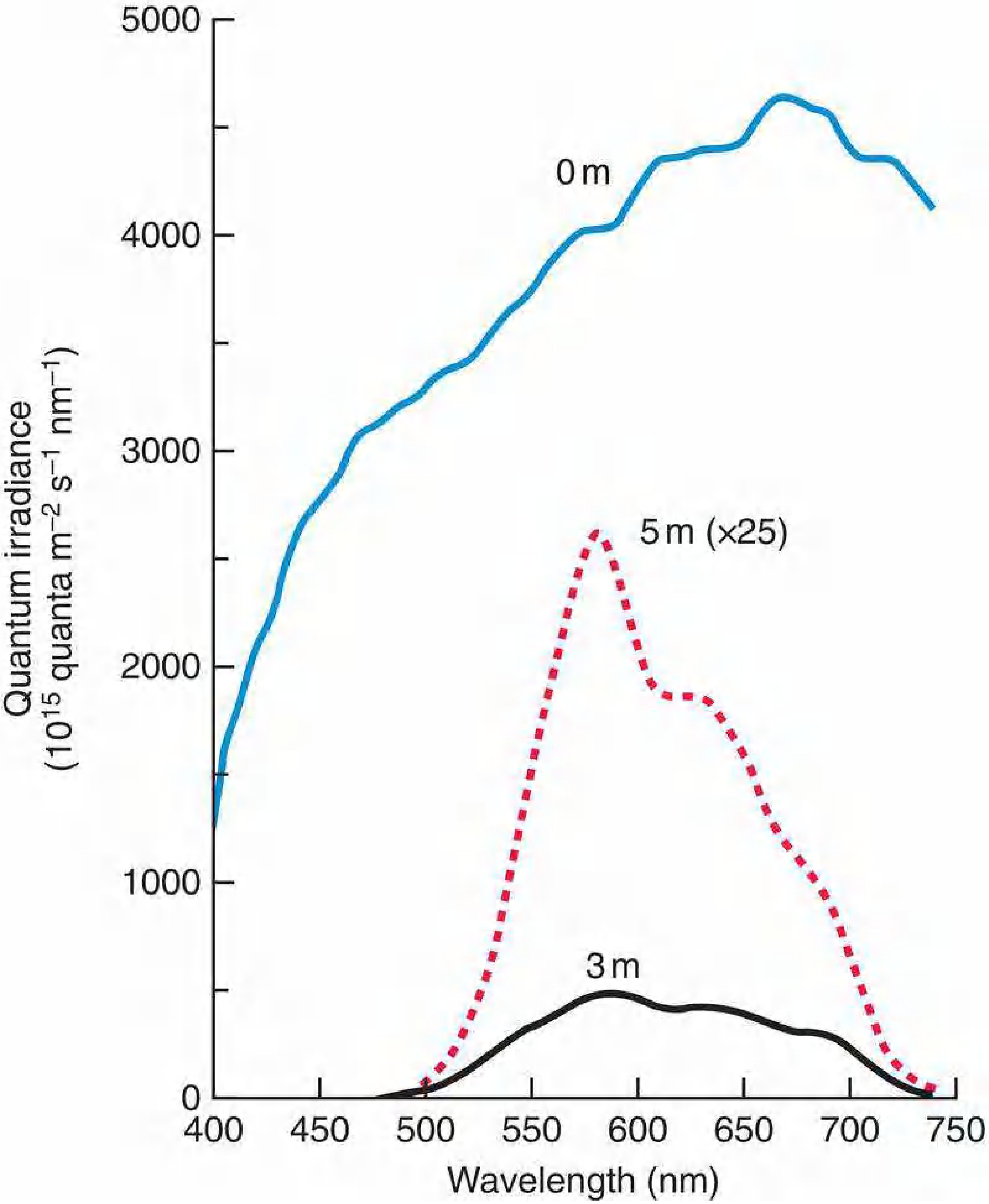


Figure 3.3 **The spectral distribution of radiation changes with depth** as shown here for Lake Burley Griffin, Australia. Note that photosynthetically active radiation lies broadly within the range 400–700 nm.

Source: After Kirk (1994).

sun and shade species

The way in which organisms react to systematic, predictable patterns in the supply of a resource reflects both their present physiology and their past evolution. At a very broad scale, the seasonal shedding of leaves by deciduous trees in temperate regions in part reflects the annual rhythm in the intensity of radiation – they are shed when they are least useful. Amongst terrestrial species, plants that are characteristic of shaded habitats generally use radiation at low intensities more efficiently than sun species, but the reverse is true at high intensities (Figure 3.4). Part of the difference between them lies in the physiology of the leaves, but the morphology of the plants also influences the efficiency with which radiation is captured. The leaves of sun plants are commonly exposed at acute angles to the midday sun, spreading an incident beam of radiation over a larger leaf area and effectively reducing its intensity (Poulson & DeLucia, 1993). The leaves of sun plants are also usually superimposed into a multilayered canopy. In bright sunshine even the shaded leaves in lower layers may have positive rates of net photosynthesis. Shade plants adopt a different strategy, commonly having leaves held near to the horizontal and in a single-layered canopy.

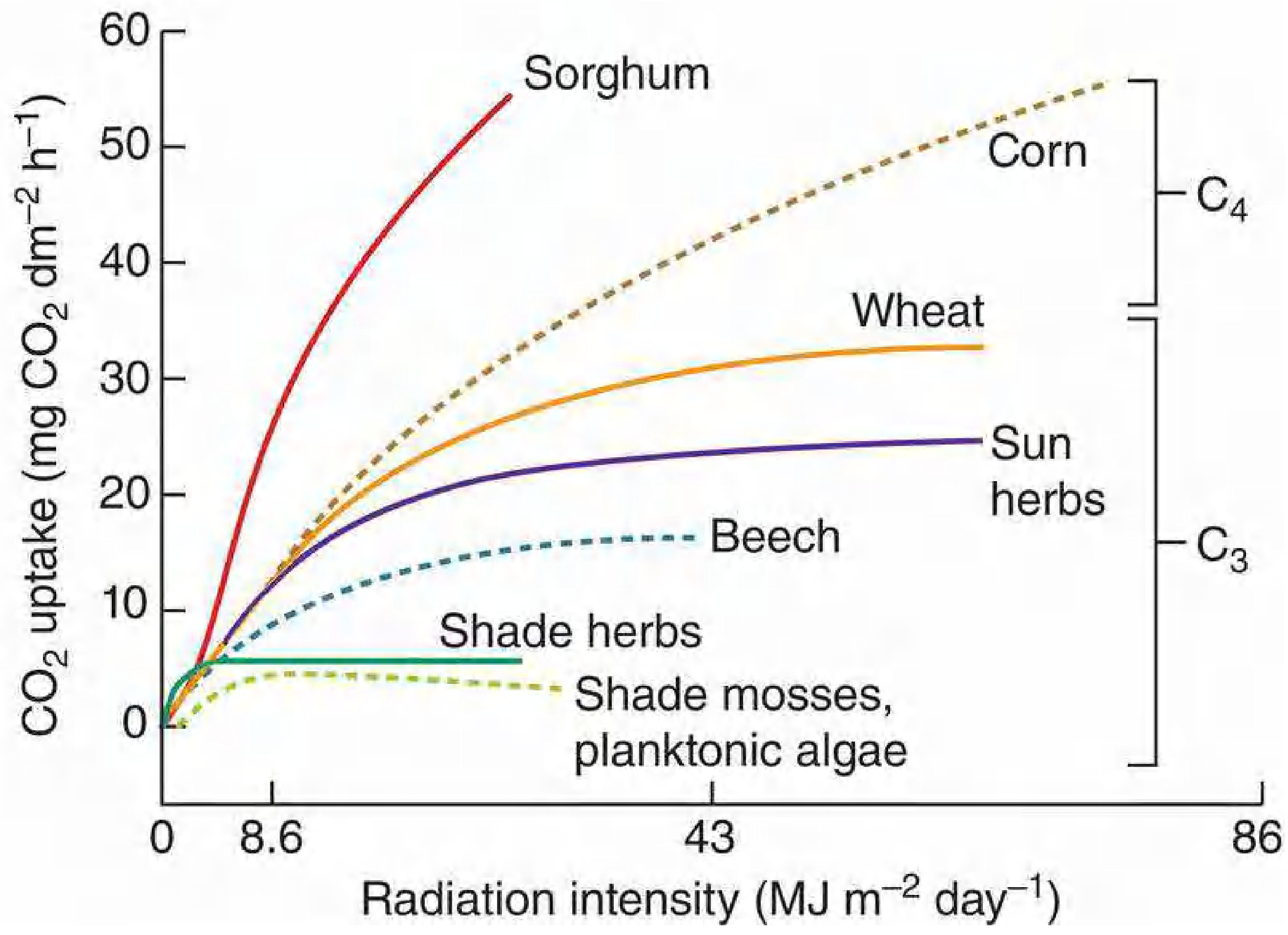


Figure 3.4 The response of photosynthesis to radiation intensity in various plants at optimal temperatures and with a natural supply of CO_2. Note that corn and sorghum are C_4 plants and the remainder are C_3 (the terms are explained in Sections 3.3.1 and 3.3.2).

Source: After Larcher (1980), and other sources.

sun and shade leaves

Plants may also respond 'tactically' to the radiation environment in which they develop, producing 'sun leaves' and 'shade leaves' within the canopy of a single plant. Sun leaves (and indeed, leaves on sun plants) are typically smaller, thicker, have more cells per unit area, denser veins, more densely packed chloroplasts and a greater dry weight per unit area of leaf. They are said to have a smaller *specific leaf area* (leaf area per unit leaf mass). Acclimation to shade typically involves increasing chlorophyll concentration and decreasing investment in the rest of the photosynthetic apparatus. This allows the leaf to maximise capture of light, but does not waste resource on a high photosynthetic capacity, which is not needed under shade conditions. In turn, this releases nitrogen for use by the upper leaves. However, these tactical manoeuvres take time. It is impossible for the plant to change its form fast enough to track the changes in intensity of radiation between a cloudy and a clear day. It can, however, change its rate of photosynthesis extremely rapidly, reacting even to the passing of a fleck of sunlight.

APPLICATION 3.1 Bioengineering crops for accelerated recovery from photoprotection

The responses of plants to changes in the quantity of light include the induction at high light intensities of photoprotective mechanisms that prevent the photosynthetic machinery from getting ‘overexcited’ and risking the generation of damaging oxidising radicals, instead dissipating excess light as heat. However, when intensities return to harmless levels, there is typically a delay before these protective mechanisms are fully switched off, such that rates of photosynthesis at these times are lower than they might otherwise be. Some calculations suggest that this could cost field crops as much as 20% of their potential yield (Kromdijk *et al.*, 2016). It would therefore clearly be valuable if that switching off of the protective mechanisms could be speeded up. Bioengineering (the insertion of new or altered genes into a plant) offers the opportunity of applying our understanding of the physiology of photoprotection to effect such an accelerated response. Results are shown in Figure 3.5 for a study in which variants of three different genes known to be instrumental in the operation of the mechanism were selected for increased expression levels, following a screen of seedlings of the model plant *Arabidopsis thaliana*. These variants were then inserted into tobacco plants, *Nicotiana tabacum*, itself used as a model for crop plants in general, since the photoprotective mechanism being altered is common to all plants.

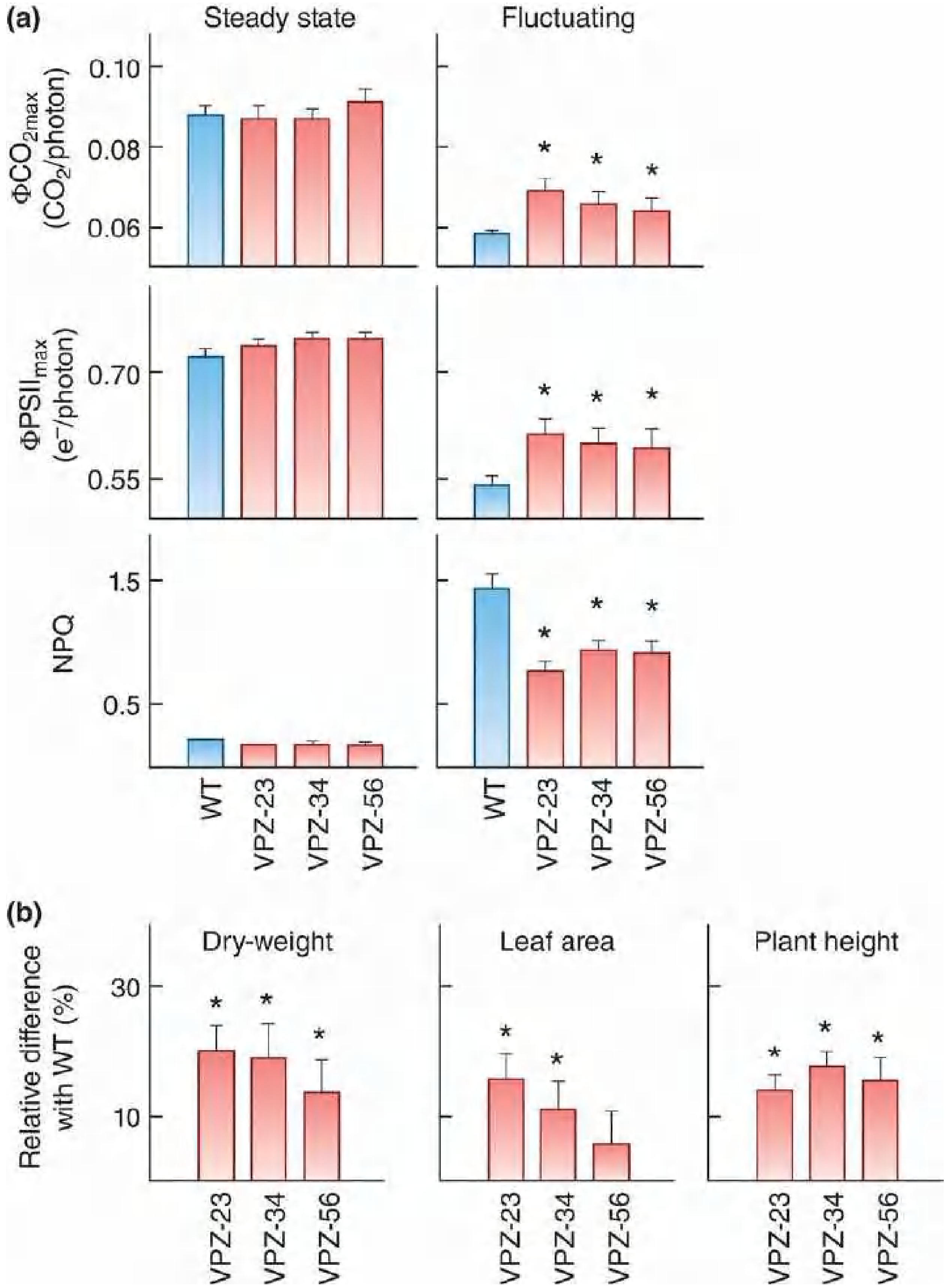

(a)
Steady state
Fluctuating
ΦCO2max (CO2/photon)
0.10
0.08
0.06
ΦPSIImax (e⁻/photon)
0.70
0.55
NPQ
1.5
0.5
WT
VPZ-23
VPZ-34
VPZ-56
*
(b)
Relative difference with WT (%)
Dry-weight
Leaf area
Plant height
30
10
VPZ-23
VPZ-34
VPZ-56

Figure 3.5 Bioengineering of photoprotection can improve crop plant performance. (a) To the left, a comparison for two measures of photosynthetic efficiency (of CO_2 uptake and of electron transport) and of the rate of harmlessly dissipating excess light as heat – the rate of 'quenching' of chlorophyll fluorescence (NPQ) – at steady levels of light, between wild type (WT) *Arabidopsis* plants and three strains bioengineered to switch off photoprotection more rapidly. There were no differences. To the right, a similar comparison but with fluctuating light levels. The bioengineered strains were all significantly more efficient in photosynthesis than the wild type because fluorescence was dampened down more rapidly. (b) The consequences for the bioengineered plants in terms of weight, leaf area and plant height, following 22 days of growth in the field. All strains grew better. In both (a) and (b), bars are SEs and * indicates a significant difference between bioengineered lines and the wild type ($P < 0.05$)

Source: After Kromdijk *et al.* (2016).

When the supply of light was constant, all three types of bioengineered plant behaved similarly to wild type plants in terms of photosynthetic efficiency and the harmless dissipation ('quenching') of excess light as heat (Figure 3.5a, left). But in the field, most leaves experience continually fluctuating light due to clouds and intermittent shading from the leaves above. It is notable, therefore, that in the fluctuating regime, photosynthetic efficiency was higher in the bioengineered plants than the wild types, and their overall level of quenching was lower, because it was compressed into a shorter period (Figure 3.5a, right). As a result, the bioengineered plants grew much better than the wild types (Figure 3.5b). Bioengineering of any sort must always be applied with caution, but these results do hold out the prospect of significant increases in yield for a wide variety of crops, since this process is common to all land plants.

pigment variation in aquatic species

In aquatic habitats, much of the variation between species is accounted for by differences in photosynthetic pigments, which contribute significantly to the precise wavelengths of radiation that can be utilised. Of the three types of pigment – chlorophylls, carotenoids and biliproteins – all photosynthetic plants contain the first two, but many algae also contain biliproteins; and within the chlorophylls, all higher plants have chlorophyll a and b, but many algae have only chlorophyll a and some have chlorophyll a and c. These different forms of chlorophyll all have slightly different absorption spectra, so that in combination, the plant or alga can trap more light. We see an example of variation in the nature of light with the concentration of dissolved organic matter in lake water, and the consequences of this for the photosynthetic microphytoplankton living there, in Figure 3.6. Of two lakes in north-western Patagonia, Argentina, one, Lake Morenito, had lower concentrations of dissolved organic matter, leading to 'greener' light (Figure 3.6a) and hence to higher densities of cryptophyte algae (Figure 3.6b). Cryptophytes have a unique combination of pigments – chlorophylls a and c, but also the carotenoid alloxanthin and one of two biliproteins – allowing them to function effectively in that range. The other, Lake Escondido, with yellower light, had a microphytoplankton community dominated by chrysophytes ('golden algae'), which lack these biliproteins.

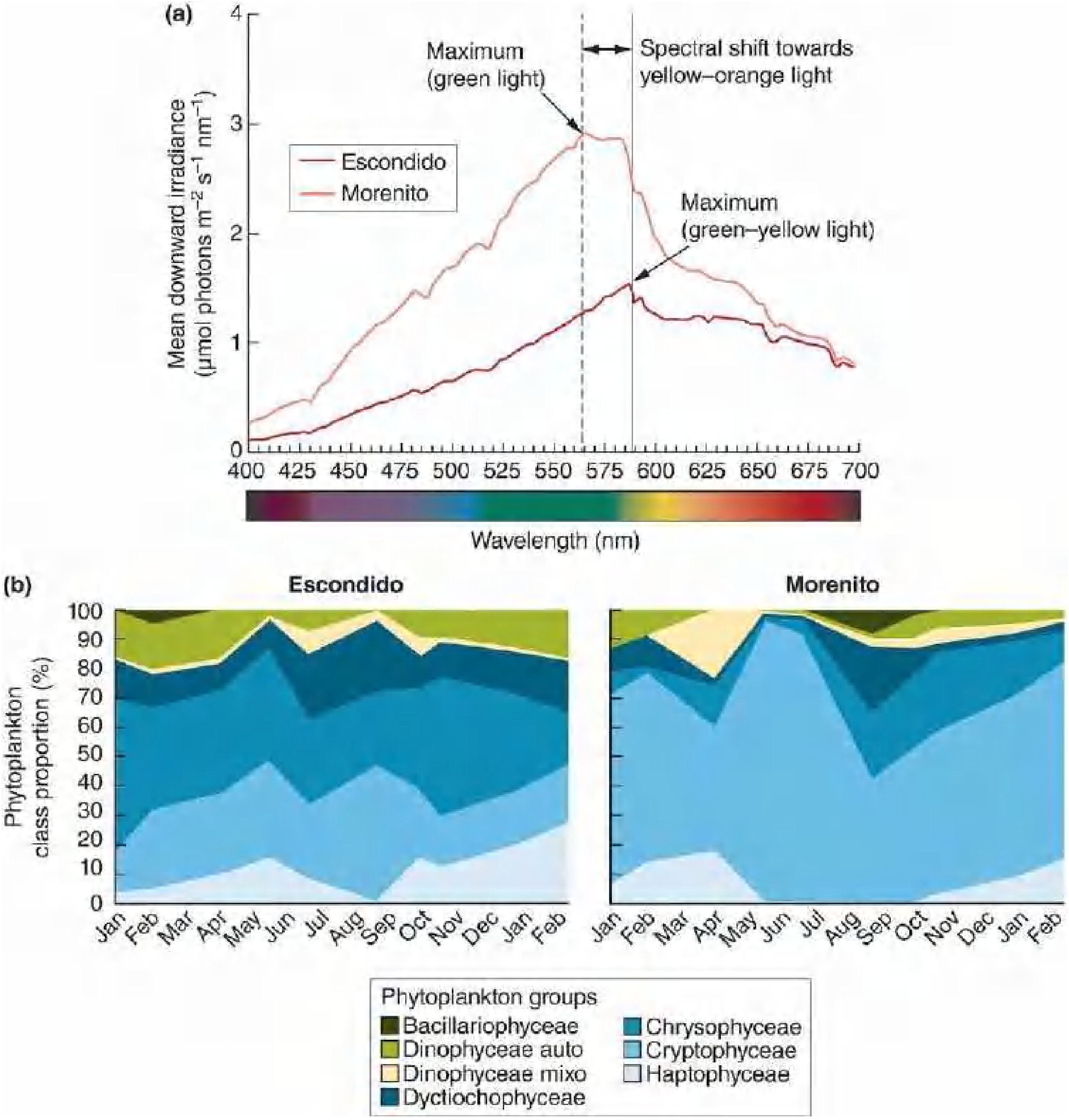


Figure 3.6 Variation in light quality in lakes can give rise to different communities of photosynthesisers. (a) The mean spectrum (wavelength distribution) of downward irradiance in the water columns of two lakes in Argentina: Lake Escondido and Lake Morenito. (b) As a result of the greener light in Lake Morenito, it supported a phytoplankton community with a higher proportion of cryptophytes that have photosynthetic pigments enabling them to function effectively at such wavelengths.

Source: After Gerea *et al*. (2017).

3.2.2 Net photosynthesis

the compensation point

Photosynthesis is measured as CO_2 uptake, which is roughly proportional to the amount of organic matter formed by the plant or alga, and in turn to the energy captured in that organic matter. However, it is often more important to consider, and very much easier to measure, the net gain. Net photosynthesis is the increase (or decrease) in dry matter that results from the difference between gross photosynthesis and the contemporary losses due to respiration. The *light compensation point* is the intensity of PAR at which the gain from gross photosynthesis exactly balances these respiratory losses. Hence, net photosynthesis increases with the intensity of PAR, and is negative below the compensation point (including, of course, in darkness), when respiration exceeds photosynthesis. An example of the way plant characteristics combine to determine the compensation point is shown in Figure 3.7. As seedlings of sugar maple (*Acer saccharum*) grow, the proportion of non-photosynthetically active stem tissue to photosynthetically active leaf tissue increases, and the degree of self-shading (upper leaves denying light to lower leaves) increases. Hence, at a given light level, taller (older) plants photosynthesise less effectively and their net carbon gain per gram of plant tissue declines (Figure 3.7a). As a consequence of this, while growing taller is often advantageous (for example, in terms of intercepting light before some other plant does) the compensation point increases in taller plants – they require more light to balance their respiratory needs (Figure 3.7b).

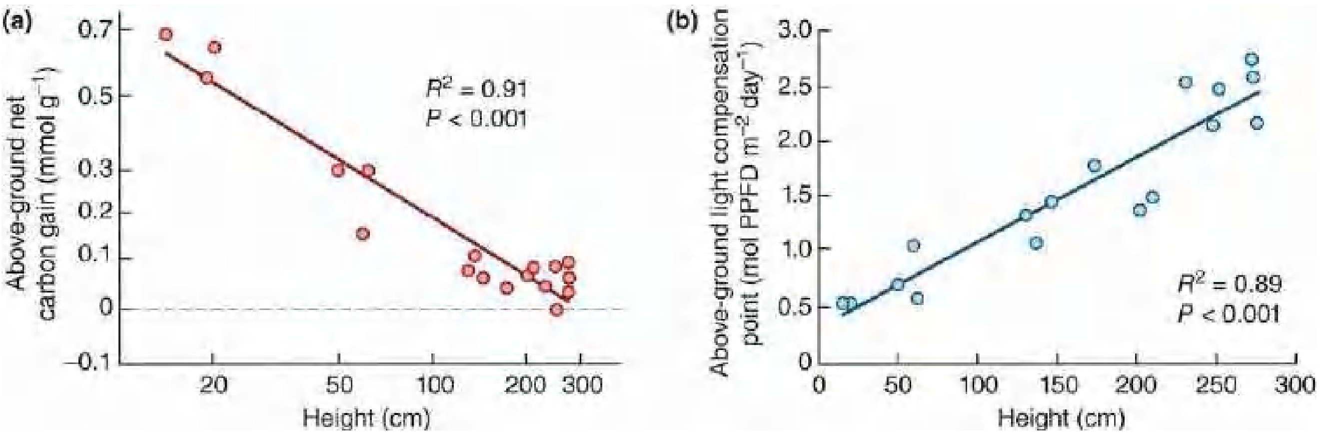


Figure 3.7 **The compensation point is higher in taller plants.** (a) The decline in the estimated above-ground net daily carbon gain (effectiveness of photosynthesis) with height for seedlings of sugar maple, *Acer saccharum*. (b) The consequent increase in the above-ground light compensation point for seedlings with height. PPFD, photosynthetic photon flux density.

Source: After Sendall *et al.* (2015).

photosynthetic capacity

There is nearly a 100-fold variation in the *photosynthetic capacity* of leaves (Mooney & Gulmon, 1979). This is the rate of photosynthesis when incident radiation is saturating, temperature is optimal, relative humidity is high, and CO_2 and oxygen concentrations are normal. When the leaves of different species are compared under these ideal conditions, the ones with the highest photosynthetic capacity are generally those from environments where nutrients, water and radiation are seldom limiting (at least during the growing season). These include many agricultural crops and their weeds, and also the primary colonisers of newly available rich habitats, for example volcanic ash, or the grass, *Echinochloa polystachya*, on Amazonian floodplains (Piedade *et al.*, 1991). Species from resource-poor environments (e.g. shade plants, desert perennials, heathland species) usually have low photosynthetic capacity – even when abundant resources are provided. Such patterns can be understood by noting that photosynthetic capacity, like all capacity, requires investment, and that the investment by plants in structures and processes that enhance photosynthetic capacity is only likely to be repaid if ample opportunity exists for that capacity to be utilised.

Needless to say, ideal conditions in which plants may achieve their photosynthetic capacity are rare. In practice, the rate at which photosynthesis actually proceeds is limited by conditions (e.g. temperature) and by the availability of resources other than radiant energy. In particular, the photosynthetic capacity of leaves is highly correlated with leaf nitrogen content, at least at a broad scale (Figure 3.8), though as we shall see below, this may not be the whole story. The rate of photosynthesis also increases with the intensity of PAR, but in most species ('C_3 plants', see later) it reaches a plateau at intensities of radiation well below that of full solar radiation.

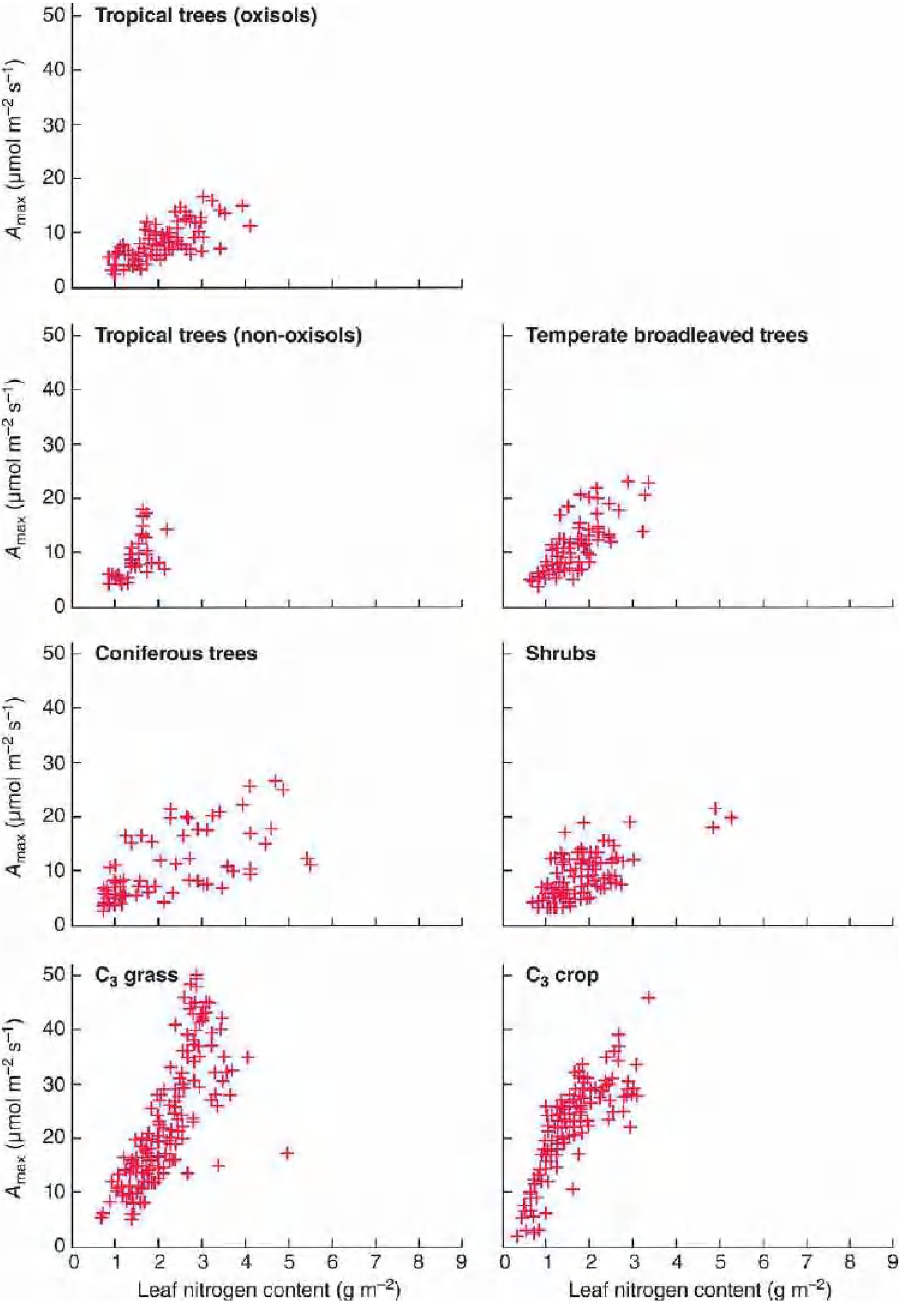

Tropical trees (oxisols)
Tropical trees (non-oxisols)
Temperate broadleaved trees
Coniferous trees
Shrubs
C$_3$ grass
C$_3$ crop
A_{max} (μmol m^{-2} s^{-1})
0
10
20
30
40
50
0
1
2
3
4
5
6
7
8
9
Leaf nitrogen content (g m^{-2})

Figure 3.8 **Photosynthetic capacity increases with leaf nitrogen content.** The positive relationship of photosynthetic capacity (A_{max}) with the leaf nitrogen content of different C_3 plant species, arranged in groups, as indicated. (Oxisols are nutrient-poor soils dominated by iron oxides, quartz and highly weathered clay minerals such as kaolinite.)

Source: Modified from Kattge *et al.* (2009).

An example of what can drive variation in photosynthetic capacity is provided by a study of tree species from the forests of Rwanda that were either characteristic of open, clear ground ('pioneer' species) or of closed, crowded canopies ('climax' species) (see Section 15.4 for a discussion of community successions). The pioneer trees, adapted to life in the sun, had much greater photosynthetic capacity, especially in the case of leaves from the top of the canopy (Figure 3.9a, b). However, contrary to the general pattern described above, this was not simply the result of a greater investment in those leaves in terms of either nitrogen or phosphorus (Figure 3.9c, d). Rather, resources in the shade leaves of climax species were directed towards improvements in the ability to harvest what light was available, by having more chlorophyll (Figure 3.9e), whereas resources in the sun leaves of pioneer species were directed towards increased carboxylation capacity – being able to make best use of high light levels (Figures 3.9a, b). Overall there was a trade-off between photosynthetic carboxylation capacity and light-harvesting ability.

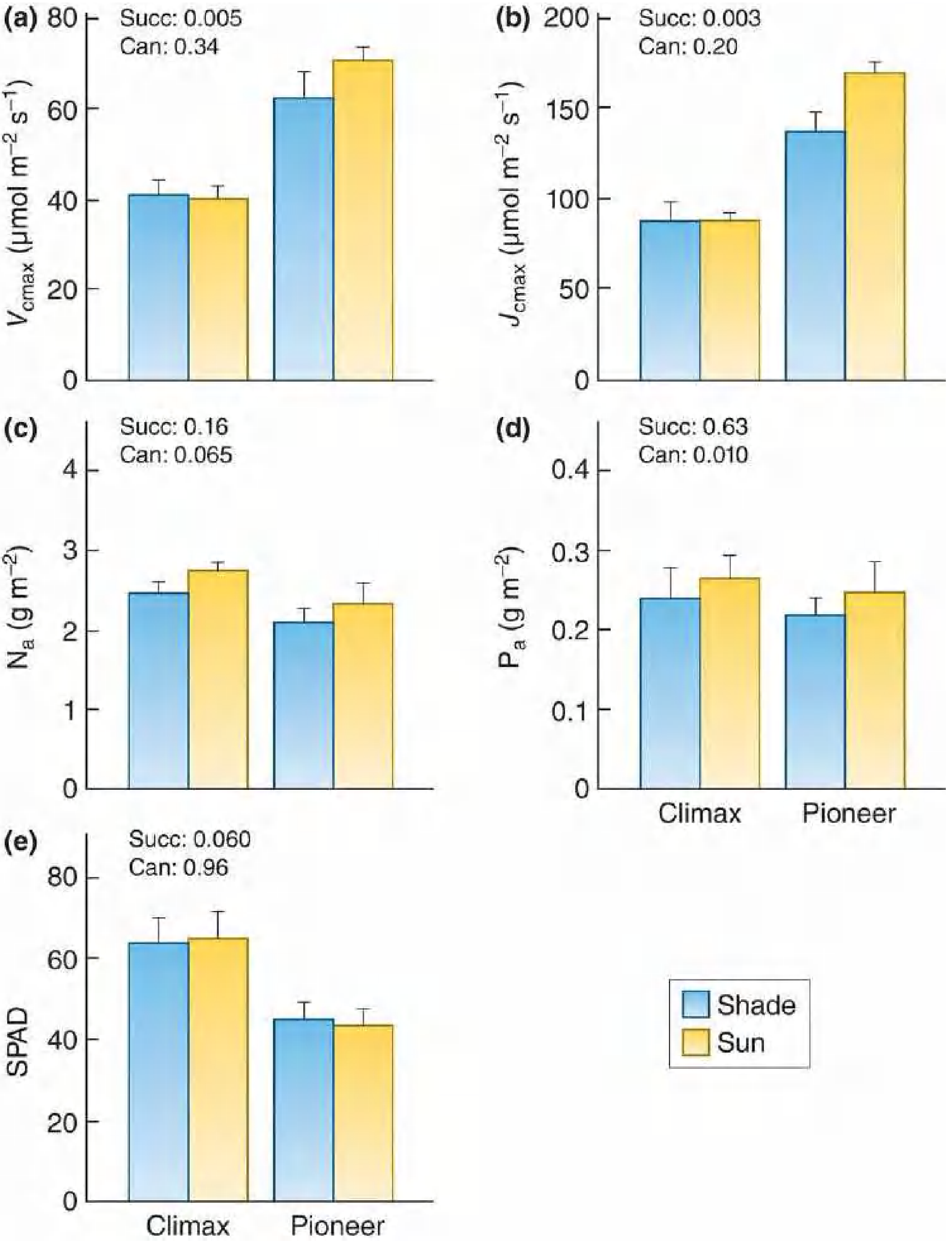

(a)
Succ: 0.005
Can: 0.34
V_{cmax} (μmol m^{-2} s^{-1})
0
20
40
60
80
(b)
Succ: 0.003
Can: 0.20
J_{cmax} (μmol m^{-2} s^{-1})
0
50
100
150
200
(c)
Succ: 0.16
Can: 0.065
N_a (g m^{-2})
0
1
2
3
4
(d)
Succ: 0.63
Can: 0.010
P_a (g m^{-2})
0
0.1
0.2
0.3
0.4
Climax
Pioneer
(e)
Succ: 0.060
Can: 0.96
SPAD
0
20
40
60
80
Climax
Pioneer
Shade
Sun

Figure 3.9 Sun and shade leaves and plants vary in their capacities and compositions. A range of comparisons of sun and shade leaves from trees from Rwandan forests characteristic either of open ground (pioneer species) or of closed, crowded canopies (climax species). Bars are SEs. Succ. and Can. in each case refer to *P* values indicating the significance in an analysis of variance of the factors ‘Succession’ (climax/pioneer) and ‘Canopy’ (sun/shade), respectively. (a) Carboxylation capacity. (b) Electron transport capacity. (c) Leaf nitrogen content. (d) Leaf phosphorus content. (e) SPAD, a proxy for leaf chlorophyll content.

Source: After Dusenge *et al*. (2015).

3.2.3 Sun and shade plants of an evergreen shrub

A number of the general points above are illustrated by a study of the evergreen shrub *Heteromeles arbutifolia*. This plant grows both in chaparral habitats in California, where shoots in the upper crown are consistently exposed to full sunlight and high temperatures, and also in shaded woodland habitats, where it receives around one-seventh as much radiation (Figure 3.10a). The leaves of sun plants have a lower specific leaf area – they are thicker and have a greater photosynthetic capacity (more chlorophyll and nitrogen) per unit leaf area than those of shade plants (Figure 3.10b). Sun-plant leaves are inclined at a much steeper angle to the horizontal, and therefore absorb the direct rays of the overhead summer sun over a wider leaf area than the more horizontal shade-plant leaves. The more angled leaves of sun plants, though, are also less likely than shade-plant leaves to shade other leaves of the same plant from the overhead rays of the summer sun (Figure 3.10c). But in winter, when the sun is much lower in the sky, it is the shade plants that are much less subject to this ‘self-shading’. As a result, the proportion of incident radiation intercepted per unit area of leaf is higher in shade than in sun plants year round: in summer because of the more horizontal leaves, and in winter because of the relative absence of self-shading.

(a)

A Sun plant
Early morning

C Sun plant
Midday

B Shade plant
Early morning

D Shade plant
Midday

(b)

	Sun		Shade		*P*
Leaf angle (degrees)	71.3	(16.3)	5.3	(4.3)	<0.01
Leaf blade thickness (μm)	462.5	(10.9)	292.4	(9.5)	<0.01
Photosynthetic capacity, area basis (μmol CO_2 m^{-2} s^{-1})	14.1	(2.0)	9.0	(1.7)	<0.01
Chlorophyll content, area basis (mg m^{-2})	280.5	(15.3)	226.7	(14.0)	<0.01
Leaf nitrogen content, area basis (g m^{-2})	1.97	(0.25)	1.71	(0.21)	<0.05

(c)

	Sun plants		Shade plants	
	Summer	Winter	Summer	Winter
Fraction self-shaded	0.22^{a}	0.42^{b}	0.47^{b}	0.11^{a}
Display efficiency	0.33^{a}	$0.38^{a,b}$	0.41^{b}	0.43^{b}
Absorption efficiency	0.28^{a}	0.44^{b}	0.55^{c}	0.53^{c}

Figure 3.10 Variations in the behaviour and properties of sun and shade leaves of an evergreen shrub. (a) Computer reconstructions of stems of typical sun (A, C) and shade (B, D) plants of the evergreen shrub *Heteromeles arbutifolia*, viewed along the path of the sun's rays in the early morning (A, B) and at midday (C, D). Darker tones represent parts of leaves shaded by other leaves of the same plant. Scale bars = 4 cm. (b) Observed differences in the leaves of sun and shade plants. Standard deviations are given in parentheses; the significance of differences is given following the analysis of variance. (c) Consequent whole-plant properties of sun and shade plants. Letter codes indicate groups that differed significantly in analyses of variance ($P < 0.05$).

Source: After Valladares & Pearcy (1998).

The properties of whole plants of *H. arbutifolia*, then, reflect both plant architecture and the morphologies and physiologies of individual leaves. The efficiency of light absorption per unit of biomass is massively greater for shade than for sun plants (Figure 3.10c). Despite receiving only one-seventh of the radiation of sun plants, shade plants reduce the differential in their daily rate of carbon gain from photosynthesis to only a half. They successfully counterbalance their reduced photosynthetic capacity at the leaf level with enhanced light-harvesting ability at the whole-plant level. The sun plants, on the other hand, can be seen as striking a compromise between maximising whole-plant photosynthesis while avoiding photoinhibition and overheating of individual leaves.

3.3 Water

Water is a critical resource. Hydration is a necessary condition for metabolic reactions to proceed, and because no organism is completely watertight, its water content needs continual replenishment. Most terrestrial animals drink free water and also generate some from the metabolism of food and body materials. There are extreme cases in which animals of arid zones may obtain all their water from their food.

3.3.1 Photosynthesis or water conservation? Strategic and tactical solutions

stomatal opening

For plants, in terrestrial habitats especially, it is not sensible to consider radiation as a resource independently of water. Intercepted radiation does not result in photosynthesis unless there is CO_2 available, and the prime route of entry of CO_2 is through open stomata (see Section 3.4). But if the stomata are open to the air, water will evaporate through them. Indeed, the volume of water that becomes incorporated in higher plants during growth is infinitesimal in comparison to the volume that flows through the plant in the transpiration stream (in through the roots, out through the stomata). If water is lost faster than it can be gained, the leaf (and the plant) will sooner or later wilt and eventually die. In most terrestrial communities, water is, at least sometimes, in short supply. The question therefore arises: should a plant conserve water at the expense of present photosynthesis, or maximise photosynthesis at the risk of running out of water? Once again, we meet the problem of whether the optimal solution involves a strict strategy or the ability to make tactical responses. There are good examples of both solutions and also compromises.

short active interludes in a dormant life

Perhaps the most obvious strategy that plants may adopt is to have a short life and high photosynthetic activity during periods when water is abundant, but remain dormant as seeds during the rest of the year, neither photosynthesising nor transpiring. Many desert annuals do this, as do annual weeds and most annual crop plants.

leaf appearance and structure

Second, plants with long lives may produce leaves during periods when water is abundant and shed them during droughts (e.g. many species of Acacia). Or they may change the nature of their leaves. Some desert shrubs in Israel (e.g. *Teucrium polium*) bear finely divided, thin-cuticled leaves during the season when soil water is freely available. These are then replaced by undivided, small, thick-cuticled leaves in more drought-prone seasons, which in turn fall and may leave only green spines or thorns (Orshan, 1963): a sequential polymorphism through the season, with each leaf morph being replaced in turn by a less photosynthetically active but more watertight structure.

Next, leaves may be produced that are long lived, transpire only slowly and tolerate a water deficit, but which are unable to photosynthesise rapidly even when water is abundant (e.g. evergreen desert shrubs). Structural features such as hairs, sunken stomata and the restriction of stomata to specialised areas on the lower surface of a leaf slow down water loss. But these same morphological features reduce the rate of entry of CO_2. Waxy and hairy leaf surfaces may, however, reflect a greater proportion of radiation that is not in the PAR range and so keep the leaf temperature down and reduce water loss.

physiological strategies

Finally, some groups of plants have evolved particular physiologies: C_4 and Crassulacean acid metabolism (CAM). We consider these in more detail in Sections 3.4.1–3.4.3. Here, we simply note that plants with 'normal' (i.e. C_3) photosynthesis are wasteful of water compared with plants that possess the modified C_4 and CAM physiologies. The water-use efficiency of C_4 plants (the amount of carbon fixed per unit of water transpired) may be double that of C_3 plants.

tactical changes in stomatal conductance

The major tactical control of the rates of both photosynthesis and water loss is through changes in stomatal 'conductance'. These may occur rapidly during the course of a day and allow a very rapid response to immediate water shortages, such that rhythms of stomatal opening may ensure that the above-ground parts of the plant remain more or less watertight except during controlled periods of active photosynthesis. Stomatal movement may even be triggered directly by conditions at the leaf surface itself – the plant then responds to desiccating conditions at the very site, and at the same time, as the conditions are first sensed.

coexisting alternative strategies in Australian savannas

The viability of alternative strategies to solve a common problem is nicely illustrated by the trees of seasonally dry tropical forests and woodlands (Eamus, 1999). Communities of this type are found naturally in Africa, the Americas, Australia and India, and as a result of human interference elsewhere in Asia. But whereas, for example, the savannas of Africa and India are dominated by deciduous species, and the Llanos of South America are dominated by evergreens, the savannas of Australia are occupied by roughly equal numbers of species from four groups (Figure 3.11a): evergreens (a full canopy all year), deciduous species (losing all leaves for at least one and usually two to four months each year), semideciduous species (losing around 50% or more of their leaves each year) and brevideciduous species (losing only about 20% of their leaves). At the ends of this continuum, the deciduous species avoid drought in the dry season (April–November in Australia) as a result of their vastly reduced rates of transpiration (Figure 3.11b), but make no net photosynthate at all for around three months, whereas the evergreens maintain a positive carbon balance throughout the year (Figure 3.11c). The alternative, contrasting strategies are clearly sufficiently viable for them to coexist in Australia. Why this is not equally true elsewhere is not known.

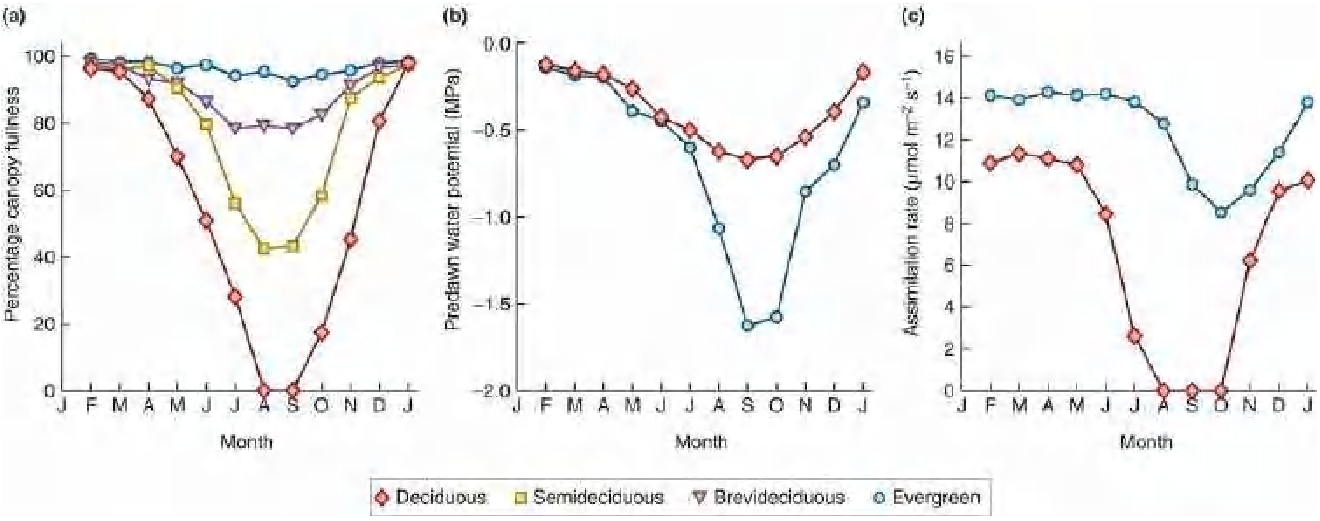


Figure 3.11 Alternative strategies for combining photosynthesis and water conservation among trees in Australian savannas. (a) Percentage canopy fullness for deciduous (red), semideciduous (yellow), brevideciduous (purple) and evergreen (blue) trees in Australian savannas throughout the year. (Note that the southern hemisphere dry season runs from around April to November.) (b) Susceptibility to drought as measured by increasingly negative values of 'predawn water potential' for deciduous and evergreen trees. (c) Net photosynthesis as measured by the carbon assimilation rate for deciduous and evergreen trees.

Source: After Eamus (1999).

3.3.2 Roots as water foragers

Counteracting loss is, of course, only one side of the balance sheet. For most terrestrial plants, the main source of water is the soil and they gain access to it through a root system. We proceed here (and in Section 3.5, on plant nutrient resources) on the basis of plants simply having 'roots'. In fact, most plants do not have simple, plant-only roots – they have mycorrhizae: associations of fungal and root tissue in which both partners are crucial to the resource-gathering properties of the whole. Mycorrhizae, and the respective roles of the plants and the fungi, are discussed in Chapter 13.

field capacity and the permanent wilting point

Water enters the soil as rain or melting snow and forms a reservoir in the pores between soil particles. What happens to it then depends on the size of the pores, which may hold it by capillary forces against gravity (Figure 3.12). If the pores are wide, as in a sandy soil, much of the water will drain away until it reaches some impediment and accumulates as a rising water table or finds its way into streams or rivers. The water held by soil pores against the force of gravity is called the *field capacity* of the soil. This is the upper limit of the water that a freely drained soil will retain. However, not all the water retained by soil is available to plants, since they must extract it from those soil pores against the surface tension holding it there, and their ability to do so depends on the structure of their root systems. Hence, there is also a lower limit to the water that can be used in plant growth, determined by the particular plant species present, known as the *permanent wilting point* – the soil water content at which plants wilt and are unable to recover. The permanent wilting point does not differ much between the plant species of mesic environments (those with a moderate amount of water) or between species of crop plants, but many species native to arid regions have very low permanent wilting points as a result of root systems that allow them to extract significantly more water from the soil, and of leaf morphological adaptations, discussed previously, that give them better water-holding capacities.

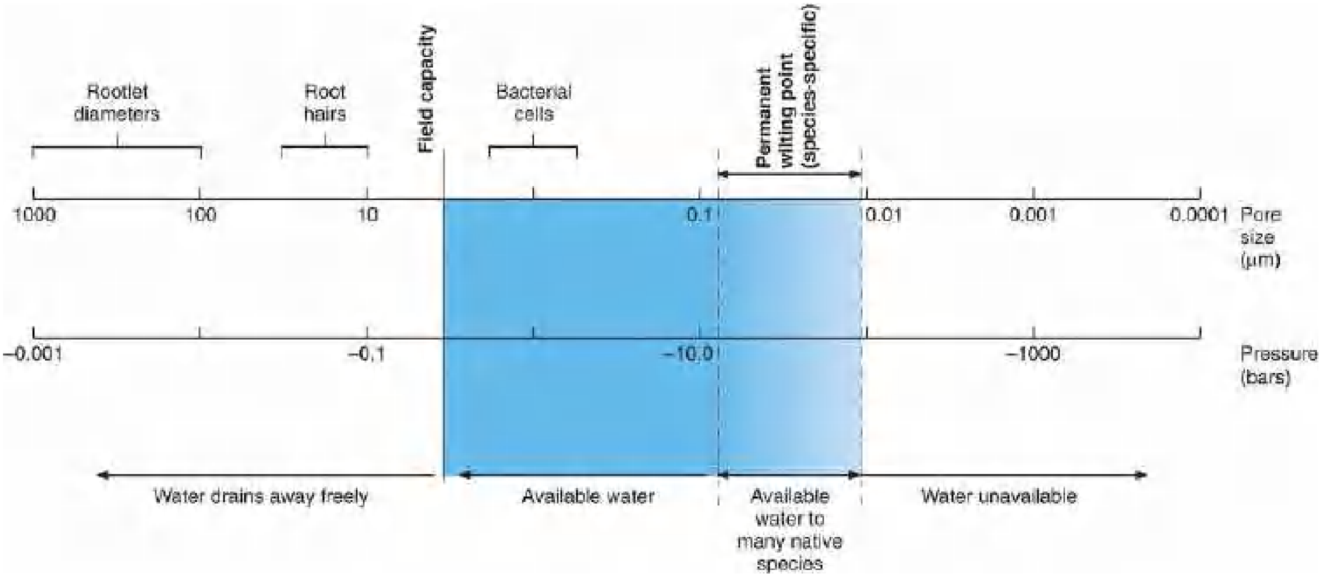


Figure 3.12 Field capacity and the permanent wilting point in soil in relation to pore size and pressure. The status of water in the soil, showing the relationship between the diameter of soil pores that remain water-filled and the pressure created by the capillary action of those pores that opposes the tendency of water to drain away under the force of gravity. Pressure values are negative because they describe the process of suction. The size of water-filled pores may be compared in the figure with the sizes of rootlets, root hairs and bacterial cells. Note that for most species of crop plant the permanent wilting point is at approximately −15 bars, but in many other species it reaches −80 bars, depending on their ability to extract water from the narrowest pores.

roots and the dynamics of water depletion zones

As a root withdraws water from the soil pores at the root's surface, it creates water-depletion zones around it – another example of the RDZs described in Section 3.2.1. These determine gradients of water potential between the interconnected soil pores. Water flows along the gradient into the depleted zones, supplying further water to the root, but this simple process is made much more complex because the more the soil around the roots is depleted of water, the more resistance there is to water flow. Thus, as the root starts to withdraw water from the soil, the first water that it obtains is from the wider pores because they hold the water with weaker capillary forces. This leaves only the narrower, more tortuous pathways, and so the resistance to water flow increases. Thus, when the root draws water from the soil very rapidly, the RDZ may become very sharply defined, because water can move across its boundary only slowly. For this reason, rapidly transpiring plants may wilt in a soil that contains abundant water.

roots as foragers

Water that arrives on a soil surface does not distribute itself evenly down the soil profile. Instead, it tends to bring the surface layer to field capacity, with further rain extending this layer deeper and deeper. This means that different parts of the same plant root system may encounter water held with quite different forces. Similar variations can occur as a result of heterogeneities in soil type – clay soils with small pores can hold far more water than sandy soils with large pores. As a root passes through such heterogeneous soil (and all soils are heterogeneous seen from a 'root's-eye view'), it typically responds by branching freely in zones that supply resources, and scarcely branching at all in less rewarding patches (Figure 3.13a). That it can do so depends on the individual rootlet's ability to react on an extremely local scale to the conditions that it meets. Strategic differences in developmental programmes can be recognised between the roots of different species (Figure 3.13b), but it is the ability of root systems to override strict programmes and be opportunistic, depending both on local conditions and their overall level of resource availability, that makes them effective exploiters of the soil (de Kroon *et al*., 2009).

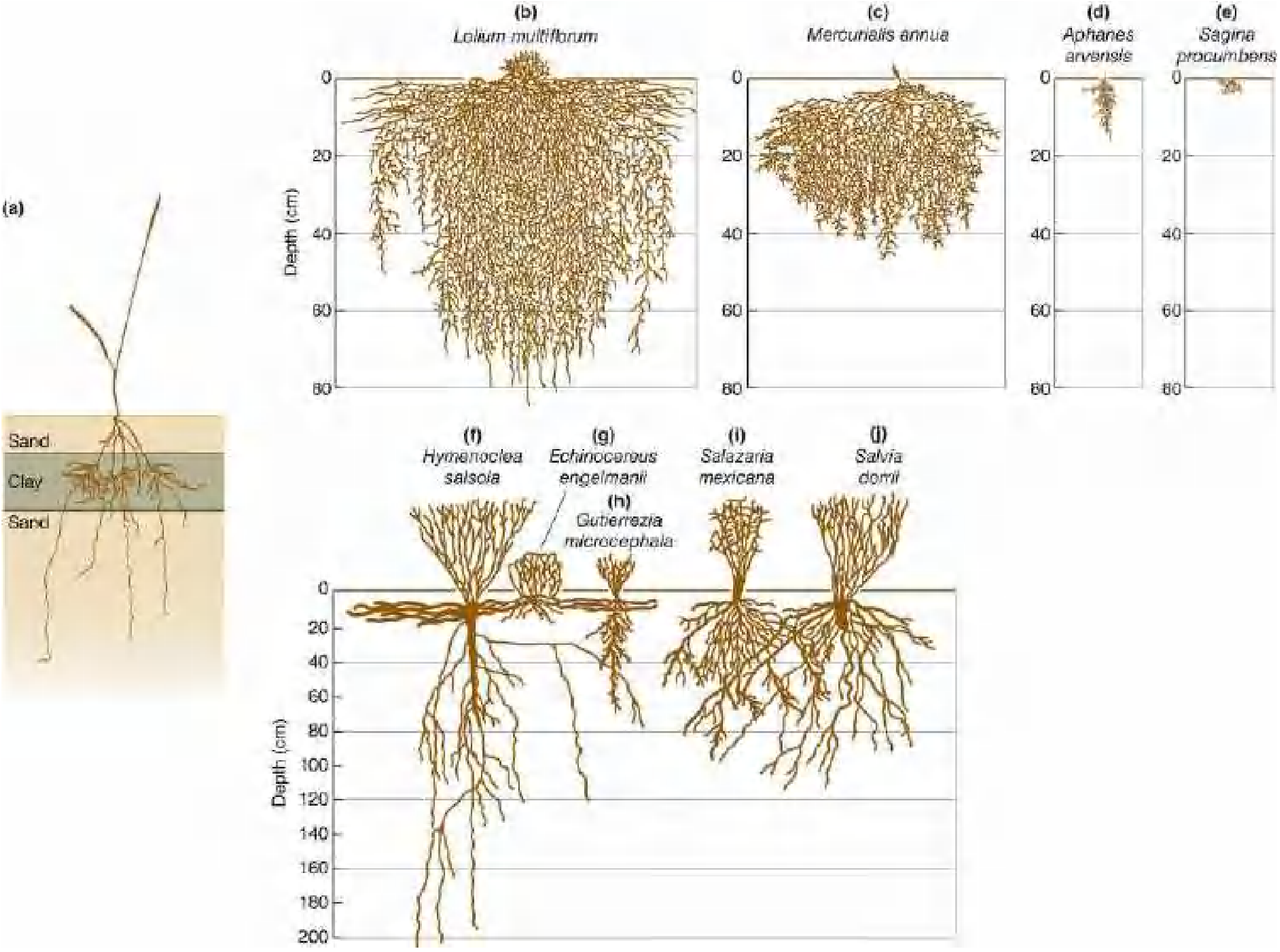


Figure 3.13 **Roots as foragers.** (a) The root system developed by a plant of wheat grown through a sandy soil containing a layer of clay. Note the responsiveness of root development to the localised environment that it encounters. (b–j) Profiles of root systems of plants from contrasting environments. (b–e) Northern temperate species of open ground: (b) *Lolium multiflorum*, an annual grass; (c) *Mercurialis annua*, an annual weed; and (d) *Aphanes arvensis* and (e) *Sagina procumbens*, both ephemeral weeds. (f–j) Desert shrub and semishrub species, Mid Hills, eastern Mojave Desert, California.

Source: (a) Courtesy of J.V. Lake. (b–e) From Fitter (1991). (f–j) Redrawn from a variety of sources.

The root system that a plant establishes early in its life can determine its responsiveness to future events. Where most water is received as occasional showers on a dry substrate, a seedling that puts its early energy into a deep taproot will gain little from subsequent showers, but in an environment in which heavy rains fill a soil reservoir to depth in the spring, followed by a long period of drought, that taproot may guarantee continual access to water. Indeed, it seems that the placement of roots with respect to water and especially nutrient availability is most important in the earlier stages of a plant's life. Later there is much greater reliance on stored resources in overcoming local or temporary shortages (de Kroon *et al.*, 2009).

3.4 Carbon dioxide

the rise in global levels

The CO_2 used in photosynthesis is obtained almost entirely from the atmosphere, where its concentration has risen from approximately 280 $\mu l\, l^{-1}$ in 1750 to about 411 $\mu l\, l^{-1}$ as we write (2018) and is still increasing by 0.4–0.5% per year (see Figure 21.22).

variations beneath a canopy

Concentrations also vary spatially. In a terrestrial community, the flux of CO_2 at night is upwards, from the soil and vegetation to the atmosphere; on sunny days above a photosynthesising canopy, there is a downward flux. Nonetheless, above a vegetation canopy, the air becomes rapidly mixed. The situation is quite different, however, within and beneath canopies. Changes in CO_2 concentration in the air within a mixed deciduous forest in summer, in Sapporo, Japan, are illustrated in Figure 3.14. Throughout the day, there was a gradient of decreasing concentration from the ground (0.5 m) to the upper canopy (24 m), reflecting the shifting balance between its production through respiration and its utilisation in photosynthesis. Indeed, an earlier study by Bazzaz and Williams (1991) had recorded levels as high as 1800 $\mu l\, l^{-1}$ near the ground, as a result of rapid decomposition of litter and soil organic matter. The gradient was steepest, and concentrations generally higher, during the night than during the day, presumably because the respiration of decomposers, especially, is relatively insensitive to the diurnal cycle. In winter, in the absence of leaves, there was no detectable variation in CO_2 concentration with height.

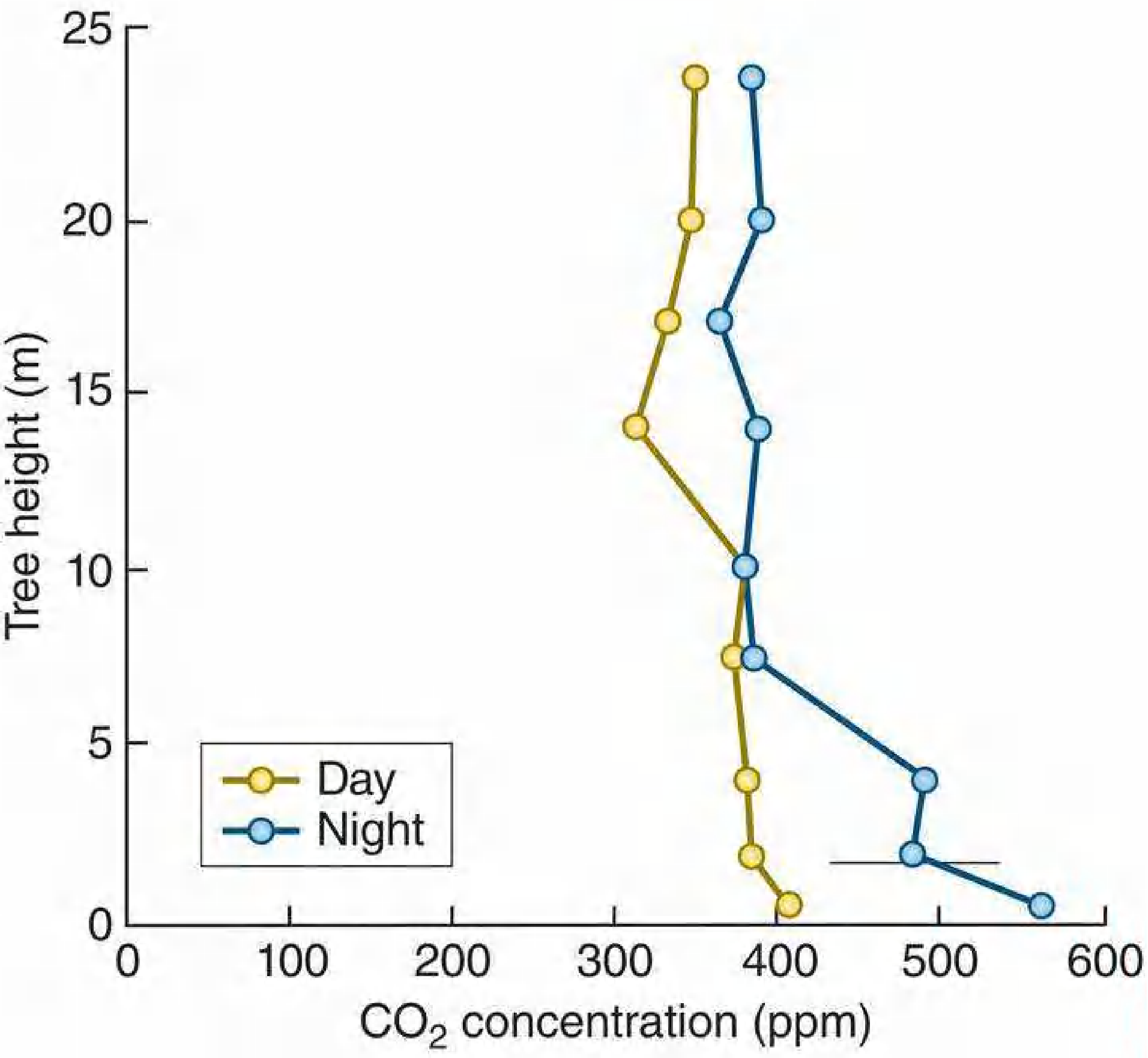


Figure 3.14 **The change in atmospheric CO_2concentration with height in a forest canopy** in Sapporo, Japan at night and in the day. The bar is the maximum SE.

Source: After Koike *et al*. (2001).

That CO_2 concentrations vary so widely within vegetation means that plants growing in different parts of a forest will experience quite different CO_2 environments. Indeed, the lower leaves on a forest shrub will usually experience higher CO_2 concentrations than its upper leaves, and seedlings will live in environments richer in CO_2 than mature trees.

variations in aquatic habitats

In aquatic environments, variations in CO_2 concentration can be just as striking, especially when water mixing is limited, for example during the summer 'stratification' of lakes, with layers of warm water towards the surface and colder layers beneath. Some examples from a study in Estonia are shown in Figure 3.15. At one extreme was the shallow Lake Äntu Sinijärv, which is supersaturated with CO_2 (CO_2 concentration higher than would result from equilibration with atmospheric CO_2) as a result of the high concentrations of bicarbonate ions in the water flowing into it. Here, there was usually virtually no vertical stratification of CO_2. Lake Saadjärv was deeper and thermally stratified and also had very high CO_2 concentrations, but in this case there was strong CO_2 stratification in the deeper layers. And finally, Lake Peipsi was a very large lake compared with the others (3555 km^2 compared with <10 km^2 for the others), similar in depth to Lake Äntu Sinijärv, but with very much lower CO_2 concentrations overall. In this case, vertical stratification of CO_2 concentrations led consistently to levels in the upper layers where the concentrations were lower than saturation, such that the lake was a sink for atmospheric CO_2, whereas the other two lakes were net CO_2 emitters.

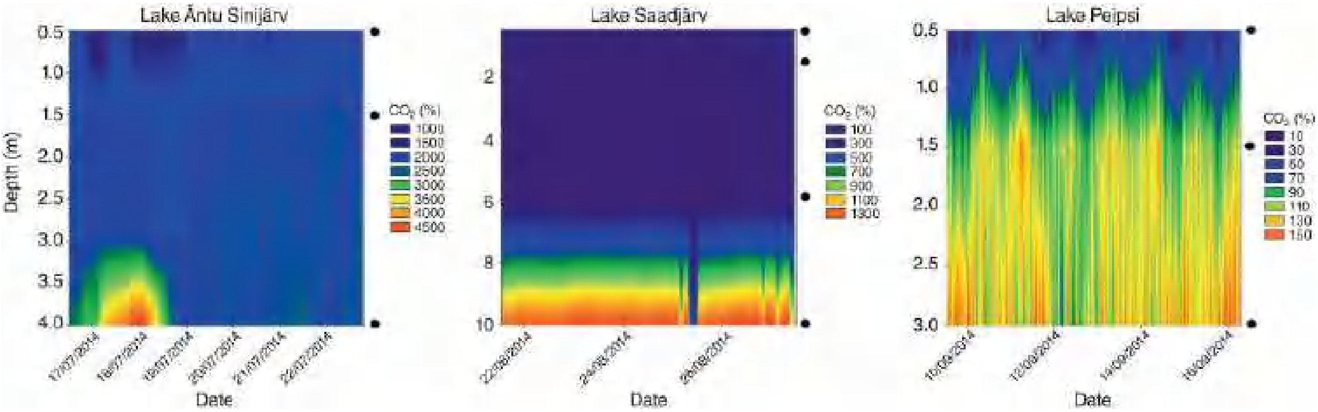


Figure 3.15 **Concentrations of CO_2vary, variably, with depth in Estonian lakes.** The profiles with depth of CO_2 concentration over a number of days (different in each case) in three lakes in Estonia, as indicated. Note that the colour-coding varies between the lakes to reflect their different concentration ranges, and that their depths are different.

Source: After Laas *et al.* (2016).

In aquatic habitats, especially under alkaline (high pH) conditions, dissolved CO_2 tends to react with water to form carbonic acid, which in turn ionises, such that 50% or more of inorganic carbon in water may be in the form of bicarbonate ions. Indeed, the Estonian lakes highlight how, in many cases, overall concentrations of CO_2 may be highly supersaturated. This may seem to suggest that aquatic plants will only rarely be limited by the availability of CO_2, but in fact they commonly are limited, due to the low rates of CO_2 diffusion in water, and around half of aquatic plants are able to use bicarbonate ions as an alternative source of dissolved organic carbon. However, since bicarbonate must ultimately be reconverted to CO_2 for photosynthesis, this is likely to be less useful as a source of inorganic carbon, and in practice, many plants will be limited in their photosynthetic rate by the availability of CO_2. We see an illustration of this in Figure 3.16. Ten species of aquatic plants, all capable of using bicarbonate as a source of CO_2, were grown in two culture conditions, both with the same overall concentration of dissolved organic carbon (0.85 mM). In one case (low-C) the water was in equilibrium with the surrounding air (saturated) and so the contribution of CO_2 itself to this was small (0.012 mM). But in the other case (high-C) the initial concentration, largely from bicarbonate ions, was much lower (0.40 mM) but gaseous CO_2 was continually added to the water, supersaturating it, and raising the overall concentration to the low-C level. All 10 species grew faster under the high-C conditions (Figure 3.16a), apparently as a result of elevated growth efficiencies at higher concentrations of CO_2, since the low-C plants were investing more in, for example, leaf nitrogen (Figure 3.16b), enabling them to make more of the limited CO_2 resource available to them. Even for these bicarbonate users, bicarbonate is good but CO_2 is better.

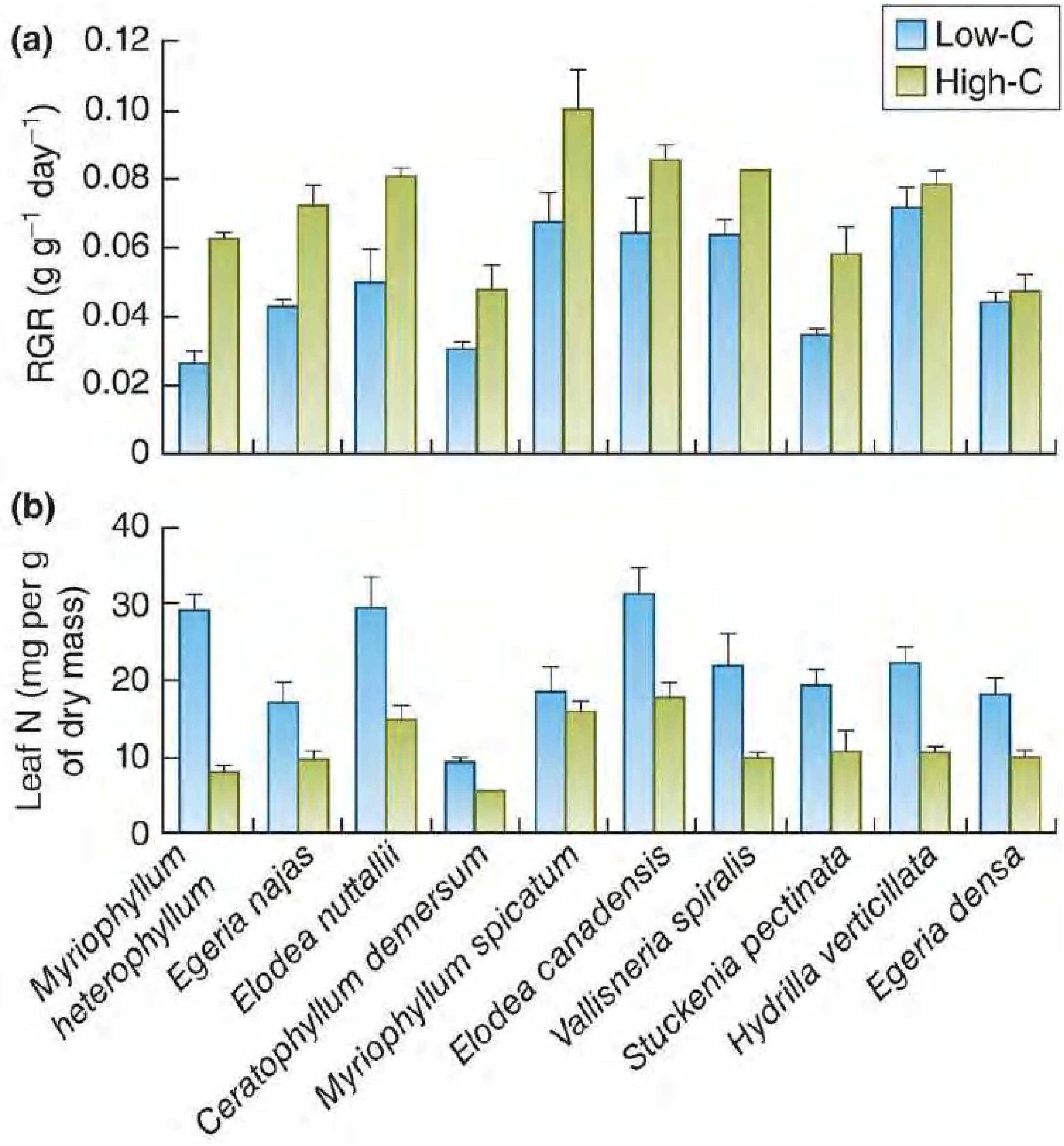


Figure 3.16 Aquatic plants may be limited in their photosynthetic ability by the availability of CO_2. (a) The relative growth rate (RGR, rate of growth per unit weight) for 10 species of aquatic plants, as indicated, when water was at equilibrium with the surrounding air with respect to CO_2 (low-C) such that the contribution of CO_2 to dissolved inorganic carbon (compared with bicarbonate) was relatively small, and when CO_2 was continually passed into the water (high-C) such that the contribution of CO_2 was large. In a two-way analysis of variance, the effects of species and treatment were both significant (respectively, $F = 11.6$, $P < 0.0001$ and $F = 52.9$, $P < 0.0001$). (b) The leaf nitrogen content of the same 10 species in the same treatments. Again, the effects of species and treatment were both significant (respectively, $F = 9.1$, $P < 0.0001$ and $F = 101.4$, $P < 0.0001$). Means and SEs are shown in both parts.

Source: After Hussner *et al*. (2016).

3.4.1 C_3, C_4 and CAM

These variations in CO_2 availability, along with associated variations in, for example, the difficulties of capturing CO_2 while avoiding the loss of water, have led to the widespread evolution of carbon concentrating mechanisms (CCMs) that increase the availability of CO_2 at the metabolic sites where it is required. Hence, while one might expect a process as fundamental to life on earth as carbon fixation in photosynthesis to be underpinned by a single unique biochemical pathway, in fact, even in higher plants there are three such pathways (and variants within them): the C_3 pathway (the most common), the C_4 pathway and CAM. The ecological consequences of the different pathways are profound, especially as they affect the reconciliation of photosynthetic activity and controlled water loss (see Section 3.3.1). Even in aquatic plants, where water conservation is not normally an issue, and most plants use the C_3 pathway, there are many CCMs that serve to enhance the effectiveness of CO_2 utilisation (Griffiths *et al.*, 2017). These CCM-based pathways are of profound importance. The C_4 and CAM pathways account for 18–30% of the 60 Pg (approximately) of carbon assimilated each year on land; while CCMs in cyanobacteria and algae account for more than half the 50 Pg of carbon assimilated each year in the oceans (Raven *et al.*, 2008).

the C_3 pathway

In the C_3 pathway, the Calvin–Benson cycle, CO_2 is fixed, through combination with ribulose 1,5-biphosphate (RuBP), into a three-carbon acid (phosphoglyceric acid) by the enzyme RuBisCO (ribulose-1,5-biphosphate carboxylase-oxygenase), which is present in massive amounts in the leaves (25–30% of the total leaf nitrogen). This same enzyme can also act as an oxygenase, as its name indicates, and this activity (photorespiration) can result in a wasteful release of CO_2 – reducing by about one-third the net amounts of CO_2 that are fixed. Photorespiration increases with temperature with the consequence that the overall efficiency of carbon fixation declines with increasing temperature.

The rate of photosynthesis of C_3 plants increases with the intensity of radiation, but reaches a plateau. In many species, particularly shade species, this plateau occurs at radiation intensities far below that of full solar radiation (see Figure 3.4). Plants with C_3 metabolism have low water-use efficiency compared with C_4 and CAM plants (see later), mainly because in a C_3 plant, CO_2 diffuses rather slowly into the leaf and so allows time for a lot of water vapour to diffuse out of it through the open stomata.

The rate of photosynthesis of C_3 plants also increases with the concentration of CO_2 within the plant, and because of the slow rate of diffusion, with the concentration of CO_2 in the atmosphere (see later). However, this rate is limited by the ability of C_3 plants to regenerate RuBP with which CO_2 can be combined, and therefore levels off as CO_2 concentrations increase.

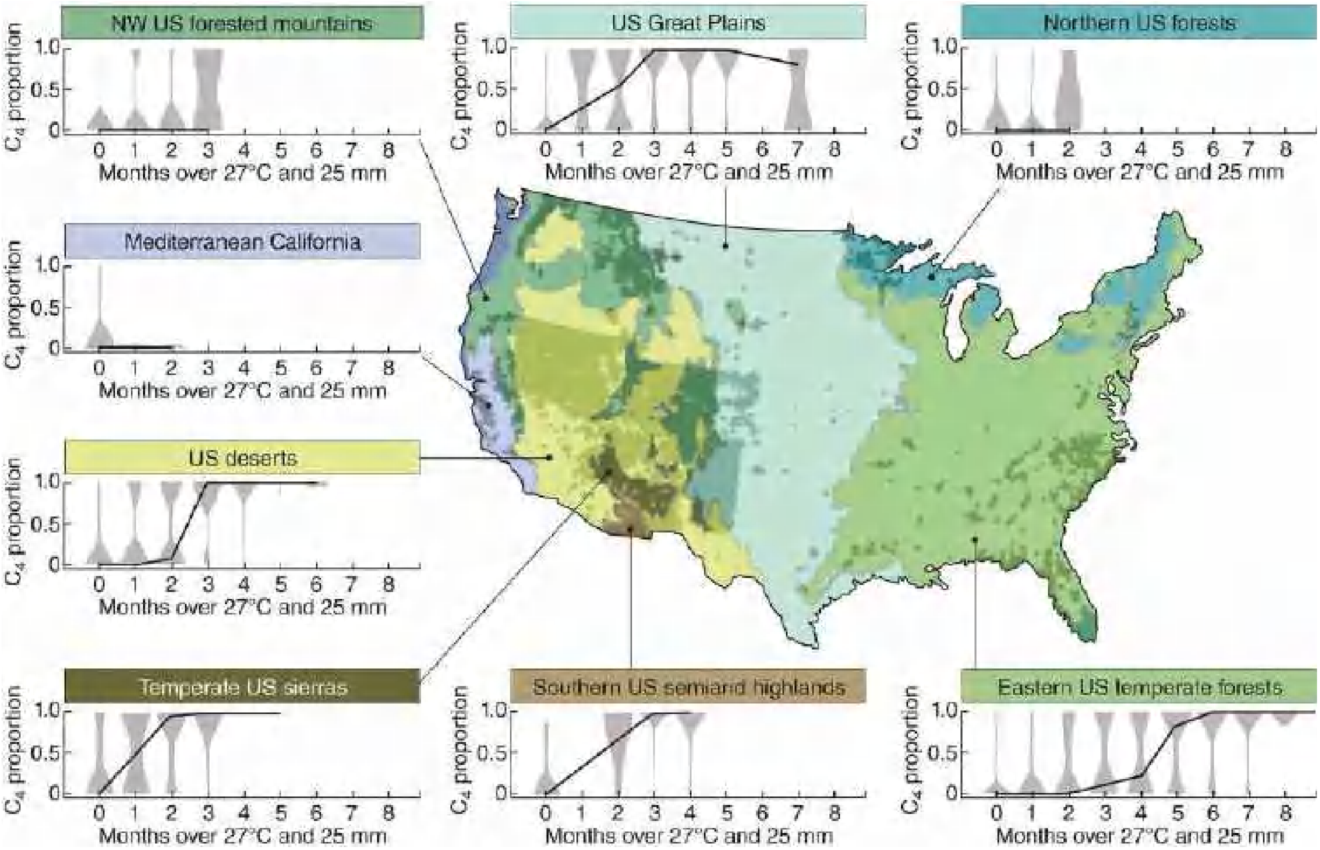
the C_4 pathway

In the C_4 pathway, the Hatch–Slack cycle, the C_3 pathway is present but it is confined to cells deep in the body of the leaf. CO_2 that diffuses into the leaves via the stomata meets mesophyll cells containing the enzyme phosphoenolpyruvate (PEP) carboxylase. This enzyme combines atmospheric CO_2 with PEP to produce a four-carbon acid. This diffuses, and releases CO_2 to the inner cells where it enters the traditional C_3 pathway. PEP carboxylase has a much greater affinity than RuBisCO for CO_2. There are profound consequences.

First, C_4 plants can absorb atmospheric CO_2 much more effectively than C_3 plants and the rate of photosynthesis is therefore much less dependent on CO_2 concentrations (but see later). Also, because of the reduced need to keep stomata open, C_4 plants may lose much less water per unit of carbon fixed. Furthermore, the wasteful release of CO_2 by photorespiration is almost wholly prevented and, as a consequence, the efficiency of the overall process of carbon fixation does not change with temperature. Finally, the concentration of RuBisCO in the leaves is a third to a sixth of that in C_3 plants, and the leaf nitrogen content is correspondingly lower. As a consequence of this, C_4 plants are much less attractive to many herbivores and also achieve more photosynthesis per unit of nitrogen absorbed.

It may seem surprising that C_4 plants, with such high water-use efficiency, have failed to dominate the vegetation of the world, but there are clear costs to set against the gains. The C_4 system has a high light compensation point and is inefficient at low light intensities; C_4 species are therefore ineffective as shade plants. Moreover, C_4 plants have higher temperature optima for growth than C_3 species: most C_4 plants are found in arid regions or the tropics. The pathway is widely distributed amongst plant families but is most prominent in grasses, where many of the attempts to account for the distributions of C_3 and C_4 species have been focused.

The most common approach to understanding the proportion of C_3 and C_4 plants in any region goes back to Collatz *et al.* (1998). It involves the identification of a climatological crossover temperature, above and below which C_4 and C_3 plants, respectively, are favoured – that is, they have a carbon gain advantage – and also a level of precipitation sufficient for plants of both types to grow. Collatz *et al.* estimated these for grasses to be a mean daytime temperature of 22°C and precipitation of at least 25 mm per month. Then, for example, the number of months in the year typically favouring C_4 growth may be used to account, statistically, for the proportion of C_4 grasses in a local flora. Subsequent refinements of the approach have re-estimated those growth criteria or acknowledged the importance of factors beyond temperature and precipitation. Thus, for instance, Griffith *et al.* (2015) explored a range of mean, minimum and maximum monthly temperatures for grasses in the USA and then found that the best fitting model was based on exceeding a monthly maximum temperature of 27°C, not a mean of 22°C (but still a mean monthly precipitation ≥25 mm; Figure 3.17). However, while this combination of temperature and precipitation thresholds was powerful in accounting for the distribution of C_4 grasses, in a number of regions, further factors were also important. In the Eastern Temperate Forest region, for example, there was a strong negative effect of tree cover on the proportion of C_4 grasses, since their shade promotes the cooler growing conditions more favourable to C_3 grasses; while in the Temperate US Sierras, there was a strong negative effect of mean annual precipitation, though whether this is favourable to C_3 grasses, unfavourable to C_4 grasses, or favourable to other plants that increase shading is uncertain.



[Figure 3.17](#) Effects of temperature and precipitation on the proportional contributions of $C_3$and C_4grasses to the floras of various regions of the USA, as indicated. Data were collected from sampling plots within each region, with their location marked as dots on the map, and these data are shown as symmetrical 'density curves', associated with the number of months at each location where conditions exceeded the estimated temperature-precipitation threshold (to the nearest month). The solid line is the predicted median proportion derived from a 'quantile regression' based on the temperature–precipitation threshold.

Source: After Griffith *et al.* (2015).

the CAM pathway

Plants with a CAM pathway also use PEP carboxylase with its strong power of concentrating CO_2. (The system is now known in a wide variety of families, not just the Crassulaceae.) In contrast to C_3 and C_4 plants, though, CAM plants open their stomata and fix CO_2 at night (as malic acid). During the daytime the stomata are closed and the CO_2 is released within the leaf and fixed by RuBisCO. However, because the CO_2 is then at a high concentration within the leaf, photorespiration is prevented, just as it is in plants using the C_4 pathway. Plants using the CAM photosynthetic pathway have obvious advantages when water is in short supply, because their stomata are closed during the daytime when evaporative forces are strongest. This appears to be a highly effective means of water conservation – water use efficiency for CAM plants is estimated to be around three times greater than for C_4 plants and more than six times greater than for C_3 plants (Borland *et al*., 2009) – but CAM species have not come to inherit the earth. One cost to CAM plants is the problem of storing the malic acid that is formed at night: most CAM plants are succulents with extensive water-storage tissues that cope with this problem. In general, CAM plants are found in arid environments where strict stomatal control of daytime water is vital for survival (desert succulents), and in habitats where CO_2 is in short supply during the daytime, for example in some submerged aquatic plants, and in photosynthetic organs that lack stomata (e.g. the aerial photosynthetic roots of orchids).

APPLICATION 3.2 Turning to CAM crops

The high water use efficiency of CAM plants makes them excellent candidates for cultivation over areas where rainfall is too little or evapotranspiration too high for the cultivation of C_3 or C_4 crops. Such areas may well expand in future. Examples include the production of pineapple, *Ananas comosus*, for consumption, of *Opuntia* spp. for animal fodder, of sisal, *Agave sisalana*, for fibre, and of *Agave tequilana* for the alcoholic beverages tequila and mescal (Borland *et al*., 2009). In addition, species of *Agave* are grown for both liquid and gaseous biofuels and indeed other products such as enzymes and lactic and succinic acid (Perez-Pimienta *et al*., 2017). Current quantities grown are small compared with the most common alternatives based around corn, sugar cane and wheat, and most of this is in Mexico. However, the much greater productivity of *Agave* (Figure 3.18), and the high water use efficiency it derives from its CAM metabolism, highlights its great potential for the future.

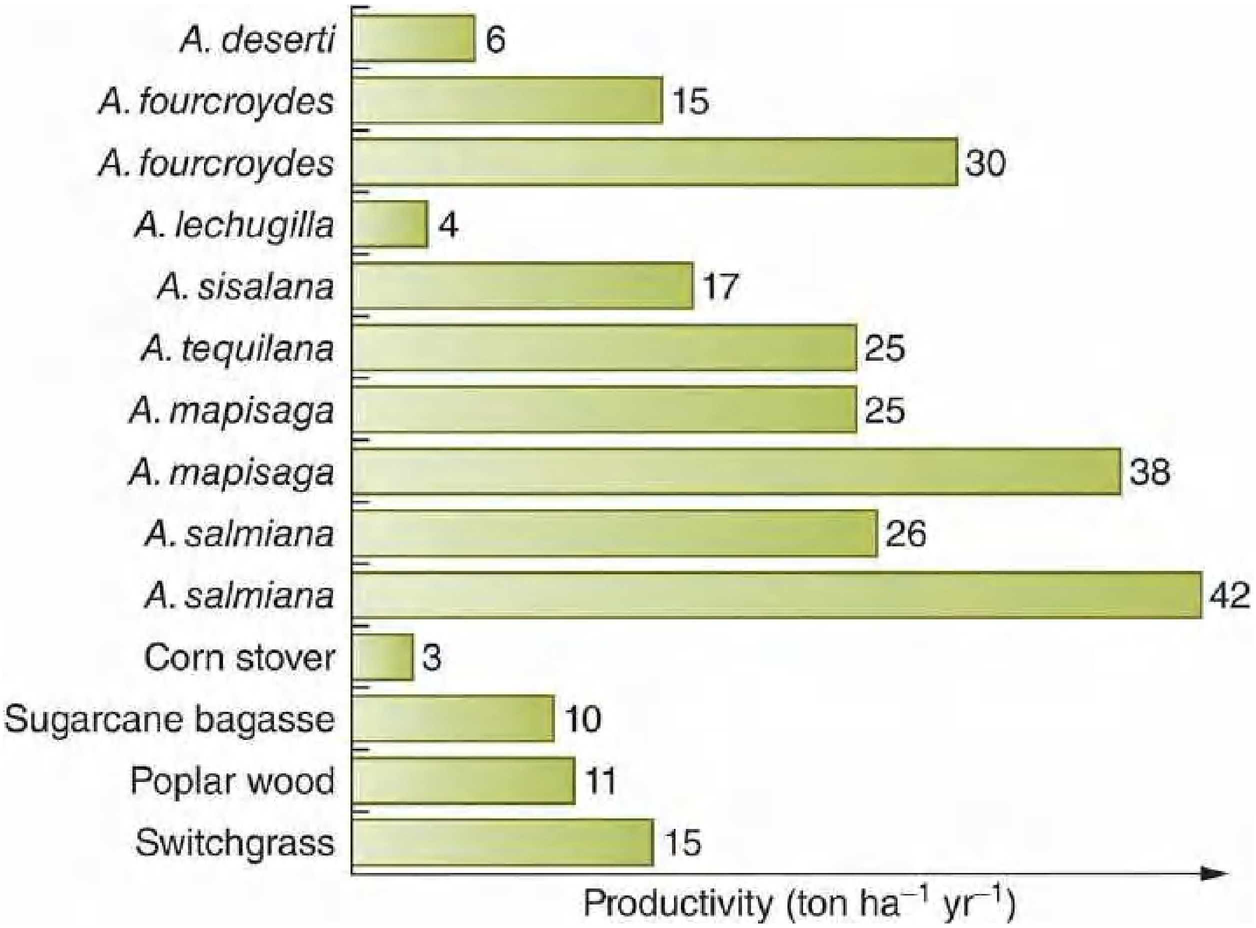


Figure 3.18 Agave species (CAM plants) exhibit high productivity due to their high water use efficiency. Comparison of the productivities of various species of *Agave* (double entries reflect separate studies) and four selected bioenergy feedstocks (raw products for the production of biofuels). Corn stover consists of the remains of plants left in the field after corn (*Zea mays*) has been harvested. Sugarcane bagasse is the fibrous matter than remains after sugarcane (*Saccharum* spp.) has been crushed to extract its juice.

Source: After Perez-Pimienta *et al*. (2017).

APPLICATION 3.3 Genetic engineering of CCMs into crops

Another way to exploit the potential of CAM would be to engineer its molecular machinery into C_3 plants (Borland *et al.*, 2014). Indeed there is great interest in general in engineering a wider range of CCMs, taken from C_4 plants and from microbes, into commercial crops. Complex metabolic alternatives, such as CAM, typically require a whole suite of co-adapted changes, and implementing them will therefore be a massive challenge. However, promising results may be obtained from less ambitious approaches. Figure 3.19, for example, illustrates the results when the photosynthetic rate of a standard variety of soybean, *Glycine max*, was compared with that of a transgenic variety that had been transformed to express the *ictB* gene taken from cyanobacteria, where it plays a crucial role in carbon uptake and is essential for their survival at anything other than very high CO_2 concentrations. In both varieties, photosynthetic rates increased with the concentrations of CO_2 in plant tissues but levelled off at higher concentrations. However, these rates, and especially the maximum rate, were significantly higher in the transgenic variety (Figure 3.19). Crucially, this difference translated into significant improvements in yield in the transgenic variety. As the authors of the study conclude, work such as this indicates that even single genes can contribute to the enhancement of yield in a major commodity crop '… and point to the significant role that biotechnological approaches to increasing photosynthetic efficiency can play in helping to meet increased global demands for food'.

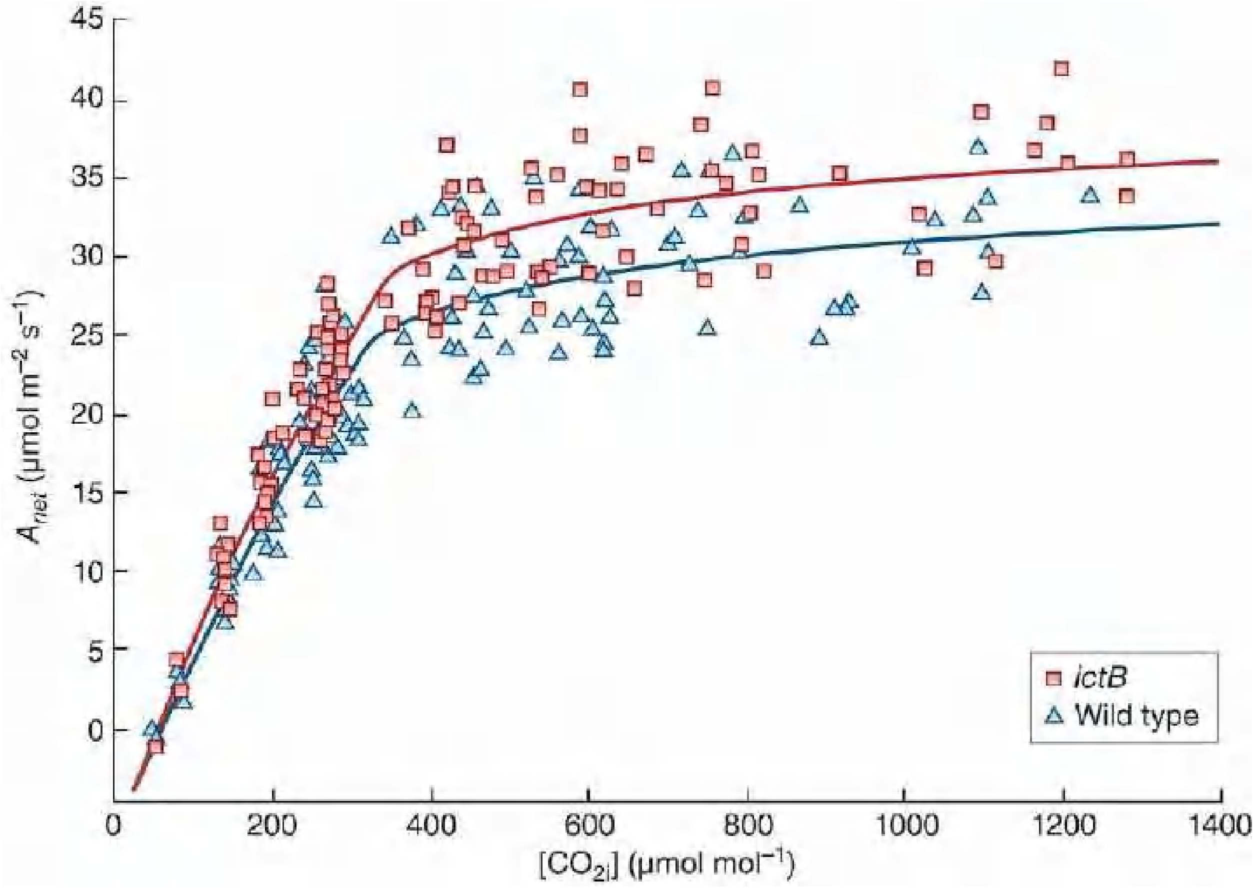


Figure 3.19 Bioengineering of a gene from cyanobacteria into soybean increases its rate of photosynthesis. A comparison of soybean (*Glycine max*) wild type (blue) and a variety genetically engineered to express the cyanobacterial membrane protein *ictB* (red) in terms of the effect of a plant's internal CO_2 ([CO_{2i}]) concentration on its net rate of photosynthesis (A_{net}). The fitted lines were non-rectangular hyperbolas reflecting current understanding of the underlying physiology.

Source: After Hay *et al*. (2017).

the evolution of C_4 and CAM

The evolution of the C_4 and CAM pathways, and of CCMs generally (that increase the concentration of CO_2 around RuBisCO), has been reviewed by Raven *et al*. (2008). They describe the very strong evidence that these mechanisms are evolutionarily primitive (the earliest appearing at least 300 million years ago) but also that the C_4 and CAM systems must have arisen repeatedly and independently during the evolution of the plant kingdom – the most recent being the appearance of C_4 in land plants only 20–30 million years ago. This prolonged coexistence of multiple paths to carbon fixation foreshadows a pattern we will see many more times in later chapters – of coexisting species utilising the same resources but in different ways.

3.4.2 The response of plants to changing atmospheric concentrations of CO_2

Of all the various resources required by plants, CO_2 is the only one that is increasing on a global scale. This rise is strongly correlated with the increased rate of consumption of fossil fuels and the clearing of forests. As Loladze (2002) points out, while consequential changes to global climate may be controversial in some quarters, marked increases in CO_2 concentration itself are not. High mixing rates in the atmosphere mean that these are changes that will affect *all* plants. Plants now are experiencing around a 30% higher concentration compared with the preindustrial period – effectively instantaneous on a geological timescale. Trees living now may experience a doubling in concentration over their lifetimes – effectively instantaneous on an *evolutionary* timescale.

changes in geological time

Putting this in a wider context, though, there is also evidence of large-scale changes in atmospheric CO_2 over much longer timescales. A range of models suggest that during the Triassic, Jurassic and Cretaceous periods (around 250–70 million years ago), atmospheric concentrations of CO_2 were four to eight times greater than at present (and both lower and very much higher prior to that), falling after the Cretaceous from between 1500 and 3000 $\mu l\, l^{-1}$ to below 1000 $\mu l\, l^{-1}$ in the subsequent Eocene, Miocene and Pliocene, and fluctuating between 180 and 280 $\mu l\, l^{-1}$ during more recent glacial and interglacial periods (the last 400 000 years; Figure 3.20). The steady rise in CO_2 since the Industrial Revolution is therefore a partial return to pre-Pleistocene conditions, more than 2.5 million years ago.

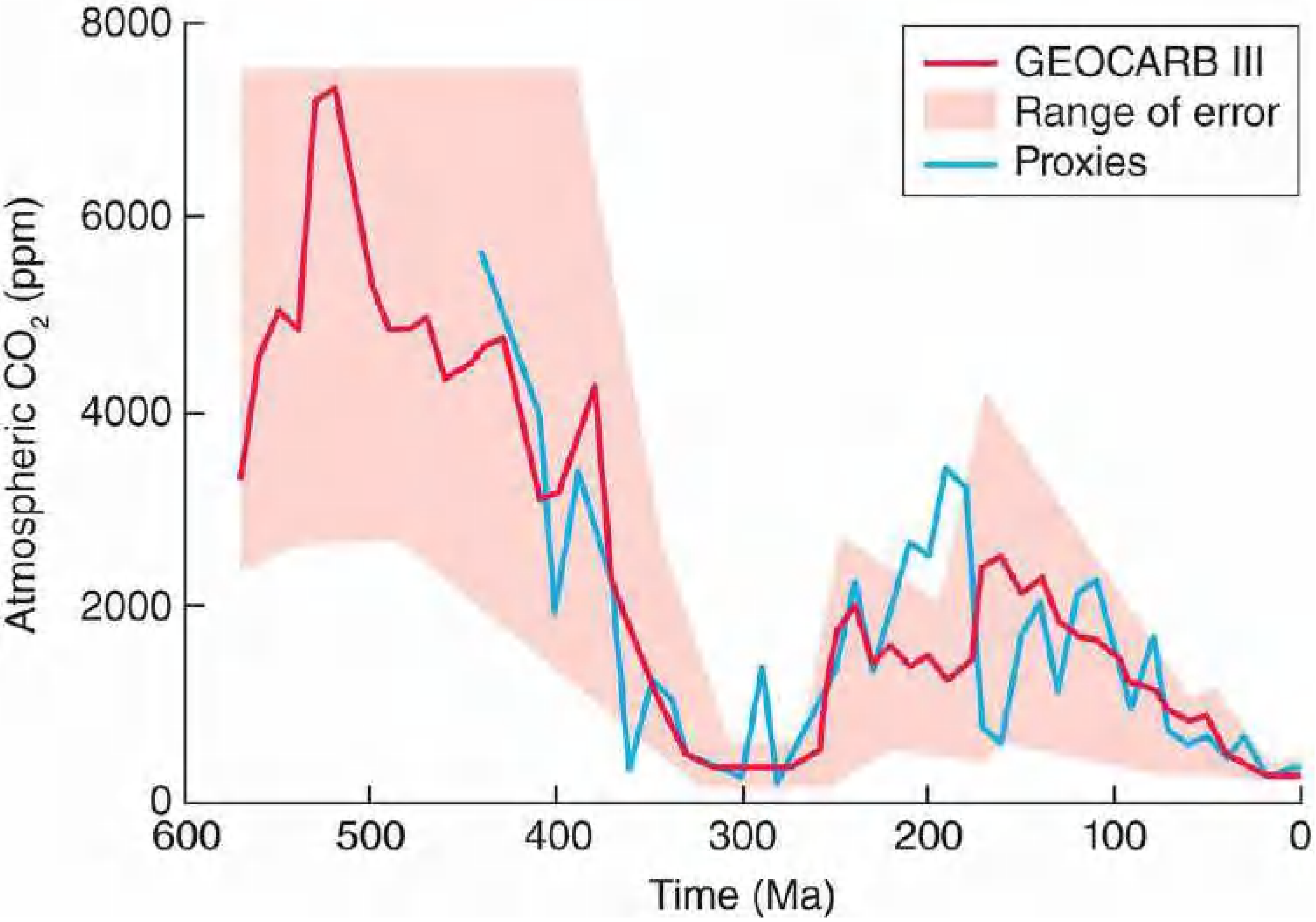


Figure 3.20 Estimates of atmospheric CO_2concentrations over the past 600 million years (Ma). The red line (with estimated range of error in pink) is the output of the GEOCARB III model, which uses estimates of rates of geological weathering, emission and burial modified by a range of factors including global temperature and continental size. The blue line is the result of averaging four proxies of CO_2 concentrations, including isotopic compositions of minerals and plankton and the distribution of stomata in plant leaves.

Source: After Royer *et al.* (2004).

what will be the consequences of current rises?

When other resources are present at adequate levels, additional CO_2 scarcely influences the rate of photosynthesis of C_4 plants but it increases the rate for C_3 plants. As atmospheric concentrations continue to rise, therefore, it is no surprise that there has been considerable interest in the effects of higher CO_2 concentrations on the productivity of individual plants and of whole crops, and natural communities including tropical rainforests (Lewis *et al*., 2009). Earlier studies often used open-top chambers into which CO_2 was blown before escaping through the top, but increasingly and now predominantly, use is made of free air CO_2 enrichment facilities (FACE – Figure 3.21) in which a ring of pipes release CO_2, at a range of heights, into an otherwise unconstrained body of plants much larger than an open top chamber (often 8–30 m in diameter). A computer-controlled system is used to regulate the flow so as to maintain the target CO_2 concentration in the FACE facility, typically 475–600 ppm.

Reviews of FACE studies have consistently shown increases in photosynthetic rates in response to elevated CO_2 concentrations (Figure 3.21a), and these responses have been markedly greater in C_3 than in C_4 plants (Figure 3.21b), as predicted. Elevations in photosynthetic rates have also often been translated into increases in yield, though it is striking that such effects are more marked under the more natural conditions of a FACE facility than in open top chambers (Figure 3.21c).

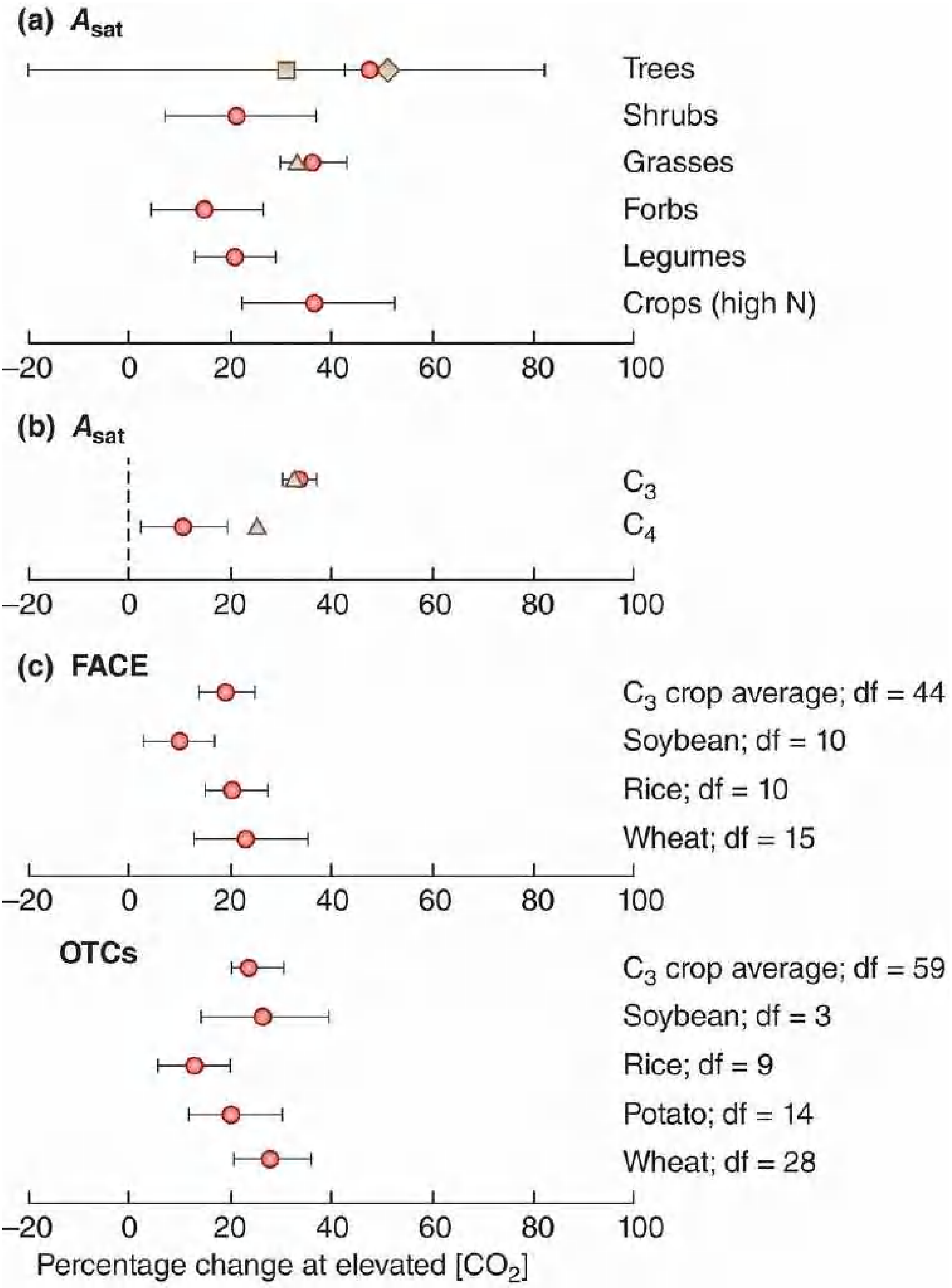
(a) A_{sat}
Trees
Shrubs
Grasses
Forbs
Legumes
Crops (high N)
–20
0
20
40
60
80
100
(b) A_{sat}
C_3
C_4
–20
0
20
40
60
80
100
(c) FACE
C_3 crop average; df = 44
Soybean; df = 10
Rice; df = 10
Wheat; df = 15
–20
0
20
40
60
80
100
OTCs
C_3 crop average; df = 59
Soybean; df = 3
Rice; df = 9
Potato; df = 14
Wheat; df = 28
–20
0
20
40
60
80
100
Percentage change at elevated [CO_2]

Figure 3.21 Photosynthetic activity is increased by enhanced CO_2concentrations in FACE experiments. (a) Mean responses from meta-analyses of light-saturated CO_2 uptake (A_{sat}), a measure of photosynthetic activity, to enhanced CO_2 concentrations ($[CO_2]$) in free air CO_2 enrichment facilities (FACE) experiments, for a variety of plant groups. Red symbols are from the current meta-analysis, brown symbols from previous meta-analyses. (b) Mean responses from these meta-analyses of light-saturated CO_2 uptake (A_{sat}) to enhanced CO_2 concentrations ($[CO_2]$) according to whether plants were C_3 or C_4. (c) Mean responses from a meta-analysis of the yields of various crops, as indicated, to enhanced CO_2 concentrations ($[CO_2]$) in FACE experiments and open top chambers (OTCs). Bars are 95% CIs in all parts.

Source: (a, b) After Ainsworth & Long (2005). (c) After Bishop *et al.* (2014).

Such responses are in part simply a reflection of a greater availability of resource (CO_2), but there is likely to be an additional effect, especially in C_3 plants, resulting from a reduced need for stomatal opening and the consequent reduction in water loss. This reduction in stomatal conductance has indeed been observed, at least in crops and especially under drought conditions (Figure 3.22a). This in turn suggests that such increases in yield at higher CO_2 concentrations should themselves be higher when there is reduced availability of water, since this is when the benefits of conserving water will be greatest. This, too, has been confirmed (Figure 3.22b).

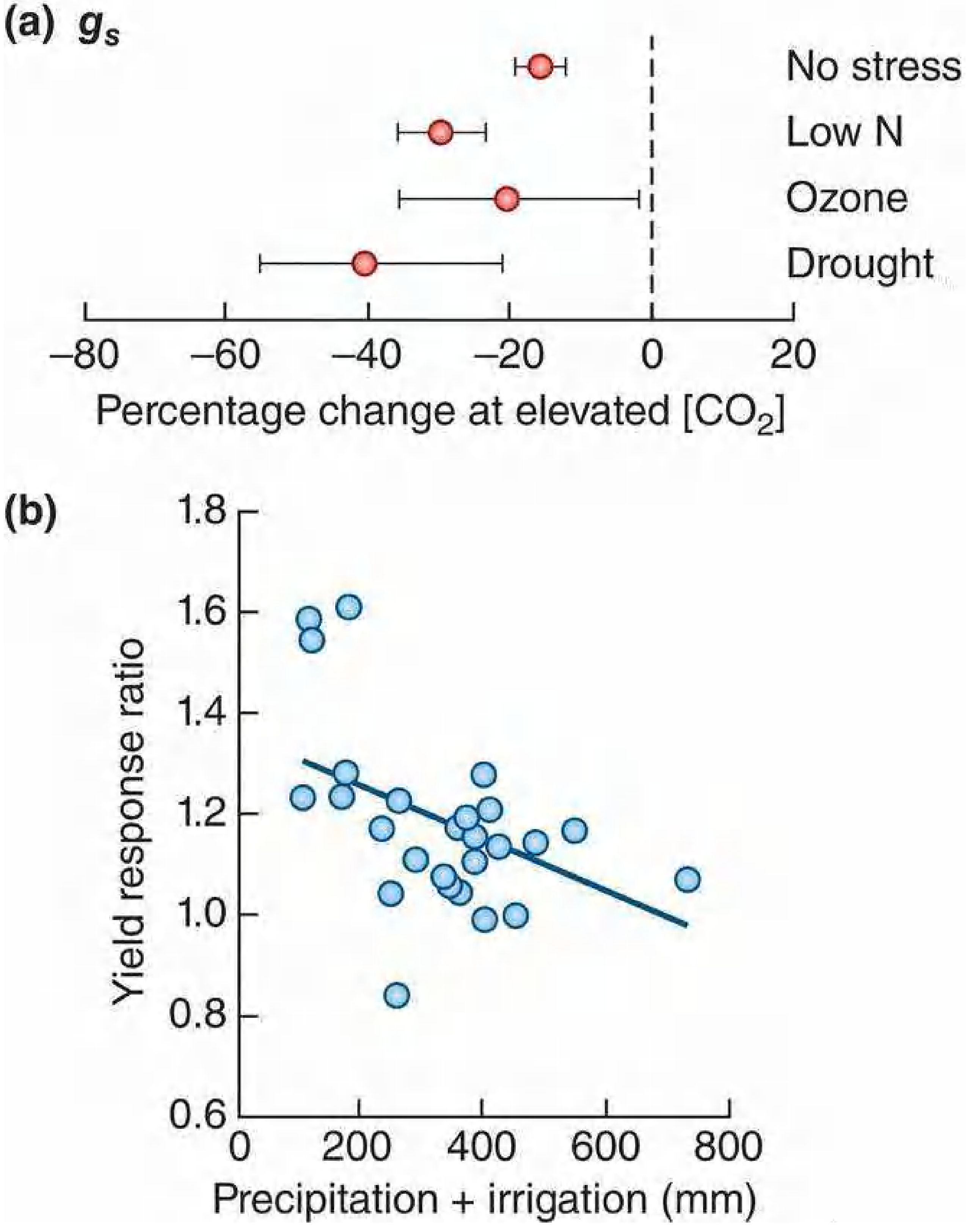

(a) g_s
No stress
Low N
Ozone
Drought
−80
−60
−40
−20
0
20
Percentage change at elevated [CO_2]
(b)
1.8
1.6
1.4
1.2
1.0
0.8
0.6
Yield response ratio
0
200
400
600
800
Precipitation + irrigation (mm)

Figure 3.22 Stomatal conductance is decreased by enhanced CO_2concentrations, leading to increased yields especially when water is scarce. (a) Mean responses from the meta-analysis in Figure 3.21c of stomatal conductance (g_s) to enhanced CO_2 concentrations ([CO_2]) in free air CO_2 enrichment facilities (FACE) experiments with crop plants, classified according to different additional treatments. Bars are 95% CIs. (b) The significant negative relationship of the yield increases in Figure 3.21c to the level of water availability ($y = 1.36 - 0.0005x$; $r^2 = 0.24$; $P = 0.009$).

Source: After Bishop *et al.* (2014).

We should not be tempted by these general patterns, however, into thinking that responses to elevated CO_2 concentrations will be straightforward and predictable. The results of a 20-year FACE experiment, in which the responses of C_3 and C_4 grasses were compared, are shown in Figure 3.23. For the first 12 years of the experiment, the results were as might conventionally have been predicted. There was a marked increase in biomass in C_3 plots compared with those experiencing ambient CO_2, but no such response in C_4 plots. However, this pattern was reversed in the subsequent eight years (Figure 3.23a). The explanation for this reversal is uncertain, but it is associated with a shifting balance in the effect of CO_2 enhancement on the availability of nitrogen in the soils of the contrasting plots. The effect was positive initially in the C_3 plots (more nitrogen available) but became increasingly negative, and was negative initially in the C_4 plots but became increasingly positive (Figure 3.23b). And overall in the experiment, positive effects of CO_2 enhancement on nitrogen availability gave rise to positive effects on biomass. The authors of the study admit, frustratingly, that the underlying basis of these nitrogen cycling responses remains an open question. But these results do at least remind us, first, that short-term responses may be misleading in predicting the consequences of what are likely to be long-term trends, and also, that the changing availability of one resource may, within the community as a whole, give rise to altered levels of other resources with equally profound consequences.

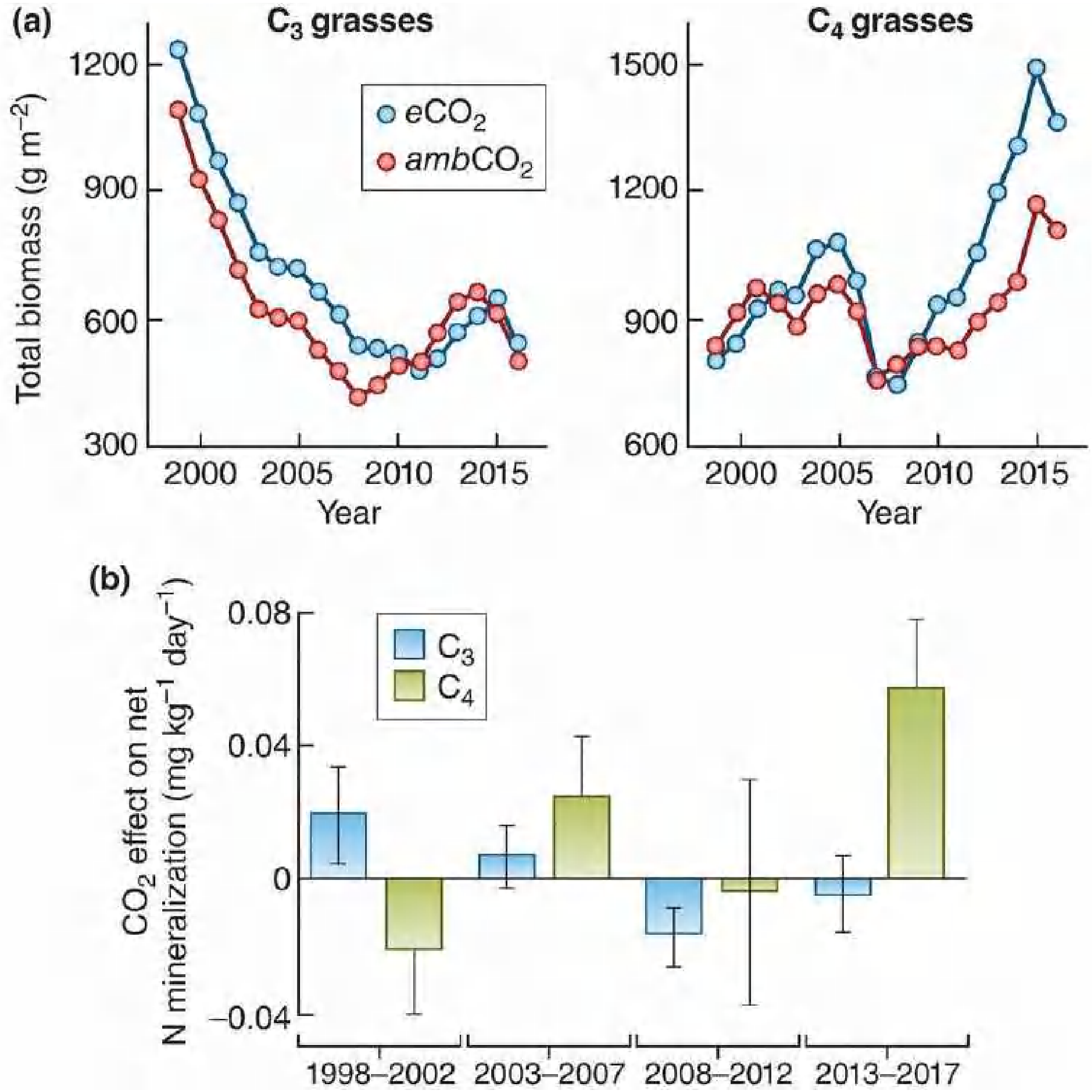


Figure 3.23 Effects of elevated CO_2concentrations on $C_3$and C_4grasses are reversed over the longer term. (a) Changes in biomass over a 20-year period in a free air CO_2 enrichment facilities (FACE) experiment comparing the responses of C_3 and C_4 grasses to enhanced CO_2 concentration (eCO_2) relative to ambient levels ($amb CO_2$). Data shown are three-year moving averages. In an analysis of variance, the interaction between year, treatment and C_3/C_4 was significant ($F = 7.2$; $P < 0.01$): the difference in response between C_3 and C_4 grasses reversed over time. (b) The effect in the experiment on net nitrogen mineralisation (a measure of nitrogen availability to the plants averaged over successive five-year periods). Bars are SEs among years. In an analysis of variance, the interaction between year, treatment and C_3/C_4 was significant ($F = 4.0$; $P < 0.05$): the difference in response between C_3 and C_4 grasses reversed over time.

Source: After Reich *et al.* (2018).

APPLICATION 3.4 Harmful effects of plants' responses to CO_2 enrichment

Along similar lines, it is well established that there is a general tendency for CO_2 enrichment to change the composition of plants, and in particular to reduce nitrogen (and hence protein) concentrations in above-ground plant tissues (Cotrufo *et al.*, 1998), which may in turn have indirect effects on plant–animal interactions, because insect herbivores may then eat more foliage to maintain their nitrogen intake and fail to gain weight as fast (Fajer, 1989). CO_2 enhancement may also reduce concentrations in plants of other essential nutrients and micronutrients (see Section 3.5), contributing in turn to 'micronutrient malnutrition', which diminishes the health and economy of more than one-half of the world's human population (Loladze, 2002).

Examples of the effects on nitrogen content are shown in Figure 3.24. First, when wheat was grown at enhanced CO_2 concentrations, effects depended on whether soils had adequate nitrogen (low-N) or this was supplemented with further fertilisation (high-N) (Figure 3.24a). In the low-N treatment, the protein concentration of individual grains was substantially reduced, but enhanced growth overall meant that the protein harvested from the grain was much less reduced. Indeed, in the high-N treatment, the effect on concentration was negligible, such that the protein harvested overall from the grain was increased by CO_2 enhancement. On the other hand, it was possible to pinpoint the reason for the lowered nitrogen content under CO_2 enhancement, namely a reduction in nitrate assimilation efficiency (Figure 3.24b) – likely to be a very common cause of nitrogen depletion in tissues following CO_2 enhancement. Thus, while it may be possible to ameliorate the effects of future (and present) rises in CO_2 concentrations by increased nitrogen fertilisation, this reduction in assimilation efficiency will substantially increase the fertilisation costs and the polluting levels of unassimilated nitrates in soil run-offs.

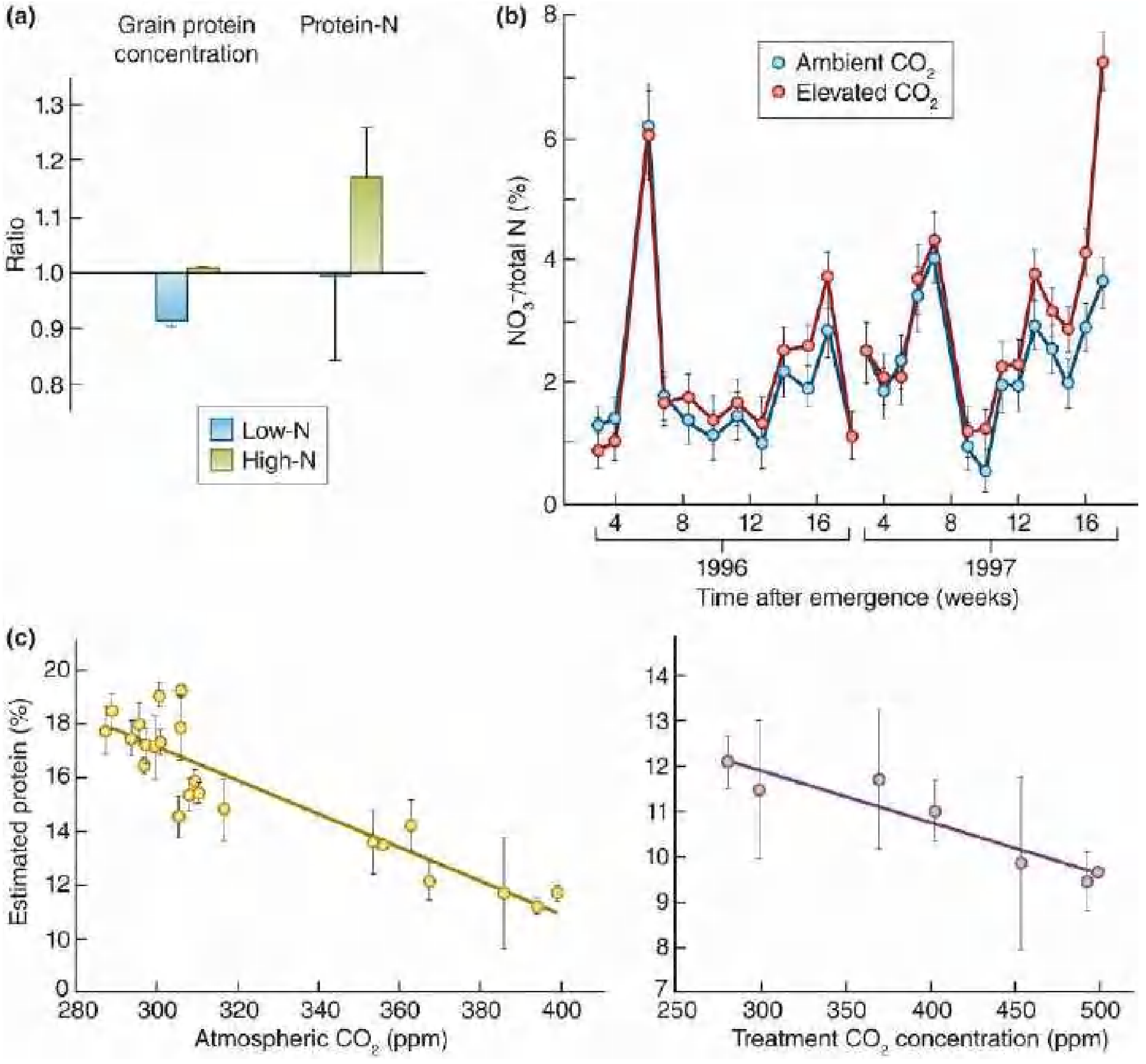


Figure 3.24 **Effects of CO_2enhancement on plant protein concentrations with potentially harmful consequences.** (a) The ratios of grain protein concentration (left) and the total protein-nitrogen harvested from wheat, *Triticum aestivum* (right) in a free air CO_2 enrichment facilities (FACE) experiment comparing enhanced to ambient CO_2 concentrations at low and high soil nitrogen levels. Bars are SEs. (b) The proportion of nitrogen in wheat leaves that was nitrate over the course of this same experiment, which was consistently higher at elevated than ambient CO_2 concentrations ($P < 0.0001$), indicative of lower nitrate assimilation at elevated concentrations. Bars are SEs. (c) The protein content of goldenrod (*Solidago* spp.) pollen in relation to atmospheric CO_2 concentration from museum collections from 1842 to 1998 (left, $r^2 = 0.81$, $P < 0.001$) and from a FACE-type experiment generating a CO_2 gradient (right, $r^2 = 0.88$, $P < 0.01$). Bars are SEs.

Source: (a) After Kimball *et al.* (2001). (b) After Bloom *et al.* (2014). (c) After Ziska *et al.* (2016).

A second example looked at the effect of increased CO_2 concentrations on the protein content of pollen from goldenrod plants, *Solidago* spp., in the USA, widely acknowledged by apiarists to be essential for the health and winter survival of both native bees (e.g. *Bombus* spp.) and honey bees (*Apis melifera*). Data both from historical records of pollen collected as CO_2 levels have risen, and from experiments that used a FACE-like facility to establish a CO_2 gradient, showed that the protein content of the pollen was reduced substantially by increases in CO_2 concentrations (Figure 3.24c). These reductions could have serious effects on bee numbers, on pollination rates and hence on plant productivity, but the generality of these effects, the abilities of bees to mitigate them through changes in their own behaviour, and indeed the extent of the harm done to bees all remain to be determined. Examples such as these, therefore, emphasise both how profound the potential implications of CO_2 increases may be, in their own right, for future food security, and how difficult these implications can be to predict.

3.5 Mineral nutrients

macronutrients and trace elements

It takes more than light, CO_2 and water to make a plant. Mineral resources are also needed and these must be obtained from the soil or, in the case of aquatic plants, from the surrounding water. These include macronutrients (i.e. those needed in relatively large amounts) – nitrogen (N), phosphorus (P), sulphur (S), potassium (K), calcium (Ca), magnesium (Mg) and iron (Fe) – and a series of trace elements, for example, manganese (Mn), zinc (Zn), copper (Cu), boron (B) and molybdenum (Mo) (Figure 3.25). Many of these elements are also essential to animals, although it is more common for animals to obtain them in organic form in their food than as inorganic chemicals. Some plant groups have special requirements. For example, aluminium is a necessary nutrient for some ferns, silicon for diatoms, and selenium for certain planktonic algae.

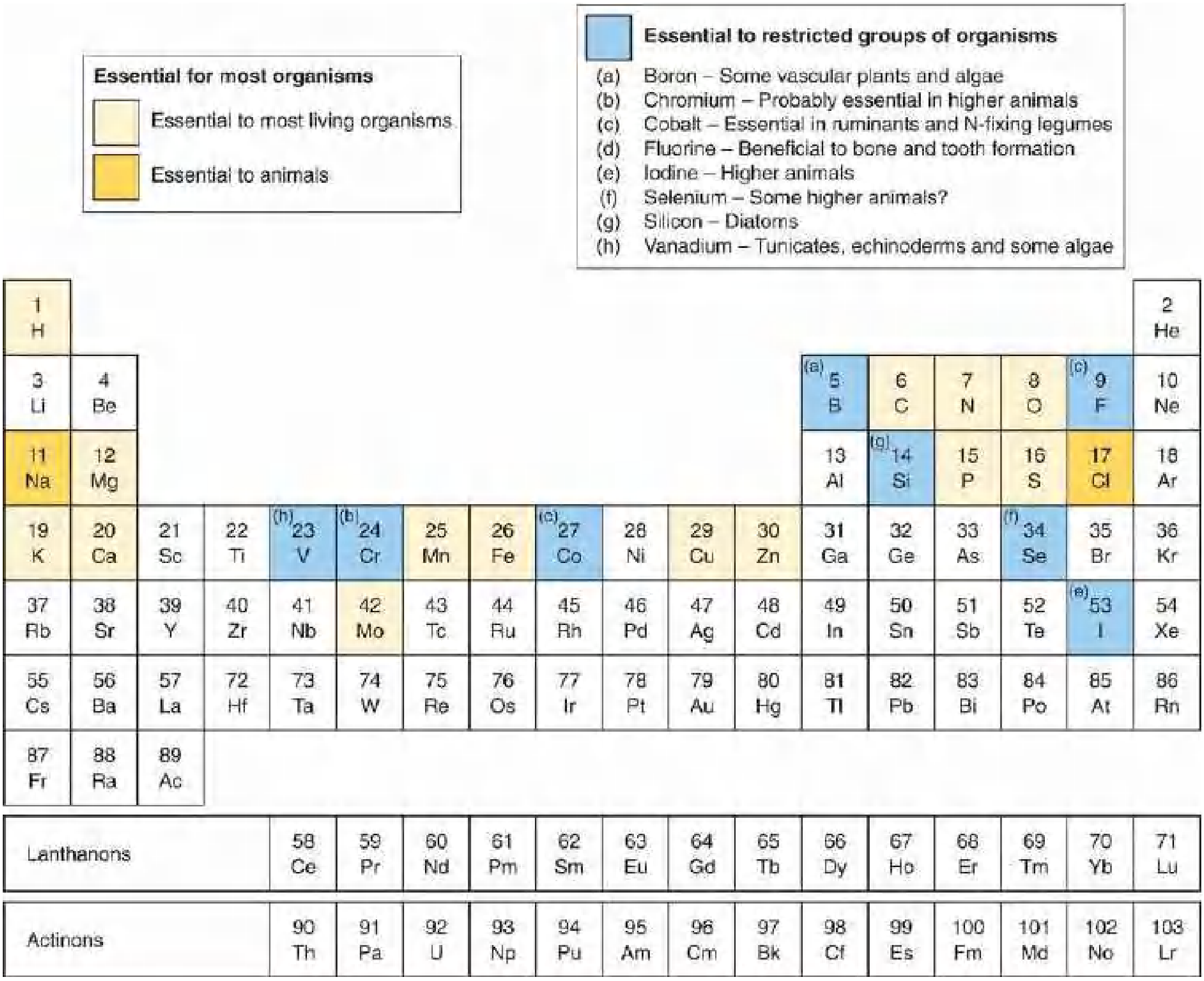


Figure 3.25 Periodic table of the elements showing those that are essential resources in the life of selected organisms.

Green plants do not obtain their mineral resources as a single package. Each element enters the plant independently as an ion or a molecule, and each has its own characteristic properties of absorption in the soil and of diffusion, which affect its accessibility to the plant even before any selective processes of uptake occur at the root membranes. All green plants require all of the 'essential' elements listed in Figure 3.25, although not in the same proportion, and there are some quite striking differences between the mineral compositions of plant tissues of different species and between the different parts of a single plant (Figure 3.26).

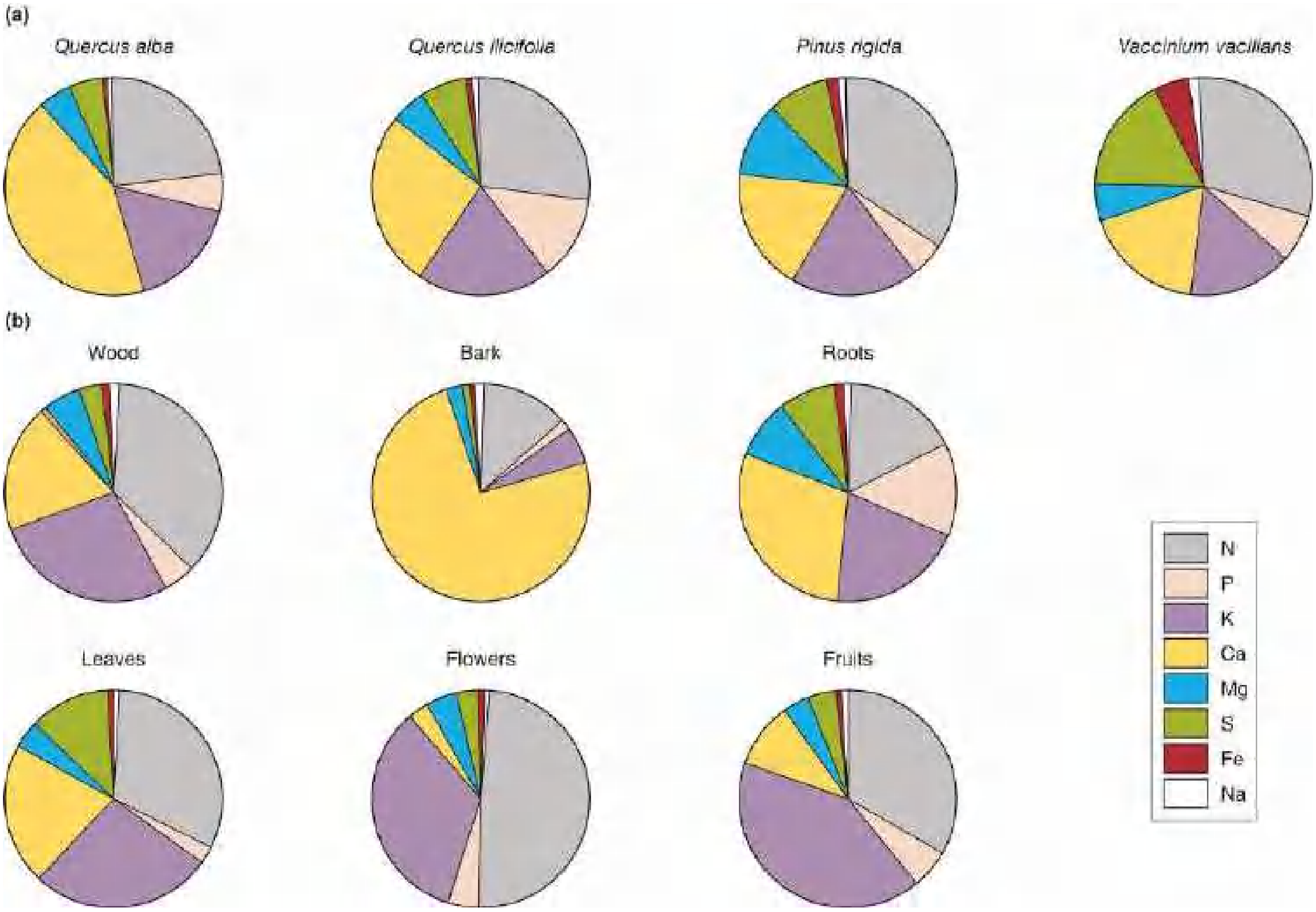


Figure 3.26 The mineral compositions of different plants and plant parts are very different. (a) The relative concentration of various minerals in whole plants of four species in the Brookhaven Forest, New York. (b) The relative concentration of various minerals in different tissues of the white oak (*Quercus alba*) in the Brookhaven Forest. Note that the differences between species are much less than between the parts of a single species.

Source: After Woodwell *et al.* (1975).

foraging for nutrients

There are strong interactions between water and nutrients as resources for plant growth. Roots will not grow freely into soil zones that lack available water, and so nutrients in these zones will not be exploited. Plants deprived of essential minerals make less growth and may then fail to reach volumes of soil that contain available water. There are similar interactions between mineral resources. A plant starved of nitrogen makes poor root growth and so may fail to 'forage' in areas that contain available phosphate or indeed contain more nitrogen. Again, plants may make both strategic and tactical responses to heterogeneities in nutrient availability (Hodge, 2004).

nitrogen

Nitrogen is the element that organisms require in the greatest amounts after carbon, hydrogen and oxygen. It is no surprise therefore that nitrogen availability often limits overall productivity in an ecosystem. Higher plants acquire nitrogen through their roots in inorganic form – as ammonium and nitrate salts – and in organic forms as urea, peptides and amino acids. This is true, too, of microorganisms, but they are best adapted to use the organic sources, followed by ammonium and then nitrate. Phytoplankton, fungi, cyanobacteria and bacteria, therefore, usually only assimilate nitrate in the absence of organic nitrogen and ammonium. Higher plants, by contrast, are much less adept at acquiring organic nitrogen, and compete strongly with soil microorganisms for ammonium sources. Nitrates are therefore the major source of nitrogen for most higher plants (Bloom, 2015).

The acquisition of nitrogen by plants is facilitated both by molecular transporters at the root surface and by root architecture (Kiba & Krapp, 2016). Of all the major plant nutrients, nitrates move most freely in the soil solution and are carried from as far away from the root surface as water is carried. Hence nitrates will be most mobile in soils at or near field capacity, and in soils with wide pores, and they will be captured most effectively by wide ranging, but not intimately branched, root systems. Their RDZs will themselves be wide, and those produced around neighbouring roots are likely to overlap such that the roots compete for the same nitrogenous molecules.

As we discuss in more detail in Section 13.9, the roots of most terrestrial plants are colonised by specialist fungi, forming *mycorrhizas*. In fact, it is these intimate, 'mutualistic' associations between the two (beneficial to both parties), rather than simply the roots alone, that are responsible for nutrient acquisition (as well as providing a series of other benefits to the plants). The fungi, for their part, are reliant on the plants for carbon. The advantages to plants of having mycorrhizal fungi are most apparent in the case of less mobile nutrients (see below), but even with nitrogen, mycorrhizas may have some role to play (Jin *et al*., 2012). Of arguably greater significance to the nitrogen economy, many plants form intimate mutualistic associations in their roots with nitrogen-fixing bacteria, overcoming the shortage of available nitrogen in the soil by harnessing the microbes' ability to convert free nitrogen in the atmosphere into ammonia, nitrate and other compounds. The most important example is the association between leguminous plants and rhizobia. These are discussed in detail in Section 13.11.

phosphorus

In many habitats, the phosphorus levels available to plants are limiting to growth, even though phosphorus itself may be abundant. It forms inert complexes, notably with iron and aluminium, and even the free phosphorus in soil solutions is relatively immobile, much of it being tightly bound on soil colloids from which its release is difficult. In contrast to nitrogen, therefore, it pays plants foraging for phosphate to explore the soil intensively rather than extensively, and the RDZs tend to be narrow. Roots or root hairs or threads from mycorrhizas will only tap common pools of free phosphorus (that is, they will compete with one another) if they are very close together.

Indeed, mycorrhizas play a crucial role in facilitating most plants' acquisition of phosphate, producing branched mycelial threads up to 100 times longer than root hairs, as well as having physiological capabilities that increase the phosphate flow (Javot *et al*., 2007). At the very lowest levels of phosphate availability, however (either because of its near-absence in the soil or because it is especially tightly bound) a number of plants lack mycorrhizas, using instead an alternative strategy, namely the production of citrate and other carboxylates in their roots, often specialised, very finely divided structures called cluster roots. The carboxylates mobilise phosphate from its tightly bound (unavailable) state, such that cluster root species can make better growth at low levels of phosphorus supply than mycorrhizal species (Lambers *et al*., 2015).

potassium

Potassium is another key mineral in plant nutrition, often abundant in the soil, but again, strongly adsorbed to soil particles and hence of potentially limiting availability. The role of mycorrhizas in potassium acquisition is relatively poorly understood but is becoming increasingly apparent (Garcia & Zimmermann, 2014).

It is clear even from these few examples that different mineral ions are held by different forces in the soil, that plants with different shapes of root system, with different root system properties, and with different mycorrhizal associations may therefore tolerate different levels of soil mineral resources, and that different species may deplete different mineral resources to different extents. This may be of great importance in allowing a variety of plant species to cohabit in the same area. We deal with the coexistence of competitors in Chapter 8.

3.6 Oxygen – and its alternatives

Oxygen is a resource for both animals and plants as the final electron acceptor in the process of aerobic respiration that provides the energy that drives their metabolism. In above-ground terrestrial environments it is rarely in limited supply, but its diffusibility and solubility in water are very low and so it can become limiting much more readily in aquatic and waterlogged environments. Because oxygen diffuses so slowly in water, aquatic animals must either maintain a continual flow of water over their respiratory surfaces (e.g. the gills of fish), which often have very large surface areas relative to body volume, or they may have specialised respiratory pigments (e.g. diving mammals and birds, see Mirceta *et al.* (2013)), or may continually return to the surface to breathe. The roots of many higher plants fail to grow into waterlogged soil, or die if the water table rises after they have penetrated deeply. Even if roots do not die when starved of oxygen, they may cease to absorb mineral nutrients so that the plants suffer from mineral deficiencies.

However, it would be wrong to adopt a higher-organism centred point of view in which respiration is predictably aerobic, reliant on oxygen as a resource that is equally predictably available. On the contrary, there are environments where oxygen is simply absent – often described as ‘extreme’, such as hot springs or deep in the ocean – and many others in which oxygen levels are depleted by biological activity at rates that cannot be counteracted by diffusion or by the activities of photoautotrophs. This is the case, for example, when organic matter decomposes in aquatic environments, and microbial respiration makes a demand for oxygen that exceeds the immediate supply. It is true, too, in water bodies that suffer eutrophication (see Section 21.1.3) when they are overly enriched with nutrients, particularly nitrates and phosphates, often as pollutants, inducing excessive growth of plants and algae that may again deplete oxygen faster than it can be replaced. Many microorganisms living in all these types of environment respire *an*aerobically, using alternative resources to oxygen as the final electron acceptor in the respiratory process: nitrates, sulphates, CO_2, ferric iron and many others. Of course, where oxygen is absent altogether, all those respiring actively must do so anaerobically.

anaerobic respiration: widespread and varied

Anaerobic respiration is generally far less efficient (produces much less energy) than aerobic respiration. Hence, when oxygen is in ready supply, aerobes thrive and anaerobes are little in evidence. However, the balance within ecological communities can change rapidly. One reason is that many microbes are facultatively anaerobic – capable of both aerobic and anaerobic respiration. We see an example in Figure 3.27a, where pitcher plants (carnivorous plants that trap their prey in pitcher-shaped modified leaves) contain a digestive liquid that supports a community of microbes. Natural communities of pitchers of the northern pitcher plant, *Sarracenia purpurea*, from Vermont, USA, were compared with pitchers that had been enriched through the repeated addition of finely ground insects (without a microbial community of their own) similar to the plants’ natural prey. Such excess loading of organic material leads to an increase in microbial activity and hypoxic (low oxygen) conditions within the experimental pitchers, as can happen naturally in pitcher plants, and as it does in much large water bodies such as ponds and lakes. The microbial activity within the communities of the control and experimental pitchers were very different, as judged by the peptides within them, which could be extracted, identified and assigned to the types of microbes producing them (Figure 3.27a). When oxygen was readily available as a resource, most of the peptides were contributed by aerobic bacteria. But when oxygen was in very short supply, most came from facultative anaerobes that could rapidly switch their metabolism from aerobic to anaerobic respiration. It is also notable, therefore, that in neither case was there a major contribution from bacteria that were obligatory anaerobes.

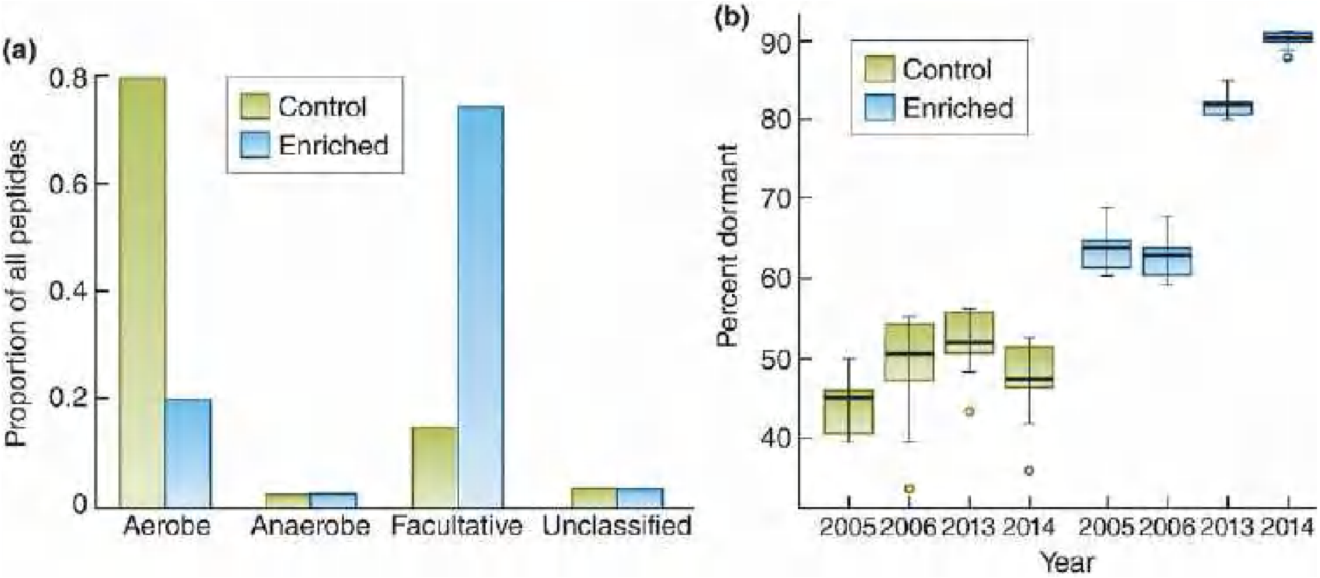


Figure 3.27 Enrichment commonly leads to a switch from oxygen to (anaerobic) alternatives as a resource for respiration. (a) The proportion of microbial peptides in communities occupying the pitchers of *Sarracenia purpurea* that were either controls or enriched, originating from microbes with different respiratory modes. (b) The percentage of taxa that were dormant (metabolically inactive) in control and nitrogen-enriched plots in saltmarshes over four years. Bold lines are medians, boxes represent 25–75 percentiles and whiskers show ranges, with outliers also shown.

Source: (a) After Northrop *et al.* (2017). (b) After Kearns *et al.* (2016).

Similarly, but on a larger scale, enrichment of salt marshes in Massachusetts, USA, led to soil microbial communities in which a much increased proportion of the taxa was dormant, that is, metabolically inactive (Figure 3.27b), but among those that were active, there was a major shift from aerobic to (in this case) obligatorily anaerobic taxa, many using sulphate or sulphur rather than oxygen as their respiratory resource. Clearly the prevalence of dormancy and the presence of facultative anaerobes mean that communities can switch rapidly from a widespread reliance on oxygen to the use of alternative resources for respiration.

APPLICATION 3.5 Permafrost, methanogenic anaerobic respiration and global warming

As the earth warms (see Section 22.2) regions of permafrost near the poles (where the soil remains frozen, year-round, for at least two consecutive years, see Section 1.5) are thawing. This is leading to a transition in these regions initially to ‘palsa’ habitats – mounds in the landscape supporting lichens and low shrubs – then to partly thawed bogs dominated by mosses (*Sphagnum* spp.), and then to fully thawed mires dominated by sedges (e.g. *Eriophorum* spp). This transition itself has potential implications for global warming, since it involves a shift from CO_2-emitting palsas to mires and fens that take up CO_2 but emit methane, a more potent greenhouse gas. High-methane emitting fen habitats contribute seven times as much greenhouse impact as palsa, per unit area (McCalley *et al.*, 2014). Our understanding of the roles played by the microbial communities of the soils in these habitats remains poor. But this is likely to be crucial if we wish to predict the trajectory of the positive feedback loop through which warming leads to thawing, leading to methane emission, more warming, more thawing, and so on. (In Section 17.3 we discuss permafrost as an example of an ecosystem that, on thawing, can pass a ‘tipping point’, shifting it from one regime to another.)

Microbes that produce methane as a respiratory by-product are Archaea, not bacteria. Most are hydrogenotrophic, using hydrogen as an electron acceptor. However, there is another smaller, but important group that are acetoclastic, cleaving acetate into methane and CO_2, and the methane produced by the two groups can be distinguished by characteristic isotopic signatures. Over a natural gradient of thawing in northern Sweden, methane emissions were greater from fully thawed mires than from partly thawed bogs, but were also more dominated by acetoclastic methanogens (Figure 3.28). Crucially, this shifting balance was associated in turn with variation in the ratio of methane-to-CO_2 production from anaerobic respiration (much higher from mires than from bogs) with consequences in turn for the models currently being used to predict future climate change, which typically assume the fraction of anaerobically metabolised carbon that becomes methane to be fixed (McCalley *et al.*, 2014). Results like those in Figure 3.28 therefore throw doubt on the validity of this simplifying assumption and press the case for further work on the dynamics of anaerobic resource use in these rapidly changing systems.

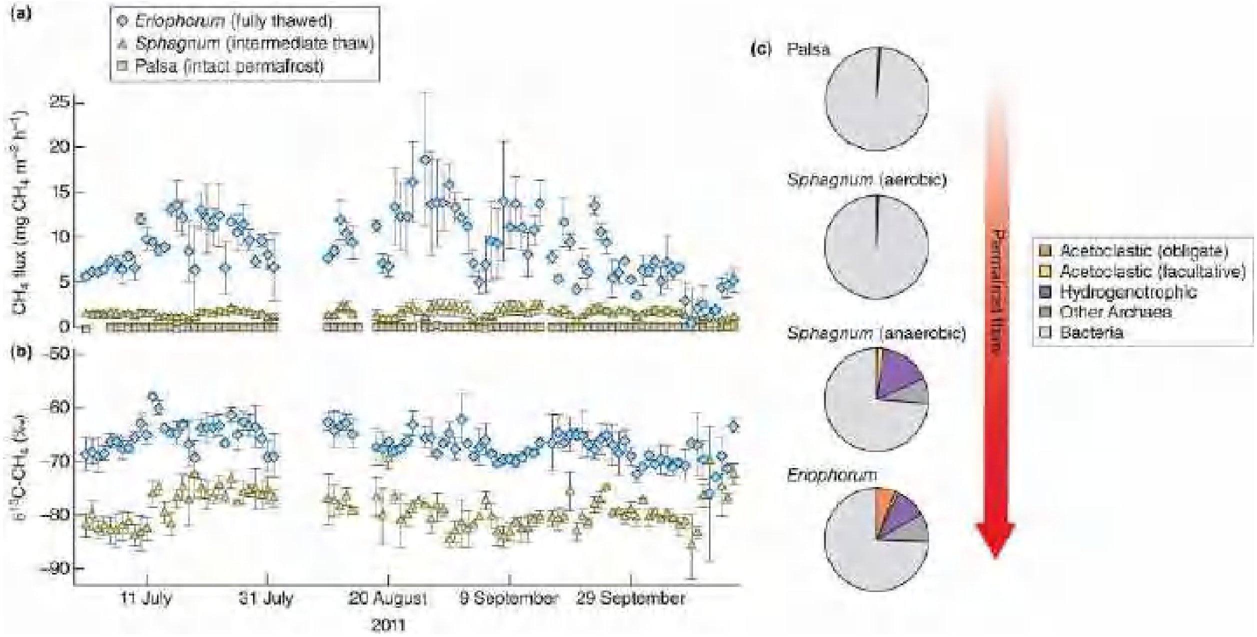


Figure 3.28 **Methane production increases when permafrost thaws, and its microbial origins change.** (a) Emissions of methane (CH_4), over time, at sites in northern Sweden at various stages of thawing from permafrost, as indicated. Bars are SEs. (b) The isotopic signatures of those methane emissions, $\delta^{13}C\text{-}CH_4$, measured as the relative difference in the ratio of ^{13}C to ^{12}C in the methane, compared with an international standard material, expressed as parts per thousand. Bars are SEs. (c) The composition of the microbial community in each case as inferred from those isotopic signatures, subdivided into bacteria and Archaea further subdivided into hydrogenotrophic and acetoclastic methanogens, and ‘others’.

Source: After McCalley *et al.* (2014).

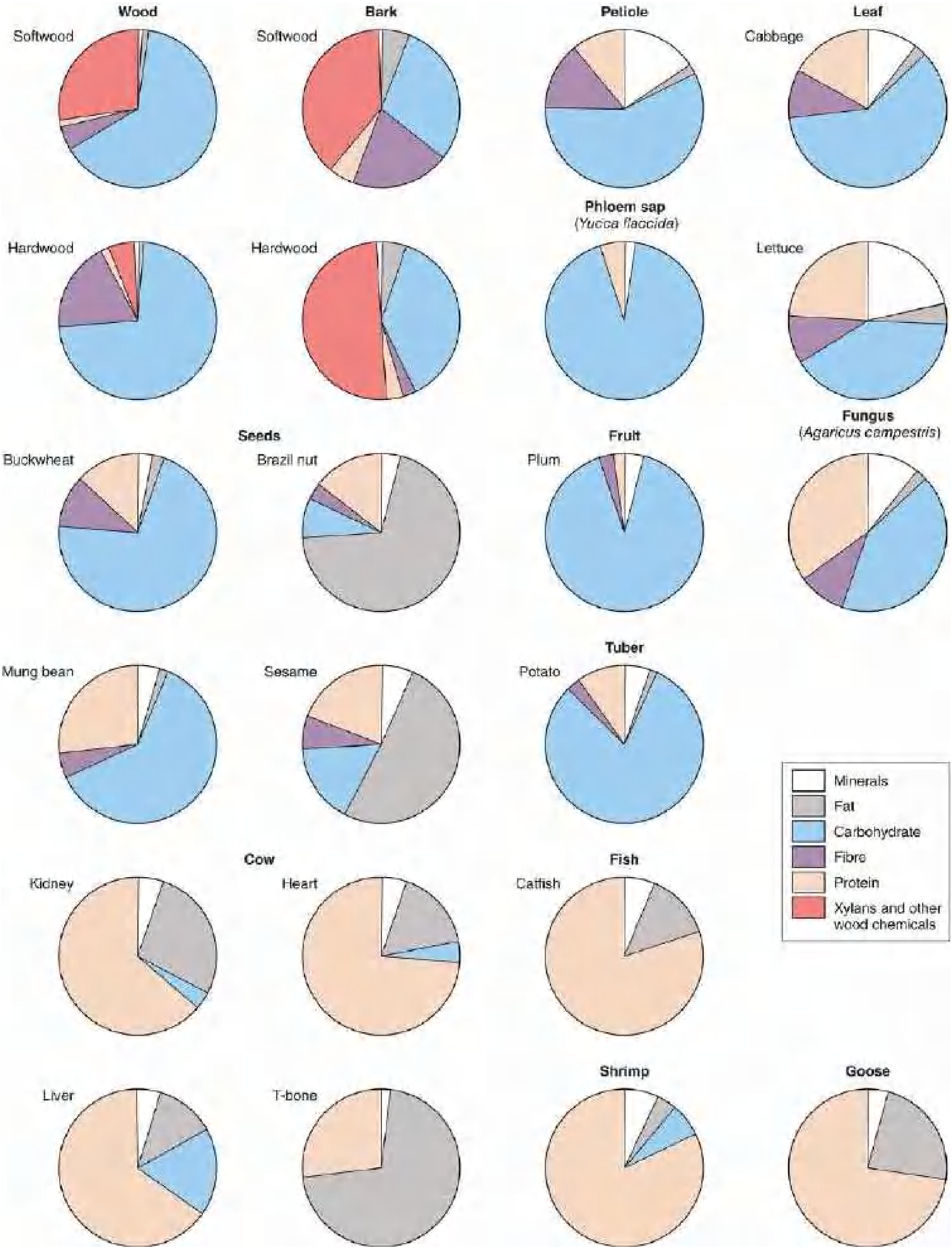
3.7 Organisms as food resources

predators, grazers and parasites

‘True’ predators predictably kill their prey. Examples include a mountain lion consuming a rabbit but also consumers that we may not refer to as predators in everyday speech: a water flea consuming phytoplankton cells, a squirrel eating an acorn (both herbivorous predators), and even a pitcher plant drowning a mosquito. *Grazing* can also be regarded as a type of predation, but the food (prey) organism is not killed. Only part of the prey is taken, leaving the remainder to live on with the potential to reproduce or regenerate. Also, grazers feed on (or from) many prey during their lifetime. Cattle and sheep are grazers of plants, but blood-sucking flies, for example, are carnivorous grazers. True predation and grazing are discussed in detail in Chapter 9. *Parasitism*, too, is a form of predation in which the consumer usually does not kill its food organism, but unlike a grazer, a parasite feeds from only one or a very few host organisms in its lifetime. Chapter 12 is devoted to parasitism.

specialists and generalists

An important distinction amongst animal consumers is whether they are specialised or generalised in their diet. Generalists (*polyphagous* species) take a wide variety of prey species, though they very often have clear preferences and a rank order of what they will choose when there are alternatives available. Some specialists consume only particular parts of their prey though they range over a number of species. This is most common among herbivores because, as we saw in Figure 3.26 and shall see again in Figure 3.29, different parts of plants are quite different in their composition. Thus, many birds specialise in eating seeds though they are seldom restricted to a particular species. Other specialists, however, may feed on only a narrow range of closely related species or even just a single species (when they are said to be *monophagous*). Examples are caterpillars of the cinnabar moth (which eat the leaves, flower buds and very young stems of species of ragwort, *Senecio*) and many species of host-specific parasites.



[Figure 3.29](#) The composition of various plant parts and of the bodies of animals that serve as food resources for other organisms.

Data from various sources.

the importance of lifespan

Many of the resource-use patterns found among animals reflect the different lifespans of the consumer and what it consumes. Individuals of long-lived species are likely to be generalists: they cannot depend on one food resource being available throughout their life. Specialisation is increasingly likely if a consumer has a short lifespan. Evolutionary forces can then shape the timing of the consumer's food demands to match the timetable of its prey. Specialisation also allows the evolution of structures that make it possible to deal very efficiently with particular resources. This is especially the case with mouthparts. Darwin's hawkmoth, *Xanthopan morganii praedicta*, with its 20 cm long proboscis, is alone in being able to take nectar and pollen from the Madagascan orchid, *Angraecum sesquipidale*, with its near-30 cm long nectary. (It is called Darwin's hawkmoth, because Charles Darwin predicted its existence on seeing the flower 20 years before the moth itself was discovered.) This can be interpreted as an exquisite product of the evolutionary process that has given the moth access to a valuable food resource – or as an example of the ever-deepening rut of specialisation that has constrained what the moth can feed on. The more specialised the food resource required by an organism, the more it is constrained to live in patches of that resource or to spend time and energy in searching for it among a mixture of resources. This is one of the costs of specialisation. We return to food preferences and diet widths in Section 9.2.

3.7.1 The nutritional contents of plants and animals and their extraction

C : N ratios in animals and plants

As a 'package' of resources, the body of a green plant is quite different from the body of an animal. This has a marked effect on the value of these resources as potential food (Figure 3.29). The most important contrast is that plant cells are bounded by walls of cellulose, lignin and/or other structural materials. It is these cell walls that give plant material its high fibre content. The presence of cell walls is also largely responsible for the high fixed carbon content of plant tissues and the high ratio of carbon to other important elements. For example, the carbon : nitrogen (C : N) ratio of plant tissues commonly exceeds 40 : 1.

In contrast, the C : N ratios in bacteria, fungi and animals are approximately 10 : 1. When plant parts are decomposed, the microbes multiplying on the decaying plant withdraw nitrogen and other mineral resources from their surroundings and build them into their own microbial bodies. Thus, plant material with a high carbon content is converted to microbial bodies with a relatively low carbon content. For this reason, and because microbial tissue is more readily digested and assimilated, plant detritus that has been richly colonised by microorganisms is generally preferred by detritivorous animals. Unlike plants, animal tissues contain no structural carbohydrate or fibre component but are rich in fat and, in particular, protein.

different plant parts represent very different resources …

The various parts of a plant, however, have very different compositions (Figure 3.29) and so offer quite different resources. Bark, for example, is largely composed of dead cells with corky and lignified walls and is quite useless as a food for most herbivores (even species of 'bark beetle' specialise on the nutritious cambium layer just beneath the bark, rather than on the bark itself). The richest concentrations of plant proteins (and hence of nitrogen) are in the meristems in the buds at shoot apices and in leaf axils. Not surprisingly, these are usually heavily protected with bud scales and defended from herbivores by thorns and spines. Seeds are usually dried, packaged reserves rich in starch or oils as well as specialised storage proteins. And the very sugary and fleshy fruits are resources provided by the plant as 'payment' to the animals that disperse the seeds. Very little of the plants’ nitrogen is spent on these rewards.

The dietary value of different tissues and organs is so different that it is no surprise to find that most small herbivores are specialists – not only on particular species or plant groups, but on particular plant parts: meristems, leaves, roots, stems, etc. The smaller the herbivore, the finer is the scale of heterogeneity of the plant on which it may specialise. Extreme examples can be found in the larvae of various species of oak gall wasps, some of which may specialise on young leaves, some on old leaves, some on vegetative buds, some on male flowers and others on root tissues.

... but the composition of all herbivores is remarkably similar

By contrast, the composition of the bodies of different herbivores is remarkably similar. In terms of the content of protein, carbohydrate, fat, water and minerals per gram, there is very little to choose between a diet of caterpillars, cod or venison. The packages may be differently parcelled (and the taste may be different), but the contents are essentially the same. Carnivores, then, are not faced with problems of digestion (and they vary rather little in their digestive apparatus), but rather with difficulties in finding, catching and handling their prey (see Chapter 9).

Differences in detail aside, herbivores that consume living plant material – and saprotrophs that consume dead plant material – all utilise a food resource that is rich in carbon and poor in protein. Hence, the transition from plant to consumer involves a massive burning off of carbon as the C : N ratio is lowered. The main waste products of organisms that consume plants are carbon-rich compounds: CO_2, fibre, and in the case of aphids, for example, carbon-rich honeydew dripping from infested trees. By contrast, the greater part of the energy requirements of carnivores is obtained from the protein and fats of their prey, and their main excretory products are in consequence nitrogenous. The crucial role of gut microbes in digestion for all animals, but for herbivores especially, is discussed in Section 13.6.

ecological stoichiometry

These ratios in the composition of organisms, and the changes in these ratios from one trophic level to the next, are the realm of *ecological stoichiometry* (Hessen *et al*., 2013), defined as ‘The balance of multiple chemical substances in ecological interactions and processes, or the study of this balance’ (Sterner & Elser, 2002) – see also Sections 11.2.5 and 20.4.3. The approach can be traced back to what is commonly called Liebig’s ‘law of the minimum’ (Liebig, 1840), which states that organismal growth rates are limited by whichever element has the lowest rate of environmental supply relative to the organisms’ demands. A more direct connection still can be traced to the work of Alfred C. Redfield going back to the 1930s, focusing especially on the apparent constancy of the ratio of nitrate to phosphate (around 16 : 1) in the biomass of phytoplankton and in seawater (Gruber & Deutsch, 2013). Indeed, ecological stoichiometry has always been most influential in aquatic, especially marine biogeochemistry, where the focus is usually on nitrogen, phosphorus and carbon, though other elements, for example iron, also have key roles to play (Tagliabue *et al*., 2017).

cellulases, which most animals lack

The large amounts of fixed carbon in plant materials mean that they are potentially rich sources of energy. Other components of the diet (e.g. nitrogen) are more likely to be limiting. Yet most of that energy is only directly available to consumers if they have enzymes capable of mobilising cellulose and lignins. An increasing number of species, especially insects, have been shown to have these enzymes themselves (Watanabe & Tokuda, 2010), but the overwhelming majority of species in both the plant and animal kingdoms lack them, the latter relying instead on the cellulases produced by gut-inhabiting, cellulolytic prokaryotes with which they form intimate, ‘mutualistic’ relationships, discussed further, for both vertebrate and invertebrate herbivores, in Chapter 13.

Because most animals lack cellulases, the cell wall material of plants hinders the access of digestive enzymes to the contents of plant cells. The acts of chewing by the grazing mammal, cooking by humans and grinding in the gizzard of birds allow digestive enzymes to reach cell contents more easily. The carnivore, by contrast, can more safely gulp its food.

Of course, one big difference between the resources of autotrophs and those of heterotrophs, at least those that consume living prey, is that the resources of the heterotrophs can fight back – on both ecological and evolutionary timescales. We pick up the story of prey defence in Chapter 9.

3.8 A classification of resources, and the ecological niche

We have seen that every plant requires many distinct resources to complete its life cycle, and most plants require the same set of resources, although in subtly different proportions. Each of these resources has to be obtained independently of the others, and often by quite different uptake mechanisms – some as ions (potassium), some as molecules (CO_2), some in solution, some as gases. Carbon cannot be replaced by nitrogen, nor phosphorus by potassium. Nitrogen can be taken up by most plants as either nitrate or ammonium ions, but there is no substitute for nitrogen itself. In complete contrast, for many carnivores, most prey of about the same size are wholly interchangeable as articles of diet. This contrast between resources that are individually *essential* for an organism, and those that are *substitutable*, can be extended into a classification of resources taken in pairs (Figure 3.30).

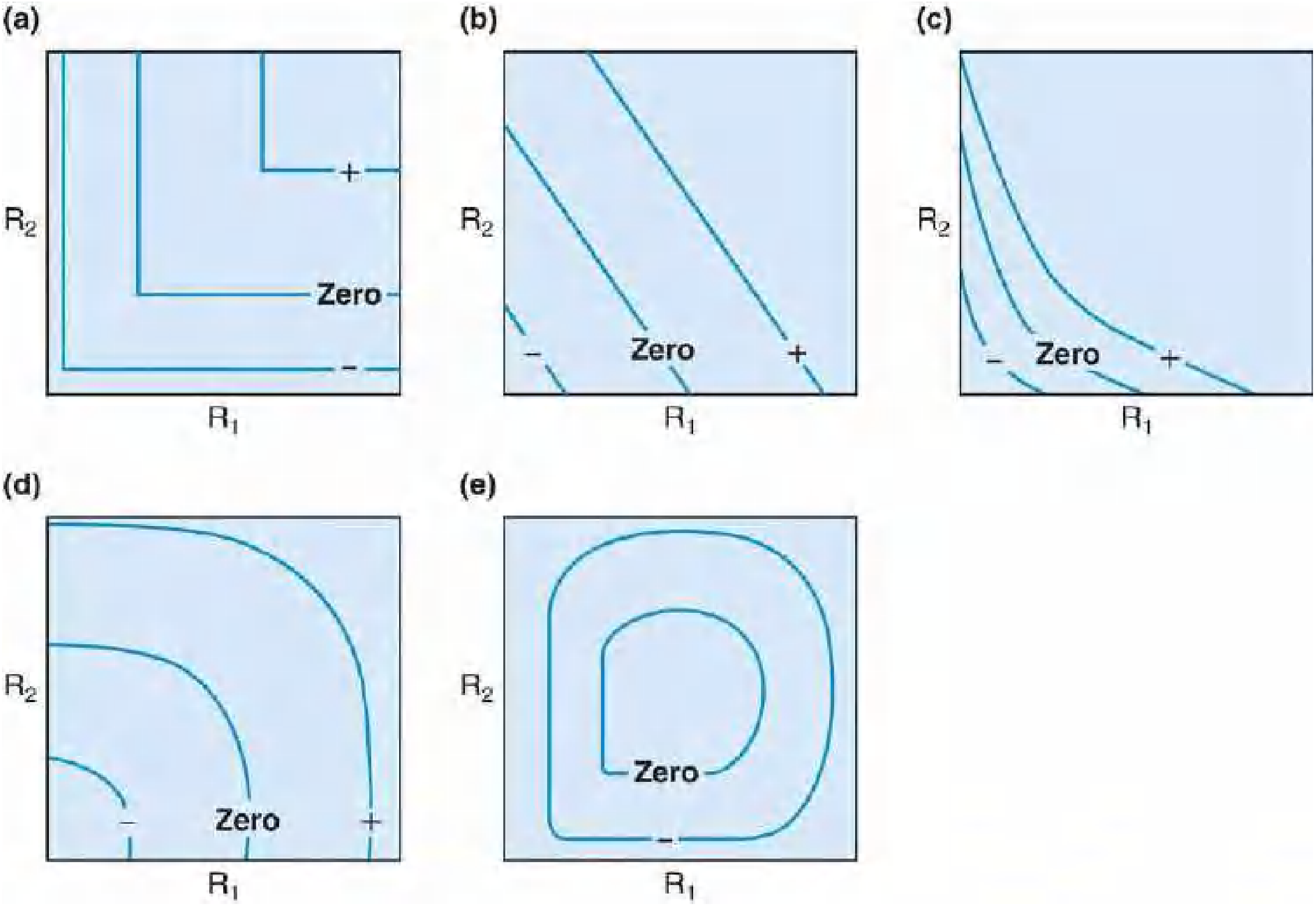


Figure 3.30 **Resource-dependent growth isoclines.** Each of the growth isoclines represents the amounts of two resources (R_1 and R_2) that would have to exist in a habitat for a population to have a given growth rate. Because this rate increases with resource availability, isoclines further from the origin represent higher population growth rates: isocline – has a negative growth rate, isocline zero a zero growth rate and isocline + a positive growth rate. In the respective figures, resources are (a) essential, (b) perfectly substitutable, (c) complementary, (d) antagonistic and (e) inhibitory.

Source: After Tilman (1982).

zero net growth isoclines

In this classification, the concentration or quantity of one resource is plotted on the x-axis, and that of the other resource on the y-axis. We know that different combinations of the two resources will support different growth rates for the organism in question (this can be individual growth or population growth). Thus, we can join together points (i.e. combinations of resources) with the same growth rates, and these are therefore contours or ‘isoclines’ of equal growth. In Figure 3.30, line ‘zero’ in each case is an isocline of zero net growth: each of the resource combinations on these lines allows the organism just to maintain itself, neither increasing nor decreasing. The ‘–’ isoclines, then, with less resources than the zero line, join combinations giving the same negative growth rate; whilst the ‘+’ isoclines, with more resources than the zero line, join combinations giving the same positive growth rate. As we shall see, the shapes of the isoclines vary with the nature of the resources.

3.8.1 Categories of resources

essential resources

Two resources are said to be *essential* when neither can substitute for the other. This is denoted in Figure 3.30a by the isoclines running parallel to both axes. They do so because the amount available of one resource defines a maximum possible growth rate, irrespective of the amount of the other resource. This growth rate is achieved unless the amount available of the other resource defines an even lower growth rate. Generally, therefore, growth rate will be determined by the resource in most limited supply. This will be true for nitrogen and potassium as resources in the growth of green plants, and for two host species in the life of a parasite that must alternate between them (see Chapter 12).

perfectly substitutable resources

Two resources are said to be *perfectly substitutable* when either can wholly replace the other. This will be true for seeds of wheat or barley in the diet of a farmyard chicken, or for zebra and gazelle in the diet of a lion. Note that we do not imply that the two resources are as good as each other. This feature (perfectly substitutable but not necessarily as good as each other) is included in Figure 3.30b by the isoclines having slopes that do not cut both axes at the same distance from the origin. Thus, in Figure 3.30b, in the absence of resource 2, the organism needs relatively little of resource 1, but in the absence of resource 1 it needs a relatively large amount of resource 2.

complementary resources

Substitutable resources are defined as *complementary* if the isoclines bow inwards towards the origin (Figure 3.30c). This shape means that a species requires less of two resources when taken together than when consumed separately. A good example is human vegetarians combining beans and rice in their diet. The beans are rich in lysine, an essential amino acid poorly represented in rice, whilst rice is rich in sulphur-containing amino acids that are present only in low abundance in beans.

antagonistic resources

By contrast, a pair of substitutable resources with isoclines that bow away from the origin are defined as *antagonistic* (Figure 3.30d). The shape indicates that a species requires proportionately more resource to maintain a given rate of increase when two resources are consumed together than when consumed separately. Though probably rare, this could arise, for example, if the resources contain different toxic compounds that act synergistically (more than just additively) on their consumer. For example, d,l-pipecolic acid and djenkolic acid (two defensive chemicals found in certain seeds) had no significant effect on the growth of the seed-eating larva of a bruchid beetle if consumed separately, but they had a pronounced effect if taken together (Janzen *et al*., 1977). Whenever resources are substitutable, whether or not they are perfectly substitutable, the growth rate is determined by their joint availability.

inhibitory resources

Finally, Figure 3.30e illustrates the phenomenon of *inhibition* at high resource levels for a pair of essential resources: resources that are essential but become damaging when in excess. CO_2, water and mineral nutrients such as iron are all required for photosynthesis, but each is lethal in excess. Similarly, light leads to increased growth rates in plants through a broad range of intensities, but can inhibit growth at very high intensities. In such cases, the isoclines form closed curves because growth decreases with an increase in resources at very high levels.

3.8.2 Resource dimensions of the ecological niche

In Chapter 2 we developed the concept of the ecological niche as an n-dimensional hypervolume. This defines the limits within which a given species can survive and reproduce for a number (n) of environmental factors, including both conditions and resources. Note, therefore, that the zero growth isoclines in Figure 3.30 define niche boundaries in two dimensions. Resource combinations to one side of this zero isocline allow the organisms to thrive – but to the other side of the line the organisms decline.

The resource dimensions of a species' niche can sometimes be represented in a manner similar to that adopted for conditions, with lower and upper limits within which a species can thrive. Thus, a predator may only be able to detect and handle prey between lower and upper limits of size. For other resources, such as mineral nutrients for plants, there may be a lower limit below which individuals cannot grow and reproduce but an upper limit may not exist (Figure 3.30a–d). However, many resources must be viewed as discrete entities rather than continuous variables. Larvae of butterflies in the genus *Heliconius* require *Passiflora* leaves to eat; those of the monarch butterfly specialise on plants in the milkweed family; and various species of animals require nest sites with particular specifications. These resource requirements cannot be arranged along a continuous graph axis labelled, for example, 'food plant species'. Instead, the food plant or nest-site dimension of their niches needs to be defined simply by a restricted list of the appropriate resources.

Together, then, conditions and resources define a species' niche. In the next chapter, we will look in more detail at the most fundamental responses of organisms to those conditions and resources: their patterns of growth, survival and reproduction.

3.9 A metabolic theory of ecology

Resources and conditions are also important insofar as they determine the metabolic rates of individuals, which determine the levels of resources available to those individuals for reproduction, growth and so on, which in turn influence their life histories, their abundance and indeed all of the processes we discuss in subsequent chapters. This perspective has generated interest in a 'metabolic theory of ecology' (Brown *et al.*, 2004). At the heart of metabolic theory are the effects on metabolic rate of temperature, and in particular of size (most often, body mass). We examined the effects of temperature in Section 2.3. We turn now to size.

metabolic scaling

The most fundamental point, perhaps, is that life is typically faster for small organisms than it is for large ones – metabolising at greater rates, and maturing and dying sooner. For example, per gram of body mass, a resting mouse metabolises about 20 times faster than an elephant. There are exceptions to this pattern, as we'll see below, but the more general rule is very widespread. Putting this more formally, we can say that the rate of a metabolic process, Y, varies with size according to the equation

$$Y = Y_0 M^b, \tag{3.1}$$

where Y_0 is referred to as the normalisation constant, reflecting the typical rate for the metabolic process concerned, M is the organism's mass and b is the so-called allometric exponent. Equation 3.1 may be said to describe *metabolic scaling*. Taking logs of both sides, we get:

$$\log Y = \log Y_0 + b \log M. \tag{3.2}$$

allometric relationships

The slowing down of metabolism at larger sizes is reflected in values of b that are less than 1. This makes them *allometric relationships* – that is, relationships in which a physical or (in this case) physiological property of an organism changes relative to the size of the organism, rather than changing in direct proportion to the changing size. That would be an *iso*metric relationship, and in that case b would be equal to 1. We can see from Equation 3.2 that b is the slope of the line when metabolism is plotted against body mass on logarithmic scales.

Examples are shown in Figure 3.31 for a wide range of taxonomic groups. As mass increases, temperature-corrected metabolic rate increases less than proportionately; $b = 0.71$, with, apparently, little variation between groups. We discuss actual values of b below. The intercept in these plots is $\log Y_0$, such that the value of Y_0 locates the relationship vertically within the plot, telling us about the absolute rate of metabolism at a given body size (and temperature). In this case, there is variation in absolute rate between the groups, despite the relationship with size being apparently the same. Such allometric relationships can be ontogenetic (changes occurring as an organism grows) or phylogenetic (changes that are apparent when related taxa of different sizes are compared).

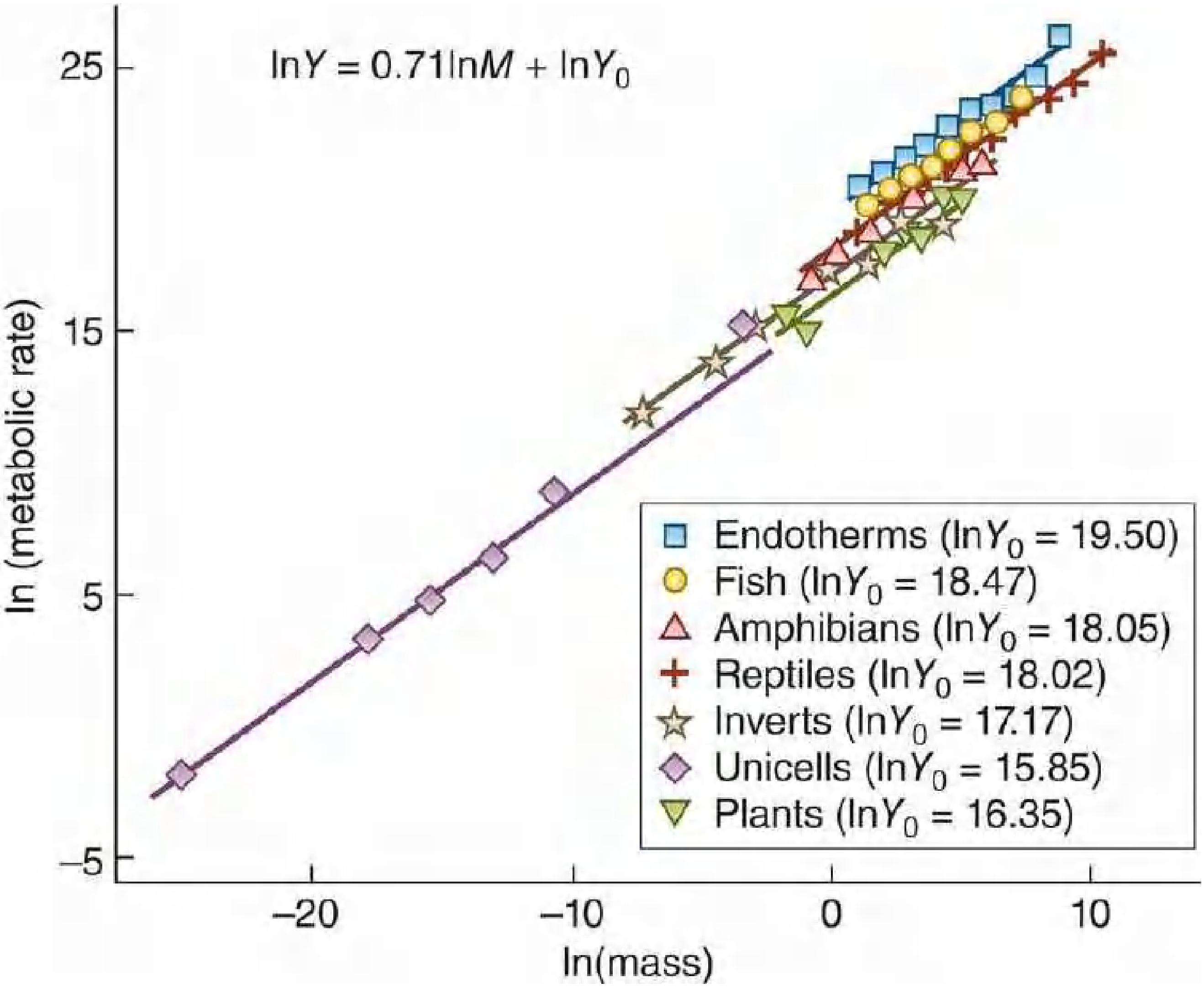


Figure 3.31 Metabolic scaling: the relationship between metabolic rate (*Y*, watts) and body mass (*M*, g) for a variety of organisms, as indicated, on logarithmic scales. The analysis sought a single slope but allowed the intercepts of different groups to vary. For clarity, the data points shown are the averages for mass-classes. The metabolic rates have been temperature-corrected to ensure different studies are comparable (see Equation 3.3), but this has been omitted from the equation, top right, for clarity.

Source: After Brown *et al.* (2004).

rates per unit mass, and times

Note that if individual metabolic rate scales with individual mass with an exponent of *b*, then metabolic rate *per unit mass* (that is, the metabolic rate of a gram of tissue) will scale with an exponent of *b*–1 (simply divide both sides of Equation 3.1 by *M*). Similarly, the time taken to complete a process (for example to reach maturity) will scale with an exponent of 1–*b*, because these times are the reciprocals of rates per unit mass (the reciprocal exponent of *b*–1 is –(*b*–1) = 1–*b*). It may sometimes be more appropriate to examine the metabolic scaling of times or rates per unit mass than individual rates.

transport or demand?

The conventional view, as we discuss later, is that plots like Figure 3.31 arise because size imposes constraints on rates of supply (of oxygen, nutrients and so on), and of transport of materials generally within the organism, which determine metabolic rates. Metabolic rates then constrain the resources available for reproduction, growth and so on (Brown *et al*., 2004). An alternative viewpoint, however, is that size is closely co-adapted with investment in reproduction and growth (and hence with the demands for these), with rates and means of transport evolving to satisfy those demands (Harrison, 2017). Size in this case is an integral part of an organism's overall life history, which has evolved to match its environment. From this perspective, a mouse, for example, metabolises rapidly, and an elephant metabolises more slowly, to fuel their respective life histories, of which size is a part. Each has evolved transport networks sufficient to service their metabolism. We discuss life history evolution more fully (and why mice might be fast, and elephants slow) in Chapter 7. The alternative viewpoints are summarised in Figure 3.32. Bringing the two together, we may conclude that there is really no single driver of metabolic scaling. Rather, we should see life histories and their metabolic demands as being co-adjusted with transport system design (Glazier, 2014).

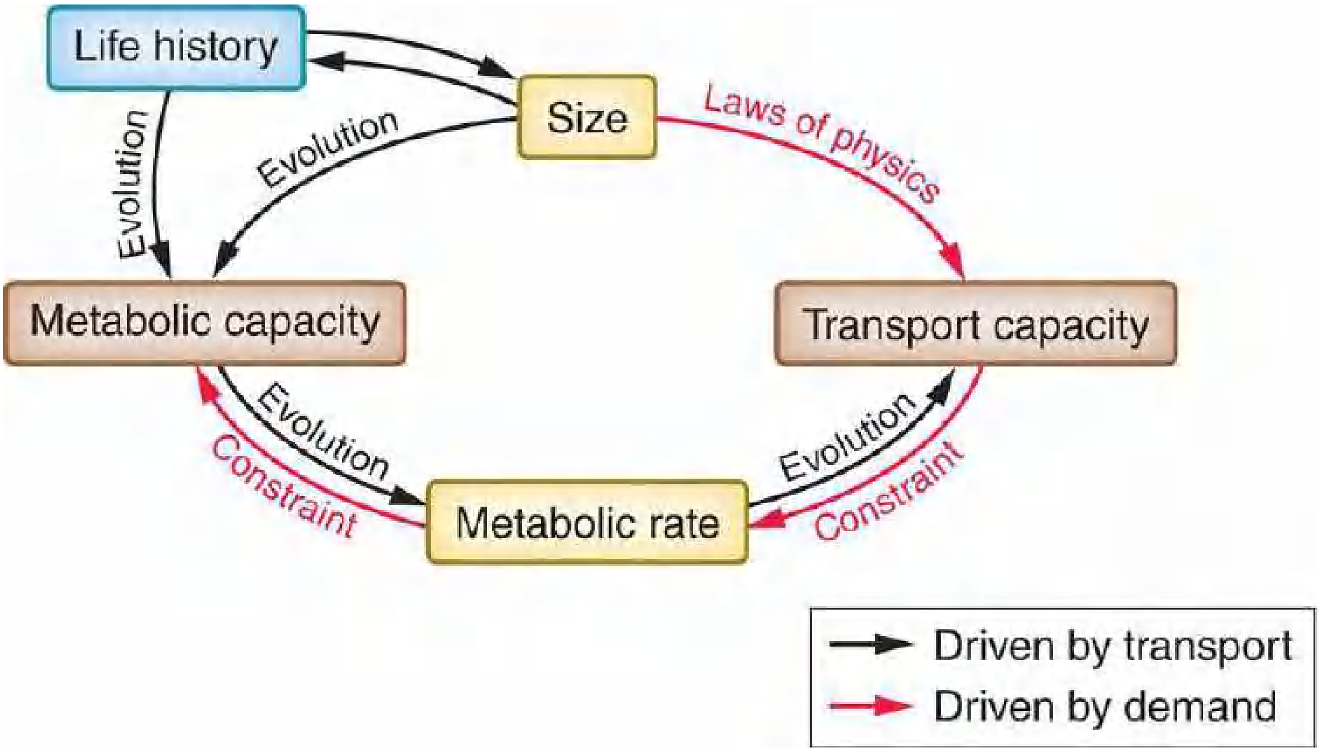


Figure 3.32 Schematic representation of the two main approaches to the relationship between metabolic rate and size. In one, indicated by the red arrows, size sets limits to rates of supply of nutrients (and of transport generally, for example of waste product excretion) and these supply routes or transport networks constrain an organism's metabolic rate. In the other approach, indicated by black arrows, an organism's life history, of which size is an integral part, evolves to match its environment, and the metabolic rate evolves to satisfy the demands (the metabolic capacity) of an organism with such a life history. The transport network, in turn, evolves to satisfy the demands of the metabolic rate.

temperature dependence

For a more complete account of metabolic theory, we must add the effects of temperature on metabolism to the effects of size. We saw in Section 2.3 that over a biologically realistic range of temperatures, the rate, Y, of a metabolic process is expected to increase exponentially, and this is conventionally described by the Arrhenius equation:

$$Y = Y_0 e^{-E/kT} \tag{3.3}$$

in which Y_0 is the normalisation constant, as above, E is the activation energy required for that process, k is the so-called Boltzmann's constant and T is the temperature in Kelvin (a scale starting at absolute zero, in which 0°C is 273 Kelvin, and increments are the same as in the centigrade scale). For our purposes, we need only note that as temperature increases, E/kT decreases, $-E/kT$ increases, and hence the metabolic rate increases exponentially.

Clearly, we can bring metabolism, size and temperature together and obtain

$$Y = Y_0 M^b e^{-E/kT}. \tag{3.4}$$

In practice, however, most studies focus on size and allow for temperature by dealing not with simple metabolic rates but temperature-compensated metabolic rates (see, for example, Figure 3.31). Here, too, therefore, we return to Equation 3.2 and the value of b.

a basis for metabolic scaling: SA and RTN theories

What should the allometric exponent be? As explained before, most answers to this question have focused on constraints on rates of transport. There have been two main types of theory: surface area theories (SA) and resource-transport network theories (RTN), both with histories stretching back to the 1800s (Glazier (2014); and see Glazier (2005) for a much fuller subdivision of theories). SA theories argue that the rate of any metabolic process is limited by the rate at which resources for that process can be transported in, or at which the heat or waste products generated by the process can be transported out. This transport occurs across a surface, either within the organism or between the organism and its environment, the extent of which increases with the square (power 2) of linear size – as too, therefore, does the metabolic rate. However, assuming no change in shape, mass itself increases with the cube (power 3) of linear size. Hence, the metabolic rate, rather than keeping up with this increase in mass (where b would be 1) lags behind, scaling with mass with an exponent (b) of $2/3$ or 0.67.

RTN theories, on the other hand, focus on the geometries of transport networks that would optimise the flow of nutrients being dispersed from a centralised hub to target tissues *within* an organism, or the flow of waste products carried away in an equivalent manner in the opposite direction. Derivations based on networks assumed to be of this type are more complex than the simple area-to-volume arguments applied above. However, we can ignore these details and note simply that initial attempts to derive a metabolic scaling rule based on such networks led to a b value of $3/4$ or 0.75 (West *et al.*, 1997), while subsequent elaborations confirmed this value if the velocity of flow itself scales with mass, but suggested a value closer to 0.67 if velocity does not vary significantly with mass (Banavar *et al.*, 2010). A value of 0.75 is attractive in that it conforms with an empirical estimate derived long ago by Kleiber (1932) from an analysis of metabolic rates in a number of birds and mammals. This had given rise to the so-called 'Kleiber's law', but the law had lacked a convincing theoretical underpinning before West *et al.*'s study.

a universal *b*?

Attempts like these to derive an 'expected' value for b have often been motivated by a wish to discover fundamental organising principles governing the world around us –universal rules linking metabolism to size – a single, common value of b (Brown *et al.*, 2004). Others have suggested that such generalisations may be oversimplified (Glazier, 2010, 2014). There need be no conflict between these two viewpoints. It can be valuable to have a single, simple theory that goes a long way towards explaining the patterns we see in nature. But it is also valuable to have a more complex, multifaceted theory that explains even more, including apparent exceptions to the simple rule. Similarly, when we examine data for these relationships, it can be valuable to focus on the general trend and fit a single line to the data, even if there is considerable variation around that general trend. But it is also valuable to treat that variation not as noise but as something requiring an explanation in its own right – for which a more complex model may be required.

A review of the data, overall, argues against a universal value for *b*. The analysis in Figure 3.31 suggested that a single value between 0.67 and 0.75 was appropriate for multicellular animals (metazoa), unicellular organisms and plants. However, a more detailed look suggests that metazoa do indeed have an exponent of around 0.75, but for unicellular eukaryotes (protists) the value is close to 1 (isometry) and for prokaryotes significantly greater than 1 (Figure 3.33) (DeLong *et al*., 2010). DeLong *et al*. hypothesise, with some empirical support, that the prokaryote value greater than one reflects an increase in genome size (and hence metabolic complexity) as organism mass increases; and that the protist value of one reflects a linear increase with size in ATP-synthesising (energy-generating) sites bound to membranes, which are surfaces. The metazoan value then reflects more conventional body surface or transport network constraints (DeLong *et al*., 2010).

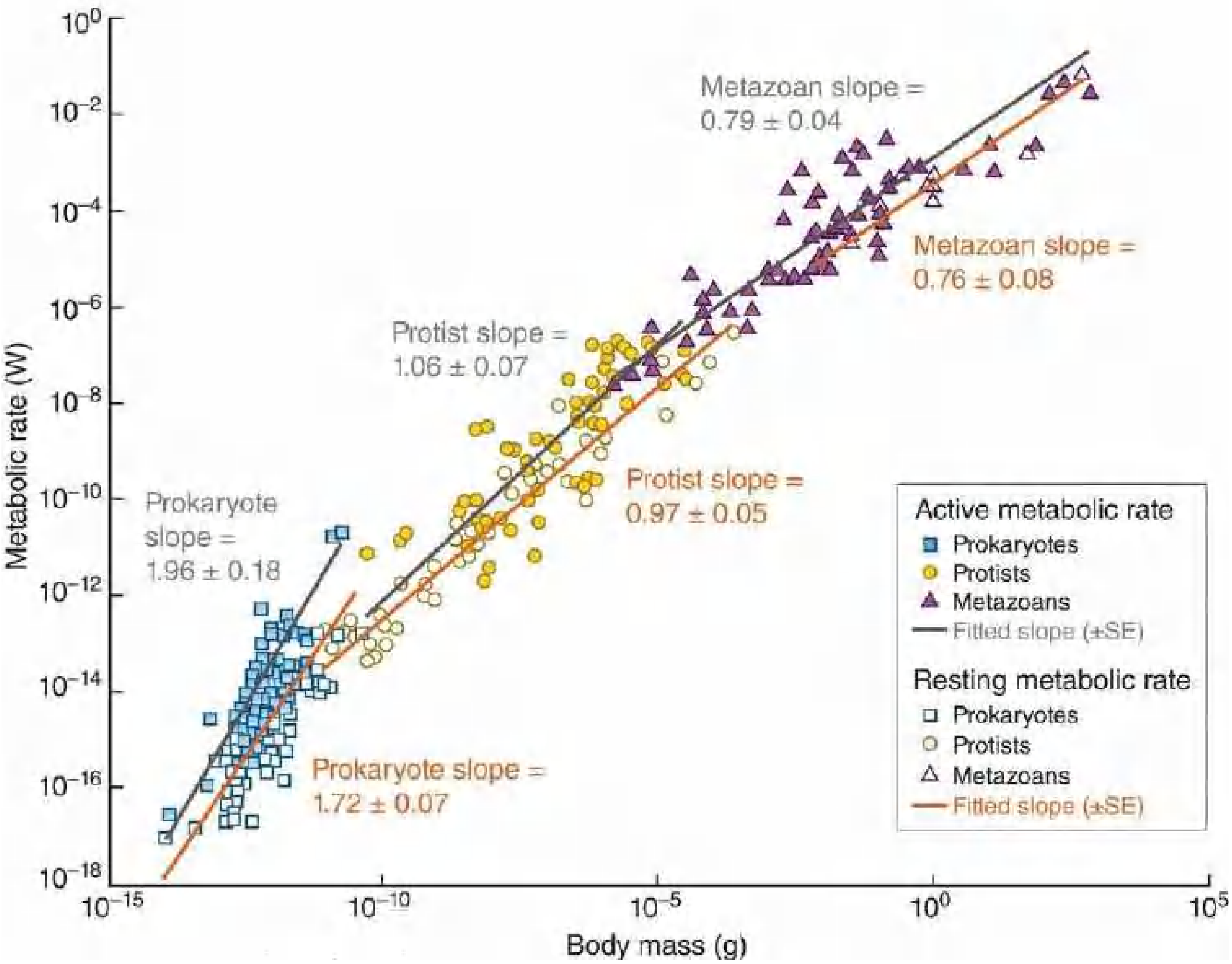


Figure 3.33 Relationships between metabolic rate and body mass for heterotrophic prokaryotes, protists and metazoans, plotted on logarithmic scales. The black lines and closed points are for active metabolic rates and the grey lines and open points for resting rates. In each case, the fitted slopes (± SE) are shown. All are significant ($P < 0.05$).

Source: After DeLong *et al*. (2010).

As another example, the allometric exponent in plants appears to be consistently different between, on the one hand, seedlings and the smallest plants, and on the other, larger saplings and adult plants – close to 1 for plants with masses less than 1 g and converging to 0.75 as masses exceed around 100 g (Figure 3.34), though the particular mass values should not be taken too literally. In this case, the authors hypothesise that for larger plants, the photosynthetic machinery is distributed across surfaces (principally of leaves), whereas for smaller plants most or all of the tissue (and hence a volume) is photosynthetically active (Mori *et al*., 2010). A curvilinear relationship has also been proposed for mammals, but this time with the opposite curvature, starting at 0.57 and rising to 0.87 (Kolokotrones *et al.,* 2010).

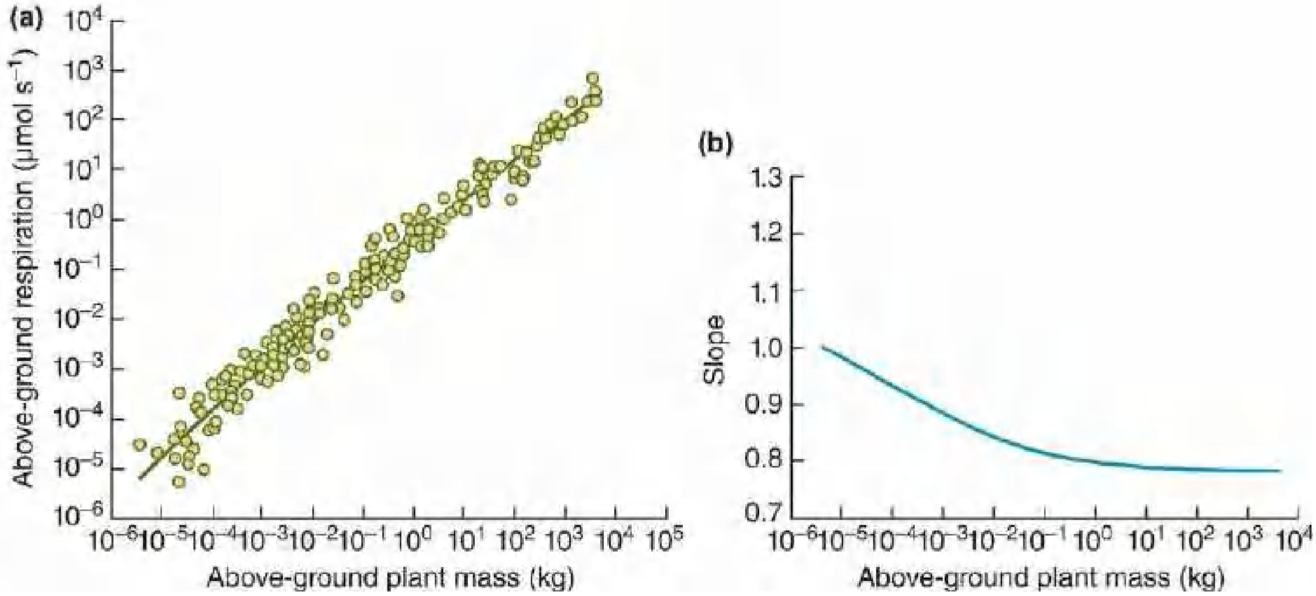


Figure 3.34 The allometric exponent of metabolism in plants decreases with plant size. (a) The relationship between temperature-adjusted respiration rate and above-ground plant mass across a wide range of masses on logarithmic scales. A curvilinear power function was fitted to the data, the changing slope of which is shown in (b).

Source: After Mori *et al*. (2010).

Note, to add a further perspective, that alongside SA and RTN theories, there is an equally long tradition of emphasising body composition as a driver of metabolic rate, with some organisms having a much higher proportion of structural, low-metabolising tissue than others (see Glazier, 2014); and other studies again have emphasised the importance of changing shape (which the simpler theories assume remains constant) and show that the shifting patterns of metabolic rates with shape support the SA but not the RTN theories of metabolic scaling (Hirst *et al*., 2016). However, particular values of *b*, and the truth or otherwise of the hypotheses proposed to explain them, are less important than the more general point that an organism's rate of metabolism reflects a whole host of constraints and demands, and different factors will therefore dominate in their effects in different organisms, and at different times, and *b* will therefore vary. It is unwise to seek a single, universal value for *b*, or a single, simple basis for all metabolic scaling. The key message, picked up again in later chapters, is that the scaling of metabolic rate plays a key role in the dynamics at all levels of ecological organisation, from the individual to the whole community.

Chapter 4
Matters of Life and Death

4.1 An ecological fact of life

Much of ecology is concerned with numbers and changes in numbers. Which species are common and which species rare? Why? Which species remain constant in abundance and which vary? Why? How can we reduce the numbers of a pest? Or prevent reductions in the numbers of a rare but valued species? At the heart of all such questions, there is a fundamental ecological fact of life:

$$N_{\text{now}} = N_{\text{then}} + B - D + I - E. \tag{4.1}$$

That is, the numbers of a particular species presently occupying a site of interest (N_{now}) is equal to the numbers previously there (N_{then}), plus the number of births between then and now (B), minus the number of deaths (D), plus the number of immigrants (I), minus the number of emigrants (E).

This defines the main aim of ecology: to describe, explain and understand the distribution and abundance of organisms. Ecologists are interested in the number of individuals, the distributions of individuals, the *demographic processes* (birth, death and migration – also often referred to as *vital rates*) that influence these, and the ways in which these demographic processes are themselves influenced by environmental factors.

4.2 Individuals

4.2.1 Unitary and modular organisms

individuals differ

Among the simplifications in our ecological fact of life is the implicit assumption that all individuals are alike (since all we need do is count them), which is patently false on a number of counts. First, almost all species pass through a number of stages in their life cycle: eggs, larvae, pupae and adults in many insects; seeds, seedlings and photosynthesising adults in plants; and so on. The different stages are likely to be influenced by different factors and to have different rates of migration, death and, of course, reproduction. Second, even within a stage, individuals can differ in 'quality' or 'condition'. The most obvious aspect of this is size, but it is also common, for example, for individuals to differ in the amount of stored reserves they possess.

unitary and modular organisms

Uniformity amongst individuals is especially unlikely when organisms are *modular* rather than *unitary*. In unitary organisms – mammals, birds, insects and so on – form and the succession of phases in the life cycle are predictable and ‘determinate’. That is, all dogs have four legs and each squid has two eyes, and dogs and squid that lived longer would not develop more of them. Similarly, we humans pass through an embryonic stage of nine months, a growth phase of around 18 years incorporating a prereproductive phase of 12 or so years, a reproductive phase lasting perhaps 30 years in females and rather longer in males, followed finally by a phase of senescence. Death can intervene at any time, but for surviving individuals the succession of phases, and even mostly the timing of phases, is, like form, entirely predictable.

But none of this is so simple for modular organisms such as trees, shrubs and herbs, chain-forming bacteria and algae, corals, sponges, and very many other marine invertebrates (Figure 4.1). These grow by the repeated production of ‘ modules’ (leaves, coral polyps, etc.) and almost always form a branching structure. Most, following a juvenile phase, are rooted or fixed, not motile, and both their structure and their precise programme of development are not predictable but ‘indeterminate’. After several years’ growth, depending on circumstances, the same germinating tree seed could either give rise to a stunted sapling with a handful of leaves or a thriving young tree with many branches and thousands of leaves. It is modularity and the differing birth and death rates of modules that give rise to this plasticity. Reviews of the growth, form, ecology and evolution of a wide range of modular organisms may be found in Harper *et al*. (1986), Hughes (1989) and Collado-Vides (2001).

(a)

(b)

(c)

(d)

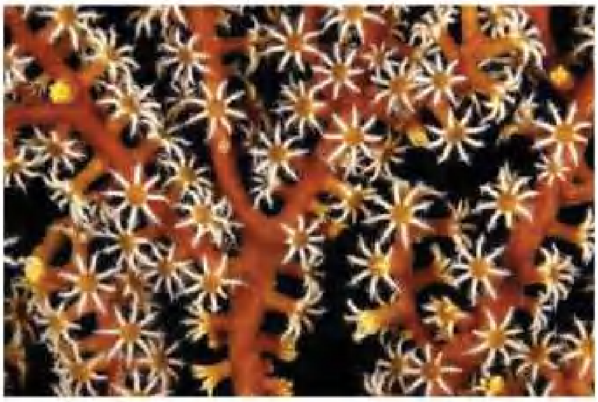


Figure 4.1 Modular plants (left) and animals (right) show the underlying parallels in the various ways they may be constructed. (a) Modular organisms that fall to pieces as they grow: duckweed (*Lemna* sp.) and *Hydra* sp. (b) Freely branching organisms in which the modules are displayed as individuals on 'stalks': a vegetative shoot of a higher plant (*Lonicera japonica*) with leaves (feeding modules) and a flowering shoot, and a hydrozoa (*Extopleura larynx*) bearing both feeding and reproductive modules. (c) Stoloniferous organisms in which colonies spread laterally and remain joined by 'stolons' or rhizomes: strawberry plants (*Fragaria*) reproducing by means of runners, and a colony of the hydroid *Tubularia crocea*. (d) Tightly packed colonies of modules: a tussock of yellow marsh saxifrage (*Saxifraga hirculus*), and a segment of the sea fan *Acanthogorgia*. (e) Modules accumulated on a long persistent, largely dead support: an oak tree (*Quercus robur*) in which the support is mainly the dead woody tissues derived from previous modules, and a gorgonian coral in which the support is mainly heavily calcified tissues from earlier modules.

what is the size of a modular population?

It follows from this that in modular organisms, the number of surviving zygotes (individuals in a genetic sense) can give only a partial and misleading impression of the 'size' of the population. Kays and Harper (1974) coined the word *genet* to describe this 'genetic individual' – the product of a zygote – and we can see that in modular organisms, the distribution and abundance of genets is important, but it is often more useful to study the distribution and abundance of modules (ramets, shoots, tillers, zooids, polyps or whatever). The amount of grass in a field available to cattle is not determined by the number of genets but by the number of leaves (modules).

4.2.2 Growth forms of modular organisms

We can see how modular organisms grow by taking higher plants as a good example. The fundamental module of construction above ground is the leaf with its axillary bud (the bud emerging where the leaf meets the stem) and the attendant section of stem. As the bud develops and grows, it produces further leaves, each bearing buds in their axils. The plant grows by accumulating these modules. At some stage in the development, a new sort of module appears, associated with reproduction (flowers in higher plants) and ultimately giving rise to new zygotes. Such specialised reproductive modules usually cease to give rise to new modules. The programme of development in modular organisms is typically determined by the proportion of modules that are allocated to different roles (e.g. to reproduction or to continued growth).

Depending on how they grow, modular organisms may broadly be divided into those that concentrate on vertical growth, and those that spread their modules laterally, over or in a substrate. Among plants that mostly extend laterally, many produce new root systems at intervals along the lateral stem: these are the rhizomatous and stoloniferous plants. The connections between the parts of such plants may die and rot away, so that the product of the original zygote becomes represented by physiologically separated parts. (Modules with the potential for separate existence are known as 'ramets'.) The most extreme examples of plants 'falling to pieces' as they grow are the many species of floating aquatics like duckweeds (*Lemna*) and the water hyacinth (*Eichhornia*). Whole ponds, lakes or rivers may be filled with the separate and independent parts produced by a single zygote.

Trees are the supreme example of plants whose growth is concentrated vertically. The peculiar feature distinguishing trees and shrubs from most herbs is the connecting system linking modules together and connecting them to the root system. This does not rot away, but thickens with wood, conferring perenniality. Most of the structure of such a woody tree is dead, with a thin layer of living material lying immediately below the bark. The living layer, however, continually regenerates new tissue, and adds further layers of dead material to the trunk of the tree. This solves, by the strength it provides, the difficult problem of obtaining water and nutrients below the ground, but also light, perhaps 50 m away at the top of the canopy.

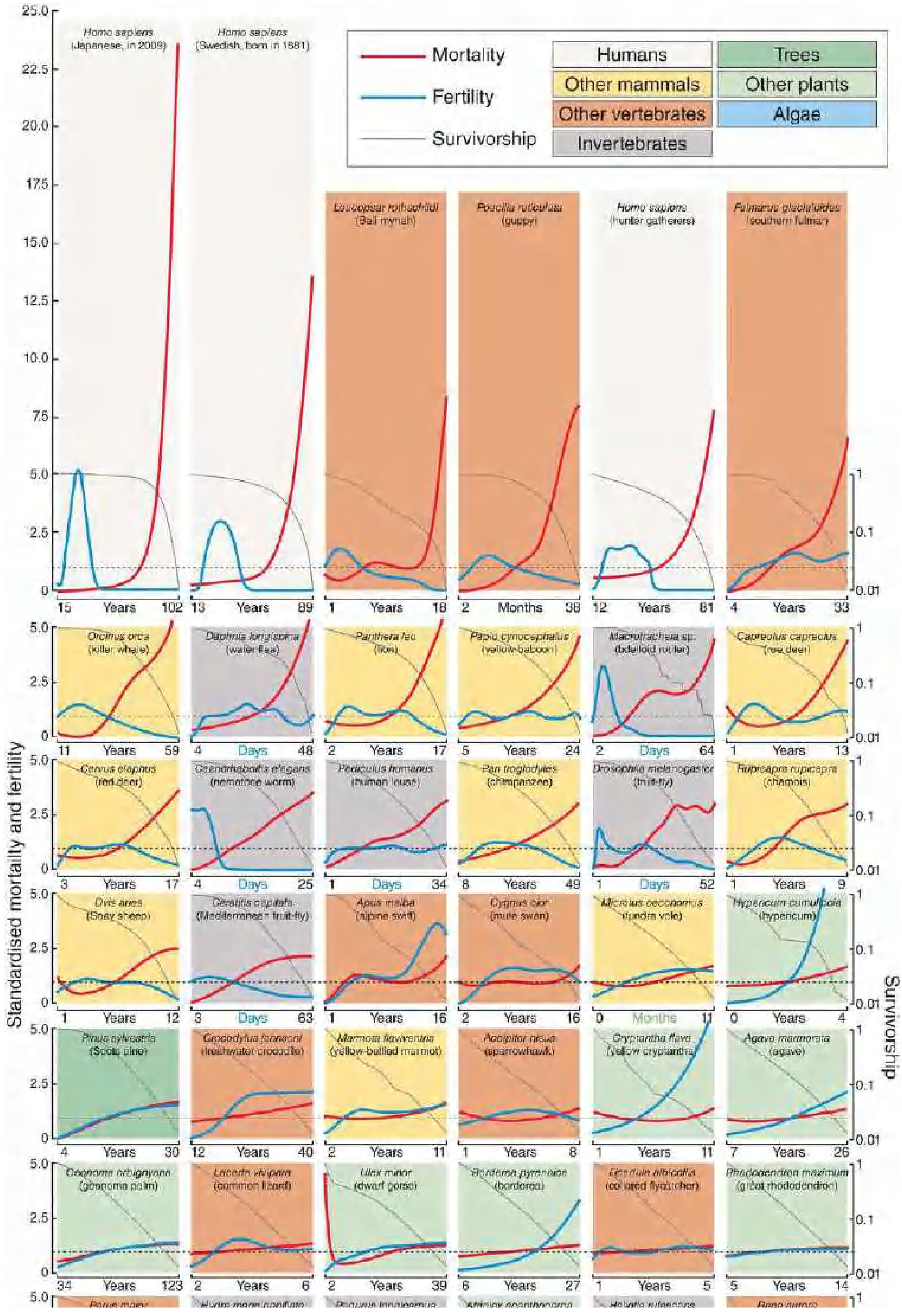
modules within modules

We can often recognise two or more levels of modular construction. The strawberry (Figure 4.1c) is a good example of this: leaves are repeatedly developed from a bud, but these leaves are arranged into rosettes. The strawberry plant grows (i) by adding new leaves to a rosette and (ii) by producing new rosettes on stolons grown from the axils of its rosette leaves. Trees also exhibit modularity at several levels: the leaf with its axillary bud, the whole shoot on which the leaves are arranged, and the whole branch systems that repeat a characteristic pattern of shoots.

Many animals, despite variations in their precise method of growth and reproduction, are as ‘modular’ as any plant. And in corals, for example, just like many plants, the individual may exist as a physiologically integrated whole, or may be split into a number of colonies – all part of one individual, but physiologically independent (Hughes *et al*., 1992).

4.2.3 Senescence – or the lack of it – in modular organisms

There is also often no programmed senescence of whole modular organisms – they appear to have perpetual somatic youth (see Thomas (2013) for a review of senescence, and its avoidance, in plants). Even in trees that accumulate their dead stem tissues, or gorgonian corals that accumulate old calcified branches, death often results from becoming too big or succumbing to disease rather than from programmed senescence. We see evidence of this in Figure 4.2, which shows how rates of mortality and birth vary with age in a wide variety of organisms. It is a figure to which we will return several times in this chapter. For now, we can note that there are a number examples there, indeed of trees and a gorgonian coral, in which there is no evidence of the increases in mortality at older ages that we see in organisms that senesce, like ourselves.


Mortality
Fertility
Survivorship
Humans
Other mammals
Other vertebrates
Invertebrates
Trees
Other plants
Algae
Standardised mortality and fertility
Survivorship
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
1
0.1
0.01
Homo sapiens (Japanese, in 2009)
15 Years 102
Homo sapiens (Swedish, born in 1881)
13 Years 89
Leucopsar rothschildi (Bali mynah)
1 Years 18
Poecilia reticulata (guppy)
2 Months 38
Homo sapiens (hunter gatherers)
12 Years 81
Fulmarus glacialoides (southern fulmar)
4 Years 33
Orcinus orca (killer whale)
11 Years 59
Daphnia longispina (water flea)
4 Days 48
Panthera leo (lion)
2 Years 17
Papio cynocephalus (yellow baboon)
5 Years 24
Macrotrachela sp. (bdelloid rotifer)
2 Days 64
Capreolus capreolus (roe deer)
1 Years 13
Cervus elaphus (red deer)
3 Years 17
Caenorhabditis elegans (nematode worm)
4 Days 25
Pediculus humanus (human louse)
1 Days 34
Pan troglodytes (chimpanzee)
8 Years 49
Drosophila melanogaster (fruit-fly)
1 Days 52
Rupicapra rupicapra (chamois)
1 Years 9
Ovis aries (Soay sheep)
1 Years 12
Ceratitis capitata (Mediterranean fruit-fly)
3 Days 63
Apus melba (alpine swift)
1 Years 16
Cygnus olor (mute swan)
2 Years 16
Microtus oeconomus (tundra vole)
0 Months 11
Hypericum cumulicola (hypericum)
0 Years 4
Pinus sylvestris (Scots pine)
4 Years 30
Crocodylus johnsoni (freshwater crocodile)
12 Years 40
Marmota flaviventris (yellow-bellied marmot)
2 Years 11
Accipiter nisus (sparrowhawk)
1 Years 8
Cryptantha flava (yellow cryptantha)
1 Years 11
Agave marmorata (agave)
7 Years 26
Geonoma orbignyana (geonoma palm)
34 Years 123
Lacerta vivipara (common lizard)
2 Years 6
Ulex minor (dwarf gorse)
2 Years 39
Borderea pyrenaica (borderea)
6 Years 27
Ficedula albicollis (collared flycatcher)
1 Years 5
Rhododendron maximum (great rhododendron)
5 Years 14
Parus major
Hydra magnipapillata
Pagurus longicarpus
Atriplex acanthocarpa
Haliotis rufescens
Rana aurora

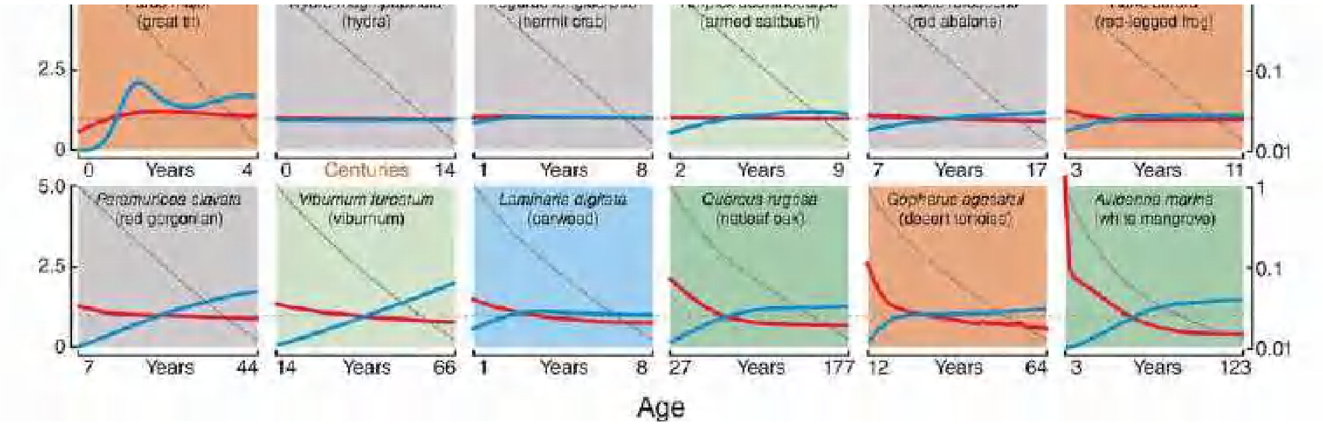


Figure 4.2 Compilation of patterns of mortality (survivorship) and reproduction from across the plant and animal kingdoms from reproductive maturity to the age where only 5% of the adult population is still alive. To emphasise variations in pattern, mortality and fertility are scaled relative to their means. Survivorship is plotted on a log scale. The plots are arranged in order of decreasing mortality at the terminal age. Note the marked contrast between organisms like ourselves (top line) that show senescence, where there is a marked increase in mortality in old age, and those like the coral and oak tree in the bottom line where there is no such increase. This is part of a more general variation in the shape of survivorship curves, picked up again in Figure 4.11.

Source: After Jones *et al.* (2014).

At the modular level, things are quite different. The annual death of the leaves on a deciduous tree is the most dramatic example of senescence – but roots, buds, flowers and the modules of modular animals all pass through phases of youth, middle age, senescence and death. The growth of the individual genet is the combined result of these processes. Figure 4.3, for example, shows that the age structure of leaves of the perennial herb, *Wedelia trilobata*, a native of central America, is changed dramatically by the application of nitrogen fertiliser. Plants are larger when they are more heavily fertilised, and the rate at which they 'give birth' to leaves is greater, but so too is the death rate of those leaves.

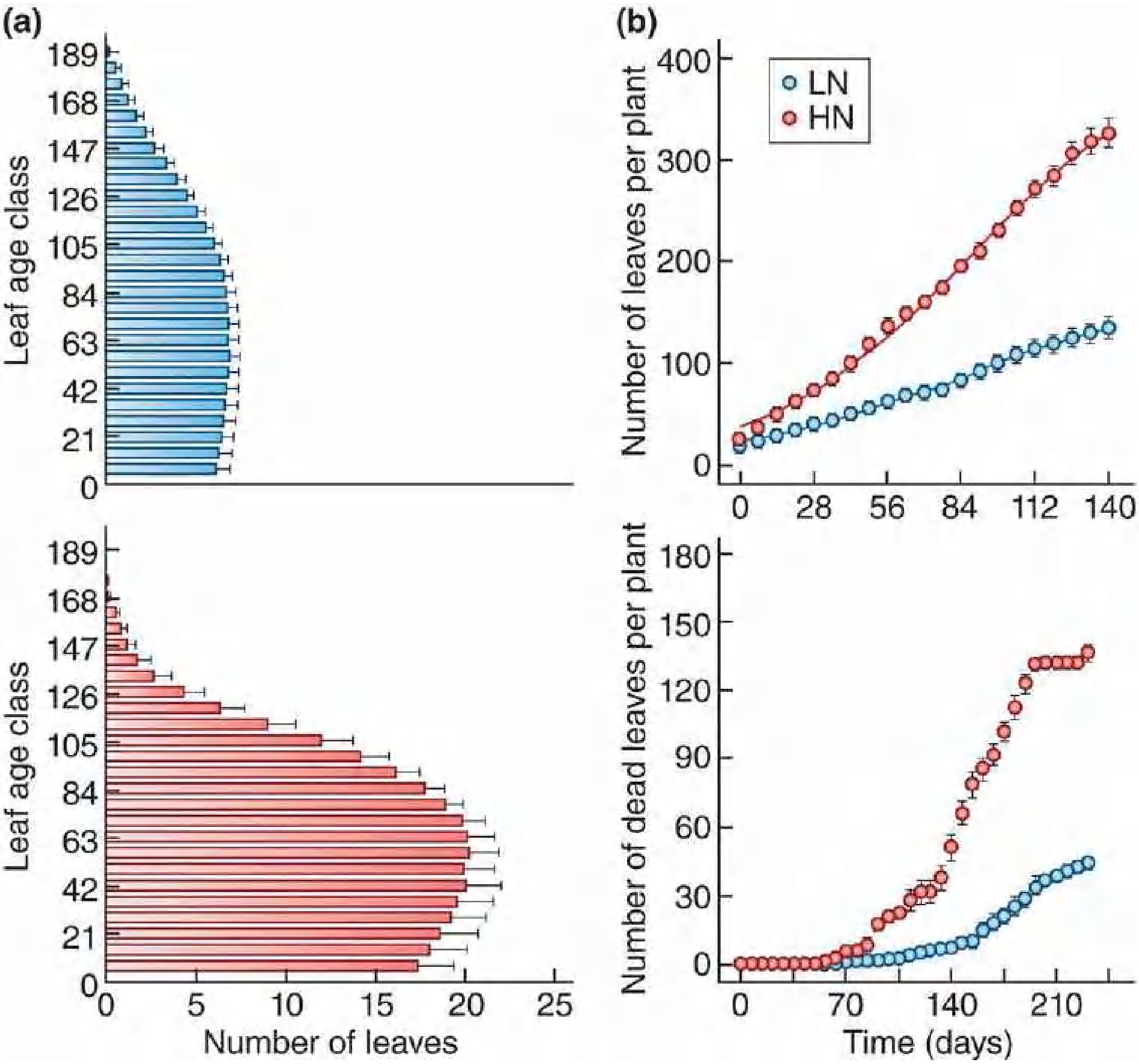


Figure 4.3 **The growth of a genet reflects the births and deaths of its component modules.** (a) Numbers of leaves of plants of *Wedelia trilobata* (means of six plants), divided into seven-day age classes, cultivated at low (above) and high (below) nitrogen availability. Bars are SEs. At high nitrogen availability, the plants are not only larger: they also have a much higher proportion of young leaves. (b) The cumulative number of newly produced (above) and dead (below) leaves in the same study. The high proportion of young leaves at high nitrogen (HN) availability seen in (a) is the result of both birth and death rates of leaves being higher. LN, low nitrogen.

Source: After Suarez (2016).

4.2.4 Integration

For many rhizomatous and stoloniferous species, this changing age structure is in turn associated with a changing level to which the connections between individual ramets remain intact. A young ramet may benefit from the nutrients flowing from an older ramet to which it is attached and from which it grew. But the pros and cons of attachment will have changed markedly by the time the daughter is fully established in its own right and the parent has entered a postreproductive phase of senescence – a comment equally applicable to unitary organisms with parental care, like ourselves (Caraco & Kelly, 1991).

The changing benefits and costs of integration have been studied experimentally in the pasture grass *Holcus lanatus*, by comparing the growth of: (i) ramets that were left with a physiological connection to their parent plant, and in the same pot, so that parent and daughter might compete (competing, connected: CC); (ii) ramets that were left in the same pot so competition was still possible but had their connection severed (competing, not connected: CN); and (iii) ramets that had their connection severed and were repotted in their parent's soil after the parent had been removed, so no competition was possible (independent plants, neither competing nor connected: NN) (Figure 4.4). These treatments were applied to daughter ramets of various ages, which were then examined after a further eight weeks' growth. For the youngest daughters, just one week old (Figure 4.4a), connection to the parent significantly enhanced growth (CC > CN), but competition with the parent had no apparent effect (CN ≈ NN). For slightly older daughters, two weeks old (Figure 4.4b), competition with the parent did depress growth (NN > CN), but physiological connection with the parent effectively negated this (CC > CN; CC ≈ NN). For even older daughters, however, the balance shifted further still. Competition with the parent again depressed growth of the daughter (NN > CN), but this time physiological connection to the parent was either not enough to fully overcome this (at four weeks, Figure 4.4c; NN > CC > CN) or eventually appeared to represent a further drain on the daughter's resources (after eight weeks, Figure 4.4d; NN > CN > CC).

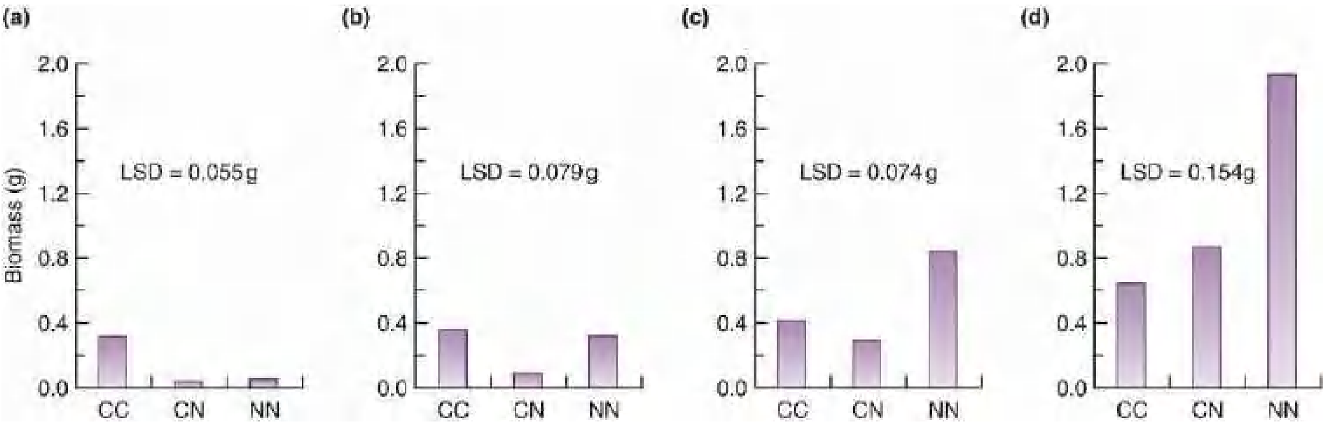


Figure 4.4 Integration within a plant leads to a shifting balance of positive and negative effects between parent and daughter modules as modules age. The growth of daughter ramets of the grass *Holcus lanatus*, which were initially (a) one week, (b) two weeks, (c) four weeks and (d) eight weeks old, and were then grown on for a further eight weeks. LSD, least significant difference, needs to be exceeded for two means to be significantly different from each other. For further discussion, see text. CC, competing, connected; CN, competing, not connected; NN, independent plants, neither competing nor connected.

Source: After Bullock *et al*. (1994).

4.3 Counting individuals

If we are going to study birth, death and modular growth seriously, we must quantify them. This means counting individuals and (where appropriate) modules. Indeed, many studies concern themselves not with birth and death but with their consequences – the total number of individuals present and the way these numbers vary with time. Whether organisms are unitary or modular, ecologists face enormous technical problems when they try to count what is happening to populations in nature. A great many ecological questions remain unanswered because of these problems.

what is a population?

It is usual to use the term *population* to describe a group of individuals of one species under investigation. What actually constitutes a population, though, will vary from species to species and from study to study. In some cases, the boundaries of a population are readily apparent: the sticklebacks occupying a small lake are 'the stickleback population of the lake'. In other cases, boundaries are determined more by an investigator's purpose or convenience: it is possible to study the population of lime aphids inhabiting one leaf, one tree, one stand of trees or a whole woodland. In yet other cases – and there are many of these – individuals are distributed continuously over a wide area, and an investigator must define the limits of a population arbitrarily. In such cases, especially, it is often more convenient to consider the *density* of a population. This is usually defined as 'numbers per unit area', but in certain circumstances 'numbers per leaf', 'numbers per host' or some other measure may be appropriate.

estimating population size

To determine the size of a population, one might imagine that it is possible simply to count individuals, especially for relatively small, isolated habitats like islands and relatively large individuals like deer. For most species, however, such 'complete enumerations' are impractical or impossible: observability – our ability to observe every individual present – is almost always less than 100%. Ecologists, therefore, must almost always *estimate* the number of individuals in a population rather than count them. They may estimate the numbers of aphids on a crop, for example, by counting the number on a representative sample of leaves, then estimating the number of leaves per square metre of ground, and from this estimating the number of aphids per square metre. For plants and animals living on the ground surface, the sample unit is generally a small area known as a quadrat (which is also the name given to the square or rectangular device used to demarcate the boundaries of the area on the ground). For soil-dwelling organisms the unit is usually a volume of soil; for lake dwellers a volume of water; for many herbivorous insects the unit is one typical plant or leaf, and so on. Further details of sampling methods, and of methods for counting individuals generally, can be found in one of many texts devoted to ecological methodology (e.g. Krebs, 1999 ; Henderson & Southwood, 2016).

For animals, especially, there are two further methods of estimating population size. The first is known as *capture–recapture*. At its simplest, this involves catching a random sample of a population, marking individuals so that they can be recognised subsequently, releasing them so that they remix with the rest of the population, and then catching a further random sample. Population size can be estimated from the proportion of this second sample that bear a mark, since, as long as the population remains the same between the two samples, that proportion should be the same as in the whole population, and the number marked in the population is known, since we, the investigators, marked and released them. In practice, there is usually a whole sequence of capture–recapture samples (not just two), and the assumption that populations remain the same between samples is relaxed. Methods of analysis are consequently more complex and more powerful (Cooch & White, 2019).

The final method is to use an *index of abundance*. This can provide information on the relative size of a population, but by itself usually gives little indication of absolute size. As an example, Figure 4.5 shows how the number of occupied ponds, and the amount of summer (terrestrial) habitat in the vicinity of a pond, affected the abundance of leopard frogs (*Rana pipiens*) near Ottawa, Canada. Here, frog abundance was estimated from the 'calling rank', based on whether there were no frogs, 'few', 'many' or 'very many' frogs calling on each of four occasions. Despite their shortcomings, even indices of abundance can provide valuable information.

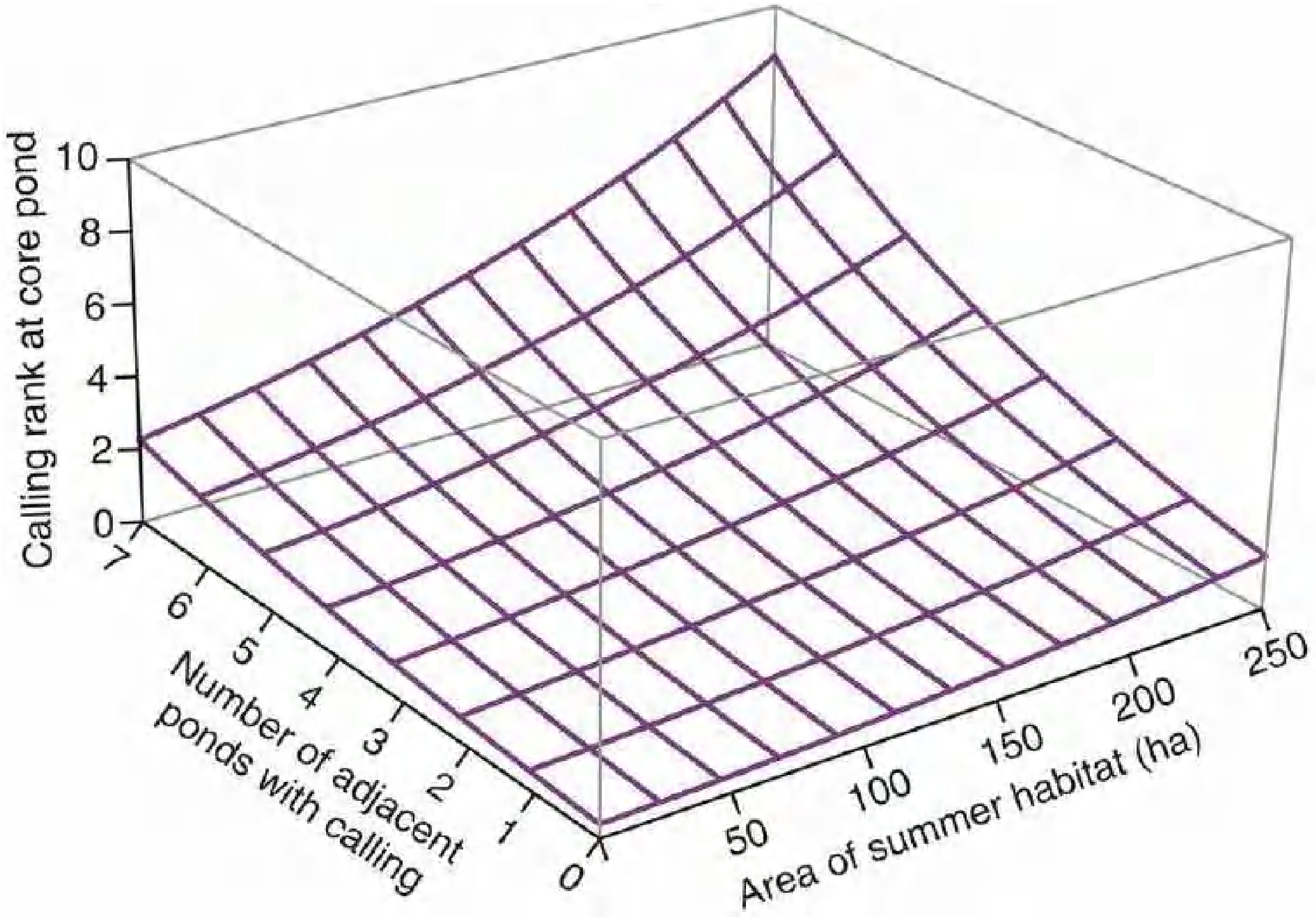


Figure 4.5 **Indices of abundance can provide valuable information.** The abundance (calling rank) of leopard frogs in ponds increases significantly with both the number of adjacent ponds that are occupied and the area of summer habitat within 1 km of the pond. Calling rank is the sum of an index measured on four occasions, namely: 0, no individuals calling; 1, individuals can be counted, calls not overlapping; 2, calls of <15 individuals can be distinguished with some overlapping; 3, calls of ≥15 individuals.

Source: After Pope *et al.* (2000).

counting births

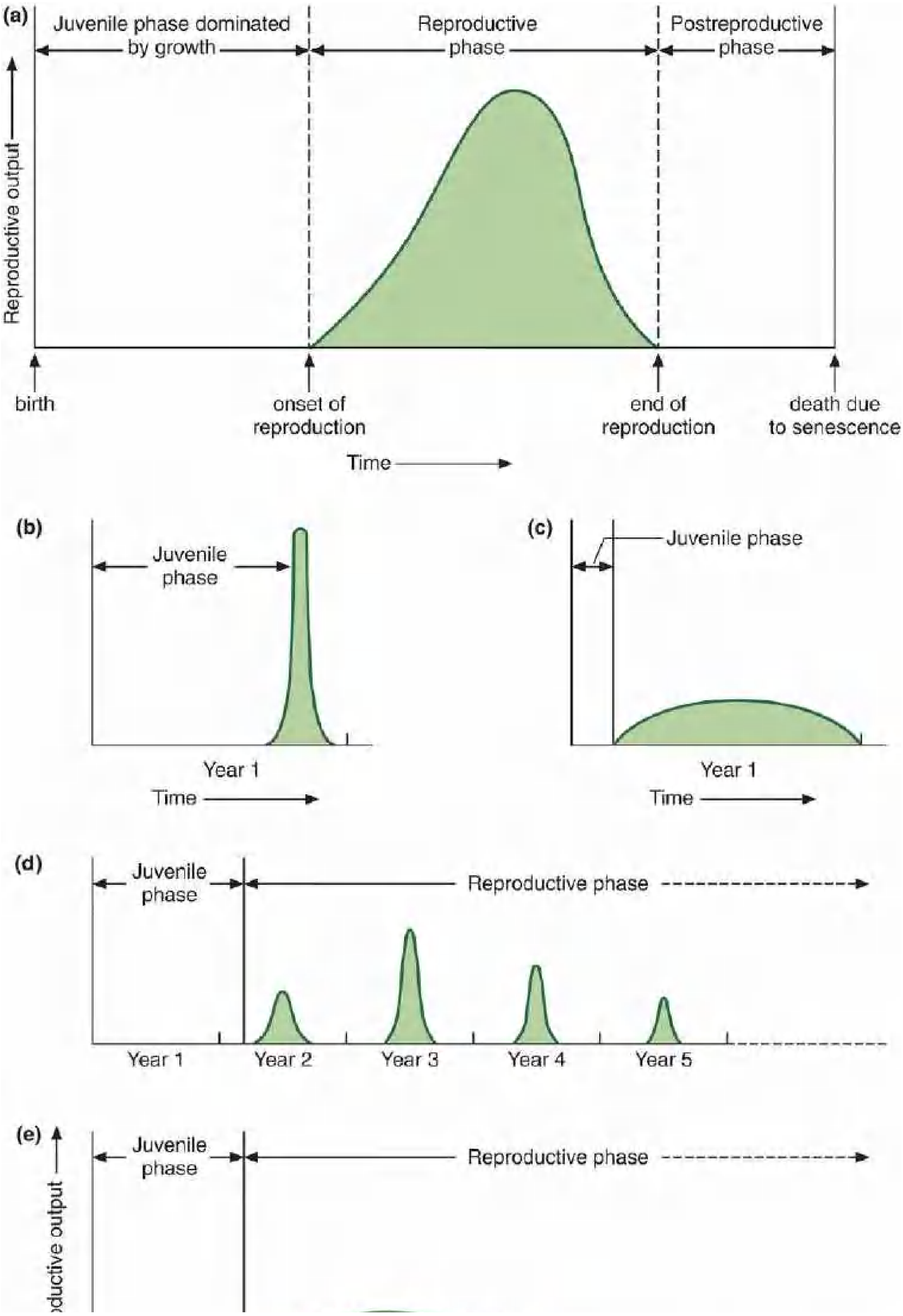
Counting births can be more difficult even than counting individuals. The formation of the zygote is often regarded as the starting point in the life of an individual. But it is a stage that is often hidden and extremely hard to study. We simply do not know, for most animals and plants, how many embryos die before 'birth', though in the rabbit at least 50% of embryos are thought to die in the womb, and in many higher plants it seems that about 50% of embryos abort before the seed is fully grown and mature. Hence, it is almost always impossible in practice to treat the start of life as the time of birth. In birds we may use the moment that an egg hatches; in mammals, perhaps, when an individual starts to be supported outside the mother as a suckling; and in plants we may use the germination of a seed as the birth of a seedling, although it is really only the moment at which a developed embryo restarts into growth after a period of dormancy. We need to remember that often half or more of a population will have died before they can be recorded as born!

counting deaths

Counting deaths poses as many problems. Dead bodies do not linger long in nature. Only the skeletons of large animals persist long after death. Seedlings may be counted and mapped one day and gone without trace the next. Mice, voles and soft-bodied animals such as caterpillars and worms are digested by predators or rapidly removed by scavengers or decomposers. They leave no carcasses to be counted and no evidence of the cause of death. Capture–recapture methods can go a long way towards estimating deaths from the loss of marked individuals from a population (they are probably used as often to measure survival as abundance), but even here it is often impossible to distinguish loss through death and loss through emigration.

4.4 Life cycles

We have noted already that counting the numbers in a population provides only an outline sketch, and one key reason for this is that virtually all organisms go through a number of stages in their lives, each with their own birth and death rates, responses to other organisms, resources and conditions, and so on. Hence, we need to understand the sequences of events that occur in those organisms' life cycles. A highly simplified, generalised life history is outlined in Figure 4.6a. It comprises birth, followed by a prereproductive period, a period of reproduction, perhaps a postreproductive period, and then death as a result of senescence (though of course other forms of mortality may intervene at any time). The variety of life cycles is also summarised diagrammatically in Figure 4.6, although there are many life cycles that defy this simple classification. Some organisms fit several or many generations within a single year, some have just one generation each year (annuals), and others have a life cycle extended over several or many years. For all organisms, though, a period of growth occurs before there is any reproduction, and growth usually slows down (and in some cases stops altogether) when reproduction starts.


(a)
Juvenile phase dominated by growth
Reproductive phase
Postreproductive phase
Reproductive output
birth
onset of reproduction
end of reproduction
death due to senescence
Time
(b)
Juvenile phase
Year 1
Time
(c)
Juvenile phase
Year 1
Time
(d)
Juvenile phase
Reproductive phase
Year 1
Year 2
Year 3
Year 4
Year 5
(e)
Juvenile phase
Reproductive phase
oductive output

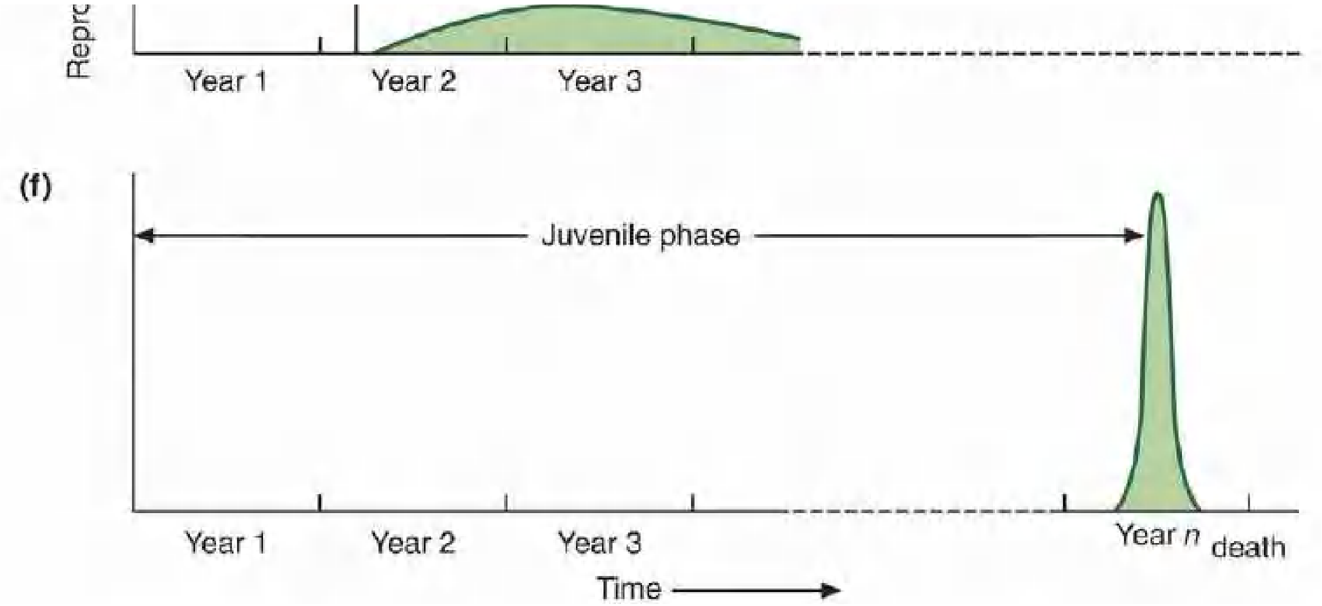


Figure 4.6 Schematic life histories for unitary organisms. (a) An outline life history for a unitary organism. Time passes along the horizontal axis, which is divided into different phases. Reproductive output is plotted on the vertical axis. The figures below (b–f) are variations on this basic theme (technical terms explained in the text). (b) A semelparous annual species. (c) An iteroparous annual species. (d) A long-lived iteroparous species with seasonal breeding (that may indeed live much longer than suggested in the figure). (e) A long-lived species with continuous breeding (that may again live much longer than suggested in the figure). (f) A semelparous species living longer than a year, where the prereproductive phase may be a little over one year (a biennial species, breeding in its second year) or longer, often much longer than this (as shown).

semelparous and iteroparous life cycles

Whatever the length of their life cycle, species may, broadly, be either *semelparous* or *iteroparous* (often referred to by plant scientists as monocarpic and polycarpic). In semelparous species, individuals have only a single, distinct period of reproductive output in their lives. Prior to this they have largely ceased to grow, during it they invest little or nothing in survival that might take them to future reproductive events, and after it they die. By contrast, in iteroparous species, an individual normally experiences several or many such reproductive events, which may in fact merge into a single extended period of reproductive activity. During each period of reproductive activity, however, the individual continues to invest in future survival and possibly growth, and it therefore has a reasonable chance of surviving beyond each bout of reproduction to reproduce again.

For example, many annual plants are semelparous (Figure 4.6b): they have a sudden burst of flowering and seed set, and then they die. This is commonly the case among the weeds of arable crops. Other annuals, such as groundsel (*Senecio vulgaris*), are iteroparous (Figure 4.6c): they continue to grow and produce new flowers and seeds through the season until they are killed by the first lethal frost of winter. They die with their buds on.

the variety of life cycles

There is also a marked seasonal rhythm in the lives of many long-lived iteroparous plants and animals, especially in their reproductive activity, with a period of reproduction once per year (Figure 4.6d). Mating (or the flowering of plants) is commonly triggered by the length of the photoperiod (see Section 2.3.7), synchronising birth, egg hatch or seed ripening with the time that seasonal resources are likely to be abundant. Here, though, unlike annual species, the generations overlap and individuals of a range of ages breed side by side. The population is maintained in part by survival of adults and in part by new births.

In wet equatorial regions, on the other hand, where there is very little seasonal variation in temperature and rainfall and scarcely any variation in photoperiod, we find species of plants that are in flower and fruit throughout the year – and continuously breeding species of animal that subsist on this resource (Figure 4.6e). There are several species of fig (*Ficus*), for instance, that bear fruit continuously and form a reliable year-round food supply for birds and primates. In more seasonal climates, humans are unusual in also breeding continuously throughout the year, though numbers of other species, cockroaches, for example, do so in the stable environments that humans have created.

Amongst long-lived (i.e. longer than annual) semelparous plants (Figure 4.6f), some are strictly biennial. Each individual takes two summers and the intervening winter to develop, but has only a single reproductive phase, in its second summer. An example is the white sweet clover, *Melilotus alba* (Klemow & Raynal, 1981). In New York State, USA this has relatively high mortality during the first growing season (whilst seedlings are developing into established plants), followed by much lower mortality until the end of the second summer, when the plants flower and survivorship decreases rapidly. No plants survive to a third summer. Thus, there is an overlap of two generations at most. A more typical example of a semelparous species with overlapping generations is the composite *Grindelia lanceolata*, which may flower in its third, fourth or fifth years. But whenever an individual does flower, it dies soon after.

A well-known example of a semelparous animal with overlapping generations (Figure 4.6f) is the Pacific salmon *Oncorhynchus nerka*. Salmon are spawned in rivers. They spend the first phase of their juvenile life in fresh water and then migrate to the sea, often travelling thousands of miles. At maturity they return to the stream in which they were hatched. Some mature and return to reproduce after only two years at sea; others mature more slowly and return after three, four or five years. At the time of reproduction the population of salmon is composed of overlapping generations of individuals. But all are semelparous: they lay their eggs and then die; their bout of reproduction is terminal.

There are even more dramatic examples of species that have a long life but reproduce just once. Many species of bamboo form dense clones of shoots that remain vegetative for many years: up to 100 years in some species. The whole population of shoots, from the same and sometimes different clones, then flowers simultaneously in a mass suicidal orgy. Even when shoots have become physically separated from each other, the parts still flower synchronously.

seed banks, ephemerals and other not-quite-annuals

For many annual plants, that description is itself, in an important sense, misleading, because their seeds accumulate in the soil in a buried *seed bank*. At any one time, therefore, seeds of a variety of ages are likely to occur together in the seed bank, and when they germinate the seedlings will also be of varying ages (age being the length of time since the seed was first produced). This is just one example of real organisms spoiling our attempts to fit them neatly into clear-cut categories. The formation of something comparable to a seed bank is rarer amongst animals, but there are examples to be seen amongst the eggs of nematodes, mosquitoes and fairy shrimps, the gemmules of sponges and the statocysts of bryozoans.

the species composition of seed banks

As a general rule, dormant seeds, which enter and make a significant contribution to seed banks, are more common in annuals and other short-lived plant species than they are in longer lived species, such that short-lived species tend to predominate in buried seed banks, even when most of the established plants above them belong to much longer lived species. Certainly, the species composition of seed banks and the mature vegetation above may be very different (Figure 4.7; see also Application 4.1).

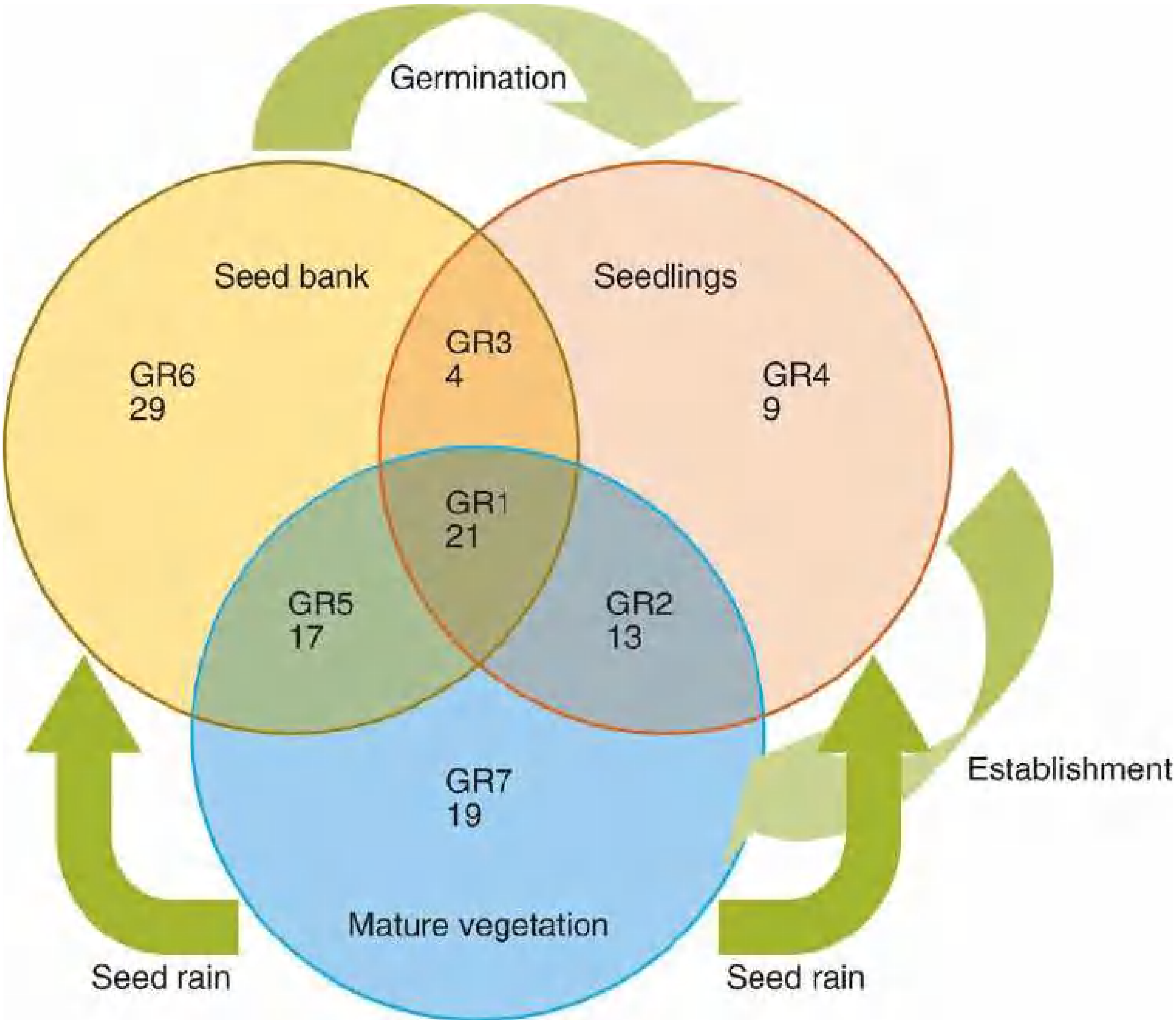


Figure 4.7 The composition of seed banks can be very different from the vegetation above them. Species recovered from the seed bank, from seedlings and from mature vegetation in a coastal grassland site on the western coast of Finland. Seven species groups (GR1–GR7) are defined on the basis of whether they were found in only one, two or all three stages (numbers of species shown below the group numbers). GR3 (seed bank and seedlings only) is an unreliable group of species that are mostly incompletely identified; in GR5 there are many species difficult to identify as seedlings that may more properly belong to GR1. Nonetheless, the marked difference in composition, especially between the seed bank and the mature vegetation, is readily apparent.

Source: After Jutila (2003).

Annual species with seed banks are not the only ones for which the term annual is, strictly speaking, inappropriate. For example, there are many annual plant species living in deserts that are far from seasonal in their appearance. They have a substantial buried seed bank, with germination occurring on rare occasions after substantial rainfall. Subsequent development is usually rapid, so that the period from germination to seed production is short. Such plants are best described as semelparous *ephemerals*.

A simple annual label also fails to fit species where the majority of individuals in each generation are annual, but where a small number postpone reproduction until their second summer. This applies, for example, to the terrestrial isopod *Philoscia muscorum* living in north-east England (Sunderland *et al.*, 1976). Approximately 90% of females bred only in the first summer after they were born; the other 10% bred only in their second summer. In some other species, the difference in numbers between those that reproduce in their first or second years is so slight that the description *annual–biennial* is most appropriate. In short, it is clear that annual life cycles merge into more complex ones without any sharp discontinuity.

APPLICATION 4.1 Seed banks and the restoration of forested wetlands

Changes to farming practices mean that there are an increasing number of cases of agricultural land being abandoned. Whenever this happens, there is an understandable hope that the natural habitat that had been destroyed to make way for farming can be restored. One example is the swampland dominated by bald cypress trees, *Taxodium distichum*, lying along the rivers and streams of the Gulf coastal floodplain of North America, running from south-eastern Texas north and east to the Atlantic Ocean, and including a study site along the Cache River in southern Illinois (Middleton, 2003). Agricultural development expanded there in the 1950s, and by the late 1980s only about half of the forested swampland remained. However, the process has been halted and reversed, in part by a crash in the soybean market, and the emphasis now is on plans to restore the original swamplands, with the provision of habitat for hunting being a particular commercial driver, though there are Nature Preserve areas where hunting is not allowed.

In fact, restoration has proved difficult. The original species-rich forests supported up to 60 or so species of trees, shrubs and vines, many of them with seeds dispersed in the seasonal floods, but the forests developing following agricultural abandonment tend to be dominated by a few species with wind-dispersed seeds. An important question, therefore, is what potential seed banks have in promoting more natural restoration. To address this, Middleton (2003) assayed the seeds from nine sites in intact bald cypress swamps and 51 sites in the area that had been farmed for between one and 50 years. She found not only that there was no relationship for the dominant swamp species between the length of time farming had been practiced and seed abundance – but actually, the seeds of many of those species, including bald cypress itself, were absent from the seed banks altogether in both the farmed and intact sites. Instead these were composed mostly of seeds from large numbers of herbaceous species. It seems, therefore, that in this case seed banks can have little part to play in habitat restoration, and that abandonment alone, even of land that has been farmed for a relatively short time, offers little prospect of a return to natural habitat. Rather, the short-lived seeds of the dominant woody species of bald cypress swamps are likely to return and promote successful restoration only if flood pulsing across the landscape is re-engineered, reconnecting pristine to abandoned sites.

4.5 Dormancy

migration in time

We will discuss in Chapter 6 how an organism gains in fitness by dispersing its progeny ‘elsewhere’, as long as the progeny are more likely to leave descendants than if they remained undispersed. Similarly, an organism gains in fitness by delaying its arrival on the scene, so long as the delay increases its chances of leaving descendants. This will often be the case when conditions in the future are likely to be better than those in the present. Thus, a delay in the recruitment of an individual to a population may be regarded as ‘migration in time’.

Organisms generally spend their period of delay in a state of dormancy. This relatively inactive state has the benefit of conserving energy, which can then be used during the period following the delay. In addition, the dormant phase of an organism is often more tolerant of the adverse environmental conditions prevailing during the delay (i.e. tolerant of drought, extremes of temperature, lack of light and so on). Dormancy can be either predictive or consequential (Müller, 1970). Predictive dormancy is initiated in advance of the adverse conditions, and is most often found in predictable, seasonal environments. It is generally referred to as ‘diapause’ in animals, and in plants as ‘innate’ or ‘primary’ dormancy (Harper, 1977). Consequential (or ‘secondary’) dormancy, on the other hand, is initiated in response to the adverse conditions themselves.

4.5.1 Dormancy in animals: diapause

Diapause has been most intensively studied in insects, where examples occur in all developmental stages. The common field grasshopper *Chorthippus brunneus* is a fairly typical example. This annual species passes through an *obligatory* diapause in its egg stage, where, in a state of arrested development, it is resistant to the cold winter conditions that would quickly kill the nymphs and adults. In fact, the eggs require a long cold period before development can start again (around five weeks at 0°C, or rather longer at a slightly higher temperature). This ensures that the eggs are not affected by a short, freak period of warm winter weather that might then be followed by normal, dangerous, cold conditions. It also means that there is an enhanced synchronisation of subsequent development in the population as a whole. The grasshoppers ‘migrate in time’ from late summer to the following spring.

the importance of photoperiod

Diapause is also common in species with more than one generation per year. For instance, the fruit-fly *Drosophila obscura* passes through four generations per year in England, but enters diapause during only one of them. This *facultative* diapause shares important features with obligatory diapause: it enhances survivorship during a predictably adverse winter period, and it is experienced by *resistant* diapause adults with arrested gonadal development and large reserves of stored abdominal fat. In this case, synchronisation is achieved not only during diapause but also prior to it. Emerging adults react to the short daylengths of autumn by laying down fat and entering the diapause state; they recommence development in response to the longer days of spring. Thus, by relying, like many species, on the utterly predictable *photoperiod* as a cue for seasonal development, *D. obscura* enters a state of predictive diapause that is confined to those generations that inevitably pass through the adverse conditions.

Consequential dormancy may be expected to evolve in environments that are relatively unpredictable. In such circumstances, there will be a disadvantage in responding to adverse conditions only after they have appeared, but this may be outweighed by the advantages of: (i) responding to favourable conditions *immediately* after they reappear; and (ii) entering a dormant state only if adverse conditions *do* appear. Thus, when many mammals enter hibernation, they do so (after an obligatory preparatory phase) in direct response to the adverse conditions. Having achieved ‘resistance’ by virtue of the energy they conserve at a lowered body temperature, and having periodically emerged and monitored their environment, they eventually cease hibernation whenever the adversity disappears.

4.5.2 Dormancy in plants

Seed dormancy is an extremely widespread phenomenon in flowering plants. The young embryo ceases development whilst still attached to the mother plant and enters a phase of suspended activity, usually losing much of its water and becoming dormant in a desiccated condition. In a few species of higher plants, such as some mangroves, a dormant period is absent, but this is very much the exception – almost all seeds are dormant when they are shed from the parent and require special stimuli to return them to an active state (germination).

Dormancy in plants, though, is not confined to seeds. Many species accumulate dormant bud banks analogous to the seed banks produced by other species. In one study of tallgrass prairies in north-eastern Kansas, USA, for example, it was estimated that more than 99% of new tiller production arose from below-ground vegetative buds rather than from seed (Benson & Hartnett, 2006); and in another prairie study, in Montana, USA, the differential responses of grass species to fire at different seasons of the year, especially their release by fire from dormancy, were crucial in driving the overall dynamics and community structure (Russell *et al.*, 2015).

Indeed, the very widespread habit of deciduousness is a form of dormancy displayed by many perennial trees and shrubs. Established individuals pass through periods, usually of low temperatures and low light levels, in a leafless state of low metabolic activity.

innate, enforced and induced dormancy

Three types of dormancy have been distinguished:

1. *Innate dormancy* is a state in which there is an absolute requirement for some special external stimulus to reactivate the process of growth and development. The stimulus may be the presence of water, low temperature, light, photoperiod, fire (see previously) or an appropriate balance of near- and far-red radiation. Seedlings of such species tend to appear in sudden flushes of almost simultaneous germination. Deciduousness is also an example of innate dormancy.
2. *Enforced dormancy* is a state imposed by external conditions (i.e. it is consequential dormancy). For example, the Missouri goldenrod *Solidago missouriensis* enters a dormant state when attacked by the beetle *Trirhabda canadensis*. Eight clones, identified by genetic markers, were followed prior to, during and after a period of severe defoliation. The clones, which varied in extent from 60 to 350 m^2 and from 700 to 20 000 rhizomes, failed to produce any above-ground growth (i.e. they were dormant) in the season following defoliation and had apparently died, but they reappeared 1–10 years after they had disappeared, and six of the eight bounced back strongly within a single season (Figure 4.8). Generally, the progeny of a single plant with enforced dormancy may be dispersed in time over years, decades or even centuries. Seeds of *Chenopodium album* collected from archaeological excavations have been shown to be viable when 1700 years old (Ødum, 1965).
3. *Induced dormancy* is a state produced in a seed during a period of enforced dormancy in which it acquires some new requirement before it can germinate. The seeds of many agricultural and horticultural weeds will germinate without a light stimulus when they are released from the parent; but after a period of enforced dormancy they require exposure to light before they will germinate. For a long time it was a puzzle that soil samples taken from the field to the laboratory would quickly generate huge crops of seedlings, although these same seeds had failed to germinate in the field. It was a simple idea of genius that prompted Wesson and Wareing (1969) to collect soil samples from the field at night and bring them to the laboratory in darkness. They obtained large crops of seedlings from the soil only when the samples were exposed to light. This type of induced dormancy is responsible for the accumulation of large populations of seeds in the soil. In nature they germinate only when they are brought to the soil surface by earthworms or other burrowing animals, or by the exposure of soil after a tree falls.

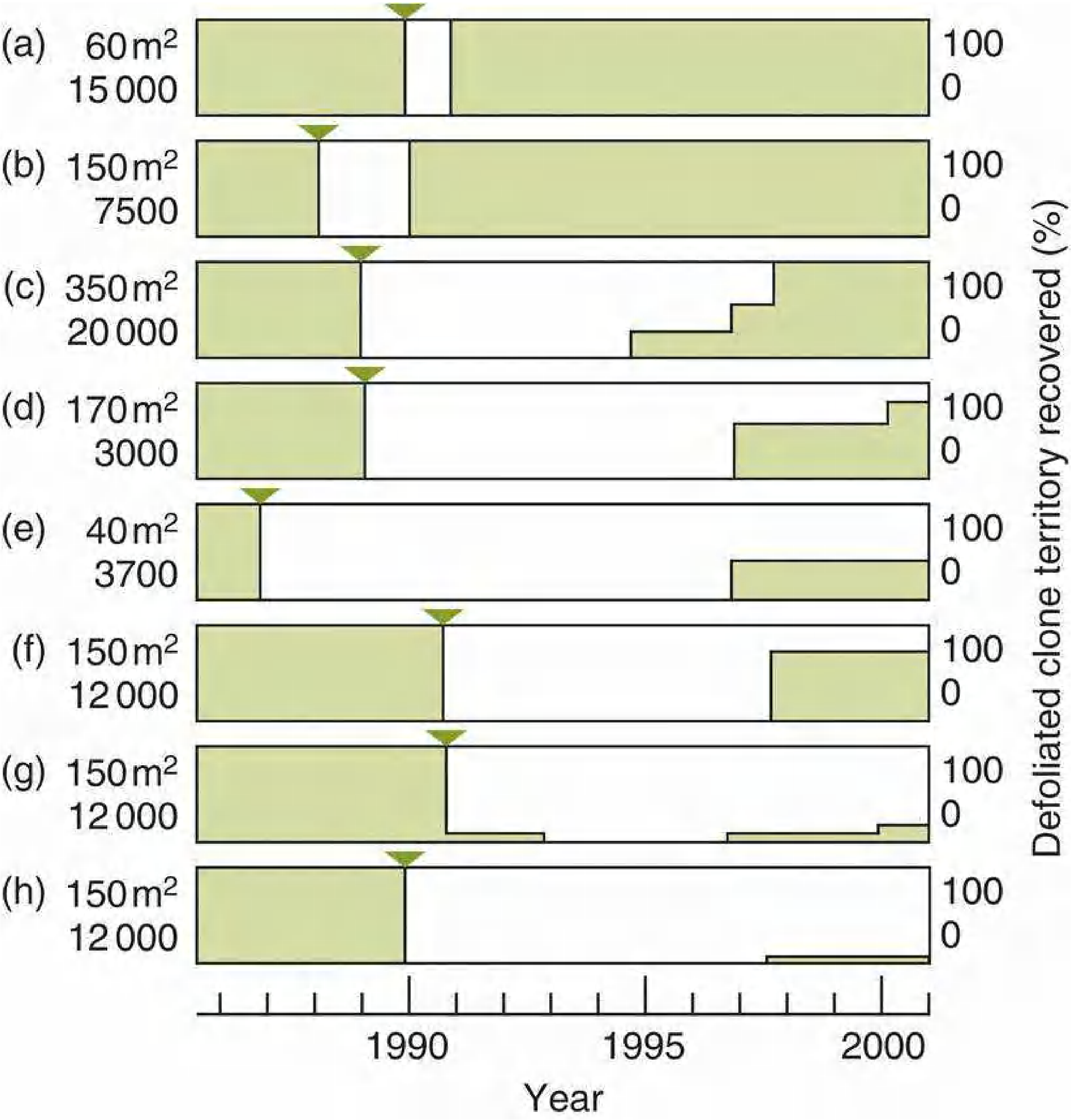


Figure 4.8 **Dormancy in goldenrods is enforced by defoliation.** The histories of eight Missouri goldenrod (*Solidago missouriensis*) clones (rows a–h). Each clone's predefoliation area (m^2) and estimated number of ramets is given on the left. The panels show a 15-year record of the presence (shading) and absence of ramets in each clone's territory. The arrowheads show the beginning of dormancy, initiated by eruption of the beetle *Trirhabda canadensis* and defoliation. Reoccupation of entire or major segments of the original clone's territory by postdormancy ramets is expressed as the percentage of the original clone's territory.

Source: After Morrow & Olfelt (2003).

Most of the species of plants with seeds that persist for long in the soil are annuals and biennials, and they are mainly weedy species – opportunists waiting (literally) for an opening. They largely lack features that will disperse them extensively in space. The seeds of trees, by contrast, usually have a very short expectation of life in the soil, and many are extremely difficult to store artificially for more than one year. The seeds of many tropical trees are particularly short-lived: a matter of weeks or even days. Amongst trees, the most striking longevity is seen in those that retain the seeds in cones or pods on the tree until they are released after fire (many species of *Eucalyptus* and *Pinus*). This phenomenon of serotiny protects the seeds against risks on the ground until fire creates an environment suitable for their rapid establishment.

4.6 Monitoring birth and death: life tables, survivorships curves and fecundity schedules

We turn now to look in more detail at the patterns of birth and death in a variety of life cycles, and at how these patterns are quantified. Often, in order to monitor and examine changing patterns of mortality with age or stage, a *life table* may be drawn up. This allows a *survivorship curve* to be constructed, which traces the decline in numbers, over time, of a group of newly born or newly emerged individuals or modules. It can also be thought of as a plot of the probability, for a representative newly born individual, of surviving to various ages. Patterns of birth amongst individuals of different ages are often monitored at the same time as life tables are constructed. These patterns are displayed in *age-specific fecundity schedules*.

The underlying principles are explained in Figure 4.9. There, a population is portrayed as a series of diagonal lines, each line representing the life 'track' of an individual. As time passes, each individual ages (moves from bottom-left to top-right along its track) and eventually dies (the dot at the end of the track). Here, individuals are classified by their age. In other cases it may be more appropriate to split the life of each individual into different developmental stages.

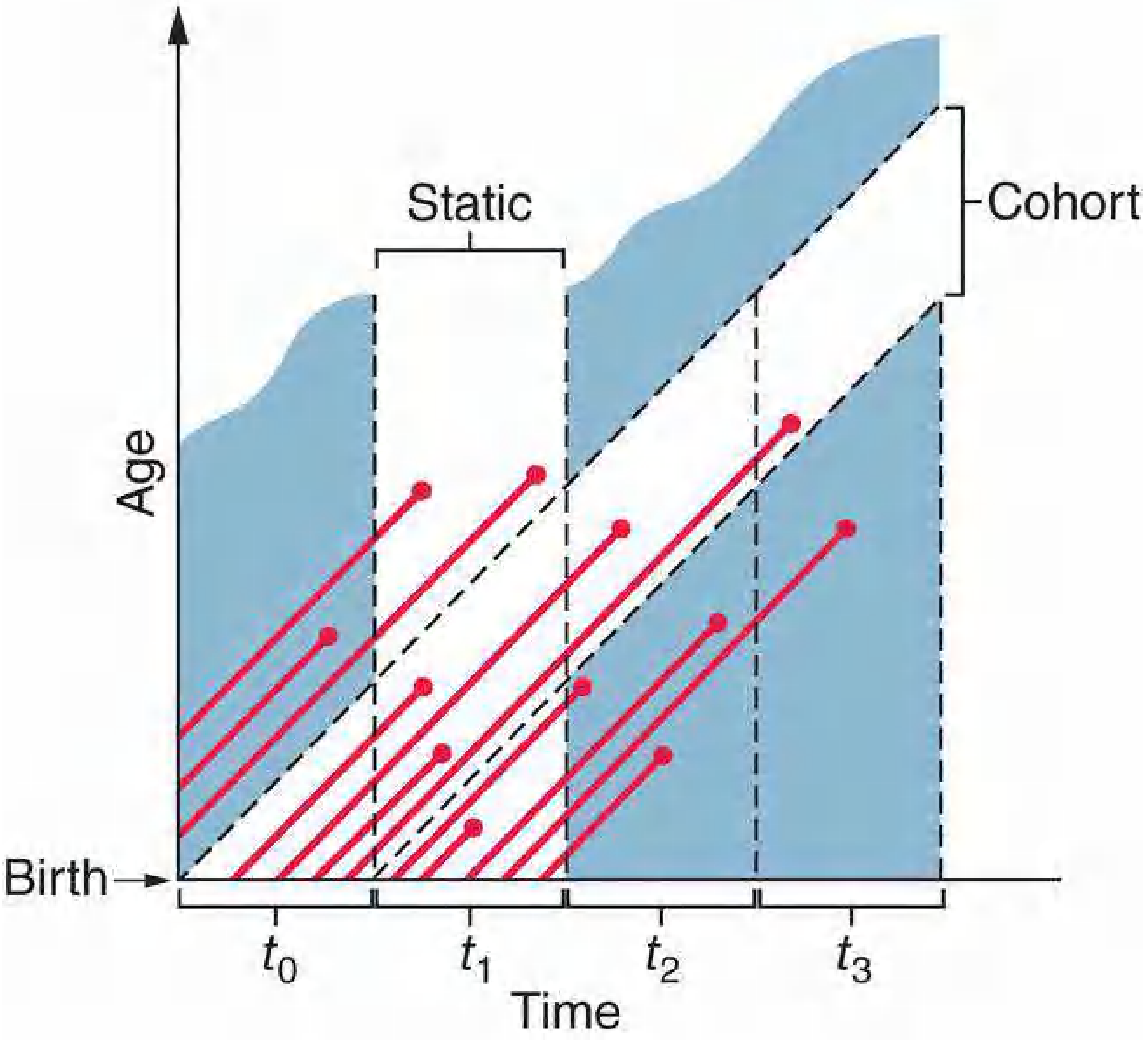


Figure 4.9 Derivation of cohort and static life tables. See text for details.

Time is divided into successive periods: t_0, t_1, etc. In the present case, three individuals were born (started their life track) prior to the time period t_0, four during t_0, and three during t_1. To construct a cohort life table, we direct our attention to a particular cohort and monitor what happens to them subsequently. Here we focus on those born during t_0. The life table is constructed by noting the number surviving to the start of each time period. So, four were there at the beginning of t_1, two of the four survived to the beginning of t_2; only one of these was alive at the beginning of t_3; and none survived to the start of t_4. The first data column of a cohort life table for these individuals would thus comprise the series of declining numbers in the cohort: 4, 2, 1, 0.

A different approach is necessary when we cannot follow cohorts but we know the ages of all the individuals in a population (perhaps from some clue such as the condition of the teeth in a species of deer). We can then, as the figure shows, direct our attention to the whole population during a single period (in this case, t_1) and note the numbers of survivors of different ages in the population. These may be thought of as entries in a life table if we assume that rates of birth and death are, and have previously been, constant – a very big assumption. What results is called a static life table. Here, of the 11 individuals alive during t_1, five were actually born during t_1 and are hence in the youngest age group, four were born in the previous time interval, two in the interval before that, and none in the interval before that. The first data column of the static life table thus comprises the series 5, 4, 2, 0. This amounts to saying that over these time intervals, a typical cohort will have started with five and declined over successive time intervals to four, then two, then zero.

4.6.1 Cohort life tables

To monitor and quantify survival, we may follow the fate of individuals from the same cohort within a population: that is, all individuals born within a particular period. The life table then records the survivorship of the members of the cohort over time (Figure 4.9). The most straightforward life table to construct is a cohort life table for an annual species. Putting to one side the caveats raised above, annual life cycles take approximately 12 months or rather less to complete (Figure 4.6b, c). Usually, every individual in a population breeds during one particular season of the year, but then dies before the same season in the next year. Generations are therefore said to be discrete, and each cohort is distinguishable from every other; the only overlap of generations is between breeding adults and their offspring during and immediately after the breeding season.

an annual life table for a plant

Two very simple life tables, for inland and coastal subspecies of the annual plant *Gilia capitata*, growing in California, USA are shown in Table 4.1. Initial cohorts of around 750 seeds were followed from seed germination to the death of the last adult.

Table 4.1 Two cohort life tables for the annual plant *Gilia capitata*. One is for the ‘inland’ subspecies, *G. capitata capitata*, and one for the ‘coastal’ subspecies, *G. capitata chamissonis*, growing at an inland site in Napa County, California, USA and being easily distinguishable morphologically, despite being cross-fertile. Cohorts of seeds were planted at the beginning of the season in 1993 and the life cycle divided simply into seeds, plants that emerged from those seeds, and emerged plants that went on to flower. Other column entries are explained in the text. *Source*: After Nagy & Rice (1997).

Stage (x)	***Number alive at the start of each age class a_x***	***Proportion of original cohort surviving to the start of each age class l_x***	d_x	q_x	***log*** a_x	***log*** l_x	k_x	***Number of female young produced by each age class F_x***	***Number of female young produced per surviving individual in each age class m_x***	***Number of female young produced per original individual in each age class $l_x m_x$***
Inland subspecies:										
Seed (0)	746	1.00	0.66	0.66	2.87	0.00	0.47	0	0	0
Emergence (1)	254	0.34	0.25	0.74	2.40	–0.47	0.59	0	0	0
Flowering (2)	66	0.09			1.82	–1.05		28,552	432.61	38.29
Coastal subspecies:										
Seed (0)	754	1.00	0.73	0.73	2.88	0.00	0.57	0	0	0
Emergence (1)	204	0.27	0.25	0.91	2.31	–0.57	1.03	0	0	0
Flowering (2)	19	0.03			1.28	–1.60		8645	455.00	11.47

a cohort life table for marmots

Even when generations overlap, if individuals can be marked early in their life so that they can be recognised subsequently, it is feasible to follow the fate of each year’s cohort separately. It may then be possible to merge the cohorts from the different years of a study to derive a single, ‘typical’ cohort life table. An example is shown in Table 4.2 of females from a population of the yellow-bellied marmot, *Marmota flaviventris*. The population was live-trapped and marmots marked individually from 1962 through to 1993 in the East River Valley of Colorado, USA and it was this that allowed each individual to be assigned, whenever it was caught, to its own cohort.

Table 4.2 A cohort life table for female yellow-bellied marmots, *Marmota flaviventris* in Colorado, USA. The columns are explained in the text. *Source*: After Schwartz *et al.* (1998).

Age class (years) x	***Number alive at the start of each age class a_x***	***Proportion of original cohort surviving to the start of each age class l_x***	***d_x***	***q_x***	***log a_x***	***log l_x***	***k_x***	***Number of female young produced by each age class F_x***	***Number of female young produced per surviving individual in each age class m_x***	***Number of female young produced per original individual in each age class l_xm_x***
0	773	1.000	0.457	0.457	2.888	0.00	0.26	0	0.000	0.000
1	420	0.543	0.274	0.505	2.623	−0.26	0.31	0	0.000	0.000
2	208	0.269	0.089	0.332	2.318	−0.57	0.18	95	0.457	0.123
3	139	0.180	0.043	0.237	2.143	−0.75	0.12	102	0.734	0.132
4	106	0.137	0.050	0.368	2.025	−0.86	0.20	106	1.000	0.137
5	67	0.087	0.030	0.343	1.826	−1.06	0.18	75	1.122	0.098
6	44	0.057	0.017	0.295	1.643	−1.24	0.15	45	1.020	0.058
7	31	0.040	0.012	0.290	1.491	−1.40	0.15	34	1.093	0.044
8	22	0.029	0.013	0.455	1.342	−1.55	0.26	37	1.680	0.049
9	12	0.016	0.006	0.417	1.079	−1.81	0.23	16	1.336	0.021
10	7	0.009	0.005	0.571	0.845	−2.04	0.37	9	1.286	0.012
11	3	0.004	0.001	0.333	0,477	−2.41	0.18	0	0.000	0.000
12	2	0.003	0.000	0.000	0.301	−2.59	0.00	0	0.000	0.000
13	2	0.003	0.000	0.000	0.301	−2.59	0.00	0	0.000	0.000
14	2	0.003	0.001	0.500	0.301	−2.59	0.30	0	0.000	0.000
15	1	0.001			0.000	−2.89		0	0.000	0.000
Total								$\overline{519}$		$\overline{0.670}$

the columns of a life table

The first column in each life table is a list of the stages or age classes of the organism's life. For *Gilia*, these are simply the stages 'seed', 'emerged plants', and 'flowering plants'. For the marmots, they are years. The second column is then the raw data from each study, collected in the field. It reports the number of individuals surviving to the beginning of each stage or age class (see Figure 4.9). We refer to these numbers as a_x, where the x in the subscript refers to the stage or age class concerned: a_0 means the numbers in the initial age class, and so on.

Ecologists are typically interested not just in examining populations in isolation but in comparing the dynamics of two or more perhaps rather different populations. This was precisely the case for the *Gilia* populations in Table 4.1. Hence, it is necessary to standardise the raw data so that comparisons can be made. This is done in the third column of the table, which is said to contain l_x values, where l_x is defined as the proportion of the original cohort surviving to the start of age class. The first value in this column, l_0 (spoken: L-zero), is therefore the proportion surviving to the beginning of this original age class. Obviously, in Tables 4.1 and 4.2, and in every life table, l_0 is 1.00 (the whole cohort is there at the start). Thereafter, in the marmots for example, there were 773 females observed in this youngest age class. The l_x values for subsequent age classes are therefore expressed as proportions of this number. Only 420 individuals survived to reach their second year (age class 1: between one and two years of age). Thus, in Table 4.2, the second value in the third column, l_1, is the proportion 420/773 = 0.543 (that is, only 0.543 or 54.3% of the original cohort survived this first step). In the next row, $l_2 = 208/773 = 0.269$, and so on. For *Gilia* (Table 4.1), $l_1 = 254/746 = 0.340$ for the inland subspecies and 204/754 = 0.271 for the coastal subspecies. That is, 34% and 27.1% survived the first step to become established plants in the two cases: a slightly higher survival rate at this inland site for the inland than for the coastal subspecies.

In the next column, to consider mortality more explicitly, the proportion of the original cohort dying during each stage (d_x) is computed, being simply the difference between successive values of l_x; for example, for the marmots, $d_3 = l_3 - l_4 = 0.180 - 0.137 = 0.043$. Next, the stage-specific mortality rate, q_x, is computed. This considers d_x as a fraction of l_x. Hence, q_3 for example is 0.24 (= 0.043/0.180 or d_3/l_3). Values of q_x may also be thought of as the average 'chances' or probabilities of an individual dying during an interval. q_x is therefore equivalent to $(1 - p_x)$ where p refers to the probability of survival.

The advantage of the d_x values is that they can be summed: thus, the proportion of a cohort of marmots dying in the first four years was $d_0 + d_1 + d_2 + d_3$ (= 0.86). The disadvantage is that the individual values give no real idea of the intensity or importance of mortality during a particular stage. This is because the d_x values are larger the more individuals there are, and hence the more there are available to die. The q_x values, on the other hand, are an excellent measure of the intensity of mortality. For instance, in the present example it is clear from the q_x column that the mortality rate declined after the first two years of life but then rose again to a peak around years 9 and 10; this is not clear from the d_x column. The q_x values, however, have the disadvantage that, for example, summing the values over the first four years gives no idea of the mortality rate over that period as a whole.

k values

The advantages are combined, however, in the next column of the life table, which contains k_x values (Haldane, 1949 ; Varley & Gradwell, 1970). k_x is defined simply as the difference between successive values of $\log_{10} a_x$ or successive values of $\log_{10} l_x$ (they amount to the same thing), and is sometimes referred to as a 'killing power'. Like q_x values, k_x values reflect the intensity or rate of mortality (as Tables 4.1 and 4.2 show); but unlike summing the q_x values, summing k_x values is a legitimate procedure. Thus, the killing power or k value for the first four years in the marmot example is 0.26 + 0.31 + 0.18 + 0.12 = 0.87, which is also the difference between $\log_{10} a_0$ and $\log_{10} a_4$ (allowing for rounding errors). Note too that like l_x values, k_x values are standardised, and are therefore appropriate for comparing quite separate studies. In this and later chapters, k_x values will be used repeatedly.

fecundity schedules

Tables 4.1 and 4.2 also include fecundity schedules for *Gilia* and for the marmots (the final three columns). The first of these in each case shows F_x, the total number of the youngest age class produced by each subsequent age class. This youngest class is seeds for *Gilia*, produced only by the flowering plants. For the marmots, these are independent juveniles, fending for themselves outside their burrows, produced when adults were between 2 and 10 years old. The next column is then said to contain m_x values, which is fecundity: the mean number of the youngest age class produced per surviving individual of each subsequent class. For the marmots, fecundity was highest for eight-year-old females: 1.68, that is, 37 young produced by 22 surviving females. We get a good idea of the range of fecundity schedules in Figure 4.2: some with constant fecundity throughout most of an individual's life, some in which there is a steady increase with age, some with an early peak followed by an extended postreproductive phase. We try to account for some of this variation in the next chapter.

… combined to give the basic reproductive rate

In the final column of a life table, the l_x and m_x columns are brought together to express the overall extent to which a population increases or decreases over time – reflecting the dependence of this on both the survival of individuals (the l_x column) and the reproduction of those survivors (the m_x column). That is, an age class contributes most to the next generation when a large proportion of individuals have survived and they are highly fecund. The sum of all the l_x m_x values, $\sum l_x m_x$, where the symbol $\sum$ means 'the sum of', is therefore a measure of the overall extent by which this population has increased or decreased in a generation. We call this the basic reproductive rate and denote it by R_0 ('*R*-nought'). That is:

$$R_0 = \sum l_x m_x. \tag{4.2}$$

We can also calculate R_0 by dividing the total number of offspring produced during one generation ($\sum F_x$, meaning the sum of the values in the F_x column) by the original number of individuals. That is:

$$R_0 = \sum F_x / a_0. \tag{4.3}$$

For *Gilia* (Table 4.1), R_0 is calculated very simply (no summation required) since only the flowering class produces seed. Its value is 38.27 for the inland subspecies and 11.47 for the coastal subspecies: a clear indication that the inland subspecies thrived, comparatively, at this inland site. (Though the annual rate of reproduction would not have been this high, since, no doubt, a proportion of these would have died before the start of the 1994 cohort. In other words, another class of individuals, 'winter seeds', was ignored in this study.)

For the marmots, R_0 = 0.67: the population was declining, each generation, to around two-thirds its former size. However, whereas for *Gilia* the length of a generation is obvious, since there is one generation each year, for the marmots the generation length must itself be calculated. We address the question of how to do this in Section 4.7, but for now we can note that its value, 4.5 years, matches what we can observe ourselves in the life table: that a 'typical' period from an individual's birth to giving birth itself (i.e. a generation) is around four and a half years. Thus, Table 4.2 indicates that each generation, every four and a half years, this particular marmot population was declining to around two-thirds its former size.

4.6.2 Survivorship curves

It is also possible to study the detailed pattern of decline in a cohort. Figure 4.10a, for example, shows the numbers of marmots surviving relative to the original population – the l_x values – plotted against the age of the cohort. However, this can be misleading. If the original population is 1000 individuals, and it decreases by half to 500 in one time interval, then this decrease looks more dramatic on a graph like Figure 4.9a than a decrease from 50 to 25 individuals later in the season. Yet the risk of death to individuals is the same on both occasions. If, however, l_x values are replaced by log(l_x) values, that is, the logarithms of the values, as in Figure 4.10b (or, effectively the same thing, if l_x values are plotted on a log scale), then it is a characteristic of logs that the reduction of a population to half its original size will always look the same. *Survivorship curves* are, therefore, conventionally plots of log(l_x) values against cohort age. Figure 4.10b shows that for the marmots, there was a steady, more or less constant rate of decline until around the eighth year of life, then three further years at a slightly higher rate (until breeding ceased), followed by a brief period with effectively no mortality, after which the few remaining survivors died.

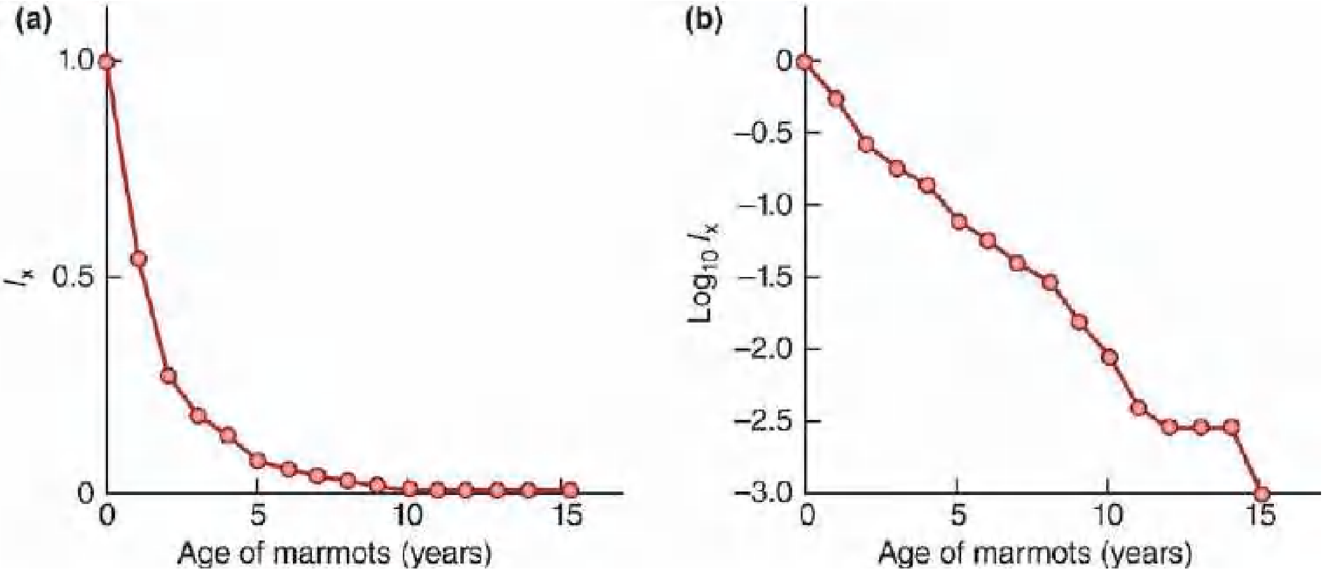


Figure 4.10 Representations of the survival of a cohort of the yellow-bellied marmot (Table 4.2). (a) When l_x is plotted against cohort age, it is clear that most individuals are lost relatively early in the life of the cohort, but there is no clear impression of the risk of mortality at different ages. (b) By contrast, a survivorship curve plotting log(l_x) against age shows a virtually constant mortality risk until around age eight, followed by a brief period of slightly higher risk, and then another brief period of low risk after which the remaining survivors died.

a classification of survivorship curves

Life tables provide a great deal of data on specific organisms. But ecologists search for generalities – patterns of life and death that we can see repeated in the lives of many species – conventionally dividing survivorship curves into three types in a scheme that goes back to 1928, generalising what we know about the way in which the risks of death are distributed through the lives of different organisms (Figure 4.11).

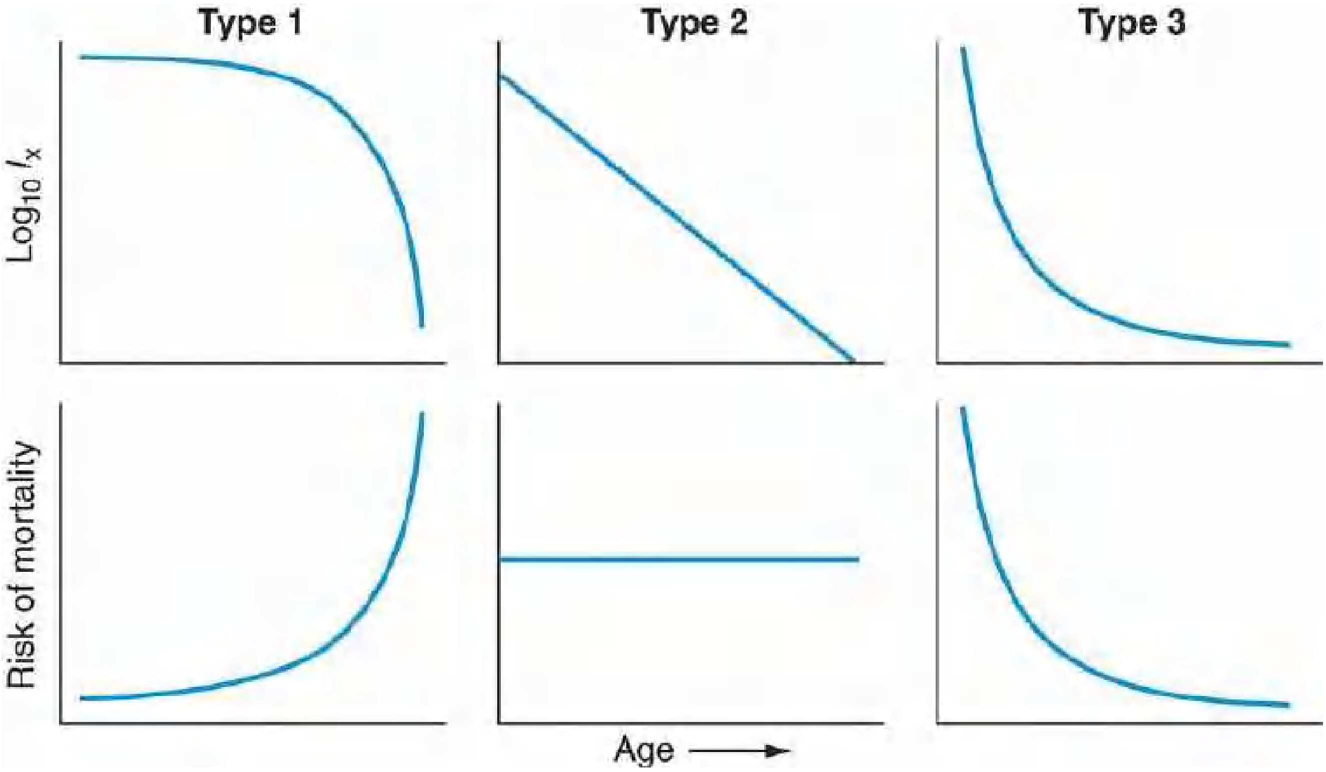


Figure 4.11 **Classification of survivorship curves** plotting $\log(l_x)$ against age, above, with corresponding plots of the changing risk of mortality with age, below. The three types are discussed in the text.

Source: After Pearl (1928) and Deevey (1947).

In a type 1 survivorship curve, mortality is concentrated toward the end of the maximum life span. It is perhaps most typical of humans in developed countries and their carefully tended zoo animals and pets. A type 2 survivorship curve is a straight line signifying a constant mortality rate from birth to maximum age. It describes, for instance, the survival of buried seeds in a seed bank. In a type 3 survivorship curve there is extensive early mortality, but a high rate of subsequent survival. This is typical of species that produce many offspring. Few survive initially, but once individuals reach a critical size, their risk of death remains low and more or less constant. This appears to be the most common survivorship curve among animals and plants in nature.

These types of survivorship curve are useful generalisations, but in practice, patterns of survival are usually more complex. We saw with the marmots, for example, that survivorship was broadly type 2 throughout much of their lives, but not at the end (Figure 4.10b). Similarly, with the dinosaurs we will meet in the next section, survivorship followed the typical type 3 pattern until they reached sexual maturity, but again failed to conform to such a simple classification thereafter (see Figure 4.13). More generally, we see examples approximating to each of the three types in the survey in Figure 4.2, but also more examples where the shape changes as individuals pass through the different phases of their lives.

APPLICATION 4.2 The survivorship curves of captive mammals

Opinions naturally differ regarding both the ethics and the practical benefits of keeping wild animals in captivity in zoos, but the current reality is that zoos play an integral role in the conservation of many species, especially those, like many mammals, that are large and inherently attractive to the general public. Hence, in managing these animals, we need to understand their patterns of survivorship, and to know in particular if there are general rules organising these patterns that would not only describe the species for which we have good data, but also allow us to predict patterns for similar or related species when currently available data are sparse. Lynch *et al.* (2010) therefore reviewed what was known about the survivorship of captive mammals – 37 species, including primates, artiodactyls (cattle, sheep, deer, etc.), carnivores, bats, seals and the giant panda – and some of their results are summarised in Figure 4.12. They were more interested in the shapes of the survivorship curves (and for example whether they were type 1, 2 or 3) than in absolute values, and all data sets were therefore scaled to the maximum longevity of the species concerned. They then fitted all datasets to a general survivorship function with two shape parameters, α and β, which allowed the different curves to be classified and either grouped together or distinguished (Figure 4.12). Broadly speaking, with increasing values of α/β, mortality shifted towards being more evenly distributed throughout life, rather than being concentrated at the start; and with decreasing values of $\alpha\beta$, mortality shifted towards including senescence – a period of increased mortality at the end of life – rather than decreasing steadily with age.

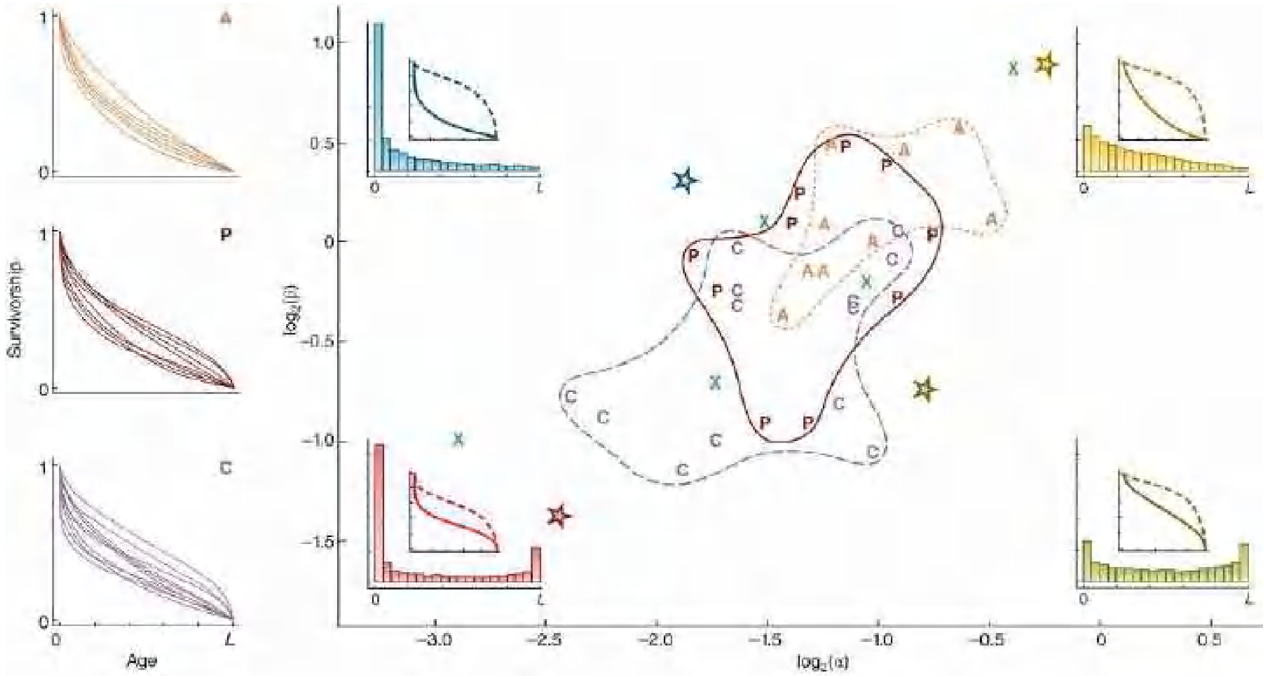


Figure 4.12 Distribution of the shapes of survivorship curves for 37 species of animals kept in zoos. (For a full list of species names, see the original text.) A generalised survivorship function with two parameters, α and β was fitted to all datasets, allowing each to be located in $\log\alpha$–$\log\beta$ space. The shapes themselves are illustrated in the insets, referring to the four starred locations, as survivorship on linear and semilogarithmic scales (solid and dashed lines, respectively; see Figure 4.11) and as distributions of mortality (histograms) between birth and maximum longevity (L). Among those species, the locations of artiodactyls (A), carnivores (C) and primates (P) are indicated, plus five further species (X).

Source: After Lynch *et al.* (2010).

Despite wide variations in size, longevity and taxonomic affiliation, most of the curves were, in essence, type 2, with some element of type 1 (senescence) or type 3 (early mortality). The variation that did exist was significantly associated with the species' taxonomic order: the artiodactyls showed the least evidence of senescence, the carnivores the most, with the primates somewhere in between (Figure 4.12). This taxonomic variation was in turn associated with variations in age to weaning (relative to lifespan) and litter size, suggesting 'syndromes' of associated life history traits. We return to the whole topic of the patterns in life histories and their possible causes in the next chapter. For now, though, the results do provide us with grounds for believing that, based on this analysis, even for species where we have little or no prior knowledge, managers in zoos can make educated predictions with some confidence about likely patterns of mortality, and act accordingly.

4.6.3 Static life tables

Many of the species that ecologists study, and for which life tables would therefore be valuable, have repeated breeding seasons like the marmots, or continuous breeding as in the case of humans, but constructing life tables here is complicated, largely because these populations have individuals of many different ages living together. Building a cohort life table is sometimes possible, as we have seen, but this is relatively uncommon. Apart from the mixing of cohorts in the population, it can be difficult simply because of the longevity of many species.

useful – if used with caution

Another approach is to construct a static life table (Figure 4.9). The data look like a cohort life table – a series of different numbers of individuals in different age classes – but these come simply from the *age structure* of the population captured at one point in time. Hence, great care is required: they can only be treated and interpreted in the same way as a cohort life table if patterns of birth and survival in the population have remained much the same since the birth of the oldest individuals – and this will happen only rarely. Nonetheless, there is often no alternative and useful insights can still be gained. This is illustrated for a population of small dinosaurs, *Psittacosaurus lujiatunensis*, recovered as fossils from the Lower Cretaceous Yixian Formation in China, where the alternative of following a cohort is obviously not available (Figure 4.13). They appear to have perished simultaneously in a volcanic mudflow, which might therefore have captured a representative snapshot of the population at the time, around 125 million years ago.

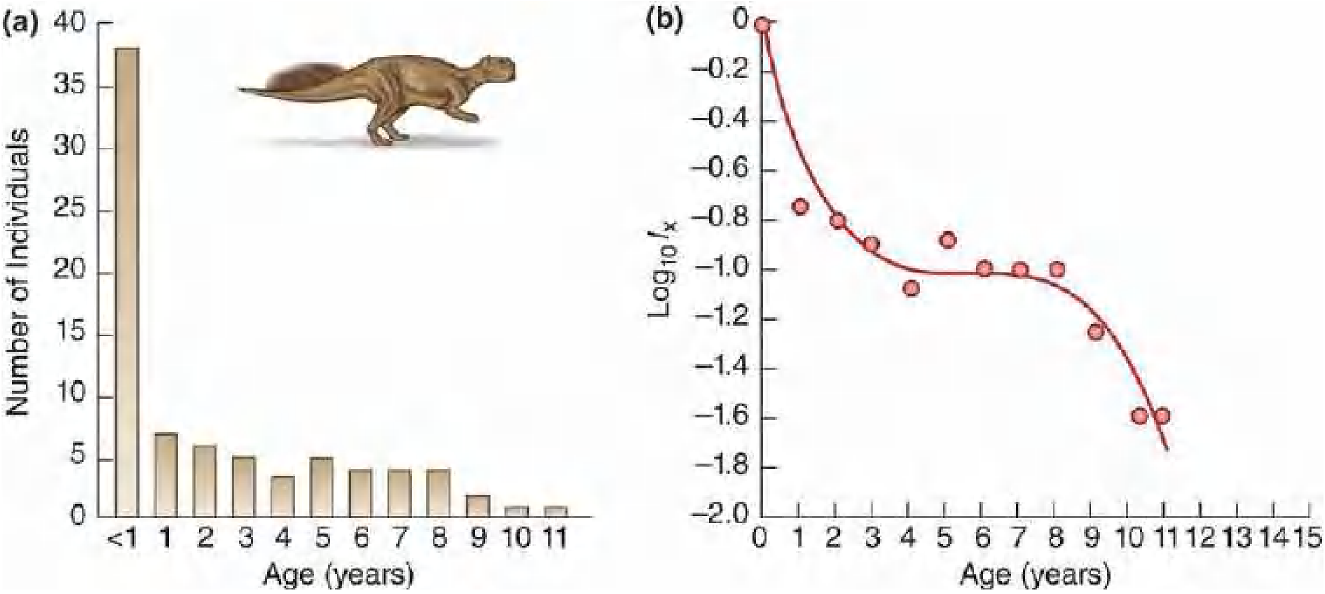


Figure 4.13 Static life tables can be informative, especially when alternatives are not available. (a) The age structure (and hence the static life table) of a population of dinosaurs, *Psittacosaurus lujiatunensis*, recovered as fossils from the Lower Cretaceous Yixian Formation in China. Age was estimated from the length of the femur, which had been shown in a subsample of specimens to correlate very strongly with the number of 'growth lines' (one per year) in the bone. (b) A survivorship curve (log(l_x) plotted against age) derived from the life table.

Source: After Erickson *et al*. (2009).

It appears that mortality rates were high amongst the dinosaurs until around the age of three, after which there was another period of around five years during which mortality rates were low even though the animals continued to grow rapidly, which they did until the age of around nine or 10 years (Figure 4.13b). Mortality rates then seem to have increased again, just as the animals were attaining their maximum size, and broadly coinciding with the appearance in the fossils of characteristics associated with sexual maturity (e.g. enlarged, flaring 'jugal' horns). As we shall see in the next chapter, many organisms suffer a cost of reproduction in terms of reductions in growth and/or survival.

Notwithstanding this successful use of a static life table, the interpretation of static life tables generally, and the age structures from which they stem, is fraught with difficulty: usually, age structures offer no easy short cuts to understanding the dynamics of populations.

4.6.4 The importance of modularity

Finally here, we turn to the difficulties of constructing any sort of life table for organisms that are not only iteroparous with overlapping generations but are also modular. We can use a study of the sedge, *Carex bigelowii*, growing in a lichen heath in Norway, to illustrate this (Figure 4.14). *C. bigelowii* has an extensive underground rhizome system that produces tillers (aerial shoots) at intervals along its length as it grows. It grows by producing a lateral meristem in the axil of a leaf belonging to a 'parent' tiller. This lateral is completely dependent on the parent tiller at first, but is potentially capable of developing into a vegetative parent tiller itself, and also of flowering, which it does when it has produced a total of 16 or more leaves. Flowering, however, is always followed by tiller death. In other words, the tillers are semelparous although the genets are iteroparous.

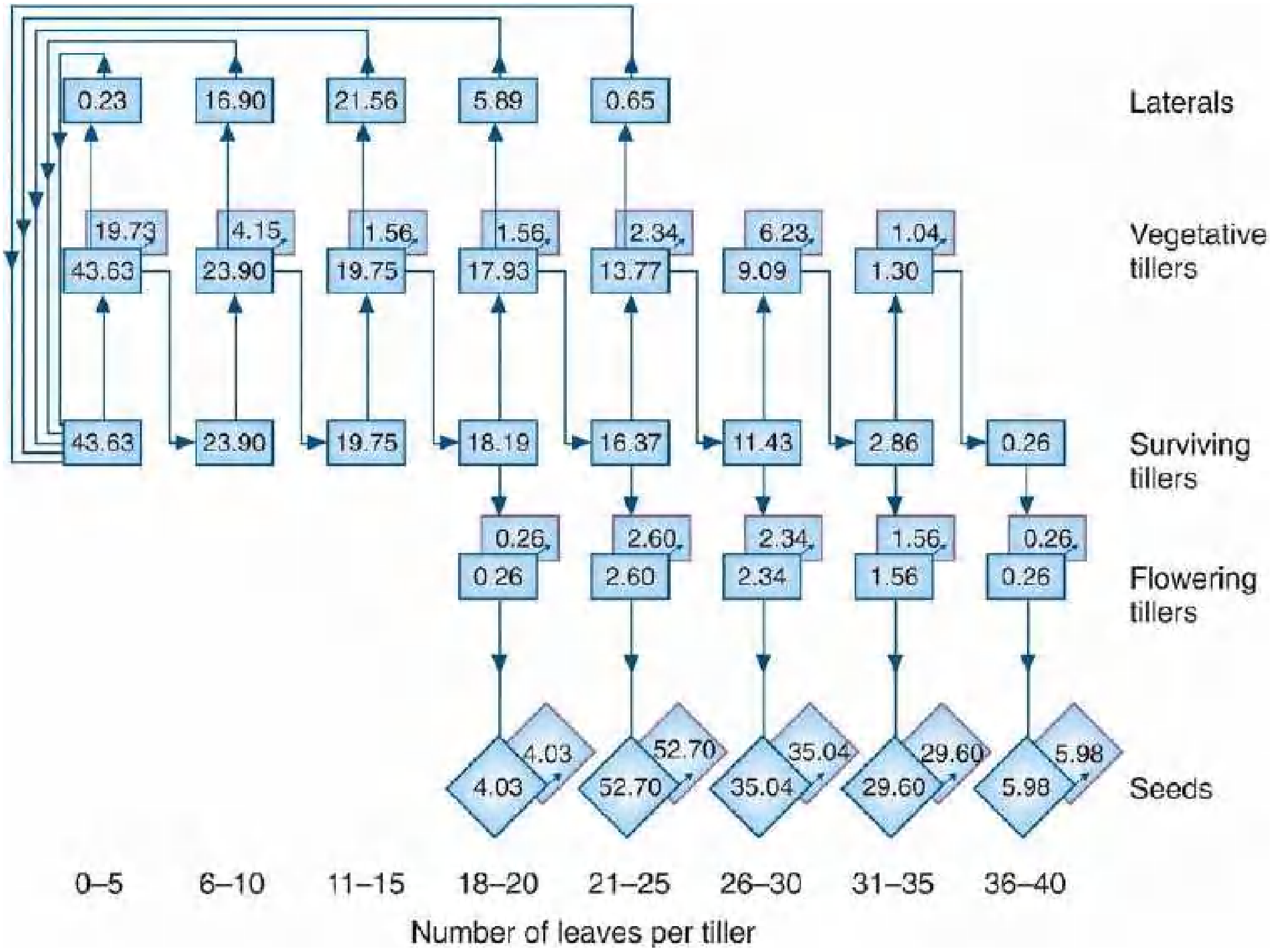


Figure 4.14 Reconstructed static life table for the modules (tillers) of a *Carex bigelowii* population. The densities per m^2 of tillers are shown in rectangular boxes, and those of seeds in diamond-shaped boxes. Rows represent tiller types, whilst columns depict size classes of tillers. Purple-walled boxes represent dead tiller (or seed) compartments, and arrows denote pathways between size classes, death or reproduction.

Source: After Callaghan (1976).

Callaghan (1976) took a number of well-separated young tillers and excavated their rhizome systems through progressively older generations of parent tillers. This was made possible by the persistence of dead tillers. He excavated 23 such systems containing a total of 360 tillers, and was able to construct a type of static life table (and fecundity schedule) based on the growth stages (Figure 4.14). There were, for example, 1.04 dead vegetative tillers (per m^2) with 31–35 leaves. Thus, since there were also 0.26 tillers in the next (36–40 leaves) stage, it can be assumed that a total of 1.30 (i.e. 1.04 + 0.26) living vegetative tillers entered the 31–35 leaf stage. As there were 1.30 vegetative tillers and 1.56 flowering tillers in the 31–35 leaf stage, 2.86 tillers must have survived from the 26–30 stage, and so on. In this way, a life table – applicable not to individual genets but to tillers (i.e. modules) – was constructed.

There appeared to be no new establishment from seed in this particular population (no new genets); tiller numbers were being maintained by modular growth alone. However, a ‘modular growth schedule’ (*laterals*), analogous to a fecundity schedule, has been constructed.

Note finally that stages rather than age classes have been used here – something that is almost always necessary when dealing with modular iteroparous organisms, because variability stemming from modular growth accumulates year upon year, making age a particularly poor measure of an individual’s chances of death, reproduction or further modular growth.

4.7 Reproductive rates, generation lengths and rates of increase

4.7.1 Relationships between the variables

In the previous section we saw that the life tables and fecundity schedules drawn up for species with overlapping generations are at least superficially similar to those constructed for species with discrete generations. With discrete generations, we were able to compute the basic reproductive rate (R_0) as a summary term describing the overall outcome of the patterns of survivorship and fecundity. Can a comparable summary term be computed when generations overlap?

Note immediately that previously, for species with discrete generations, R_0 described two separate population parameters. It was the number of offspring produced on average by an individual over the course of its life; but it was also the multiplication factor that converted an original population size into a new population size, one generation hence. With overlapping generations, when a cohort life table is available, the basic reproductive rate can be calculated using the same formula:

$$R_0 = \sum l_x m_x, \tag{4.4}$$

and it still refers to the average number of offspring produced by an individual. But further manipulations of the data are necessary before we can talk about the rate at which a population increases or decreases in size, or, for that matter, about the length of a generation. The difficulties are much greater still when only a static life table (i.e. an age structure) is available (see later).

the fundamental net reproductive rate, *R*

We begin by deriving a general relationship that links population size, the rate of population increase, and time – but which is not limited to measuring time in terms of generations. Imagine a population that starts with 10 individuals, and which, after successive intervals of time, rises to 20, 40, 80, 160 individuals and so on. We refer to the initial population size as N_0 (meaning the population size when no time has elapsed). The population size after one time interval is N_1, after two time intervals it is N_2, and in general after *t* time intervals it is N_t. In the present case, $N_0 = 10$, $N_1 = 20$, and we can say that:

$$N_1 = N_0 R, \tag{4.5}$$

where *R*, which is 2 in the present case, is known as the *fundamental net reproductive rate* or the *fundamental net per capita rate of increase*. Clearly, populations will increase when $R > 1$, and decrease when $R < 1$. (Unfortunately, the ecological literature is somewhat divided between those who use '*R*' and those who use the symbol λ for the same parameter. Here we stick with *R*, but we sometimes use λ in later chapters to conform to standard usage within the topic concerned.)

R combines the birth of new individuals with the survival of existing individuals. Thus, when $R = 2$, each individual could give rise to two offspring but die itself, or give rise to only one offspring and remain alive: in either case, *R* (birth plus survival) would be 2. Note too that in the present case *R* remains the same over the successive intervals of time, i.e. $N_2 = 40 = N_1 R$, $N_3 = 80 = N_2 R$, and so on. Thus:

$$N_3 = N_1 R \times R = N_0 R \times R \times R = N_0 R^3, \tag{4.6}$$

and in general terms:

$$N_{t+1} = N_t R, \tag{4.7}$$

and:

$$N_t = N_0 R^t. \tag{4.8}$$

R, R_0 and *T*

Equations 4.7 and 4.8 link together population size, rate of increase and time; and we can now link these in turn with R_0, the basic reproductive rate, and with the generation length (defined as lasting *T* intervals of time). In Section 4.6.1, we saw that R_0 is the multiplication factor that converts one population size to another population size, one generation later, i.e. T time intervals later. Thus:

$$N_T = N_0 R_0. \tag{4.9}$$

But we can see from Equation 4.8 that:

$$N_T = N_0 R^T. \tag{4.10}$$

Therefore:

$$R_0 = R^T, \tag{4.11}$$

or, if we take natural logarithms of both sides:

$$\ln R_0 = T \ln R. \tag{4.12}$$

r, the intrinsic rate of natural increase

The term ln R is usually denoted by r, the *intrinsic rate of natural increase*. It is the rate at which the population increases in size – the change in population size per individual per unit time. Clearly, populations will increase in size for $r > 0$, and decrease for $r < 0$; and we can note from the preceding equation that:

$$r = \ln R_0 / T. \tag{4.13}$$

Summarising so far, we have a relationship between the average number of offspring produced by an individual in its lifetime, R_0, the increase in population size per unit time, r (= ln R), and the generation time, T. Previously, with discrete generations (see Section 4.5.2), the unit of time *was* a generation. It was for this reason that R_0 was the same as R.

4.7.2 Estimating the variables from life tables and fecundity schedules

In populations with overlapping generations (or continuous breeding), r is the intrinsic rate of natural increase that the population has the *potential* to achieve; but it will only actually achieve this rate of increase if the survivorship and fecundity schedules remain steady over a long period of time. If they do, r will be approached gradually (and thereafter maintained), and over the same period the population will gradually approach a stable age structure (i.e. one in which the proportion of the population in each age class remains constant over time; see below). If, on the other hand, the fecundity and survivorship schedules alter over time – as they almost always do – then the rate of increase will continually change, and it will be impossible to characterise in a single figure. Nevertheless, it can often be useful to characterise a population in terms of its potential, especially when the aim is to make a comparison, for instance comparing various populations of the same species in different environments, to see which environment appears to be the most favourable for the species.

The most precise way to calculate r is from the equation:

$$\sum e^{-rx} l_x m_x = 1, \tag{4.14}$$

where the l_x and m_x values are taken from a cohort life table, and e is the base of natural logarithms. However, this is a so-called 'implicit' equation, which cannot be solved directly (only by iteration, usually on a computer). It is therefore customary to use instead an approximation to Equation 4.13, namely:

$$r \approx \ln R_0 / T_c, \tag{4.15}$$

where T_c is the *cohort generation time* (see below). This equation shares with Equation 4.13 the advantage of making explicit the dependence of r on the reproductive output of individuals (R_0) and the length of a generation (T). Equation 4.15 is a good approximation when $R_0 \approx 1$ (i.e. population size stays approximately constant), or when there is little variation in generation length, or for some combination of these two things (May, 1976).

We can estimate r from Equation 4.15 if we know the value of the cohort generation time T_c, which is the average length of time between the birth of an individual and the birth of one of its own offspring. This, being an average, is the sum of all these birth-to-birth times, divided by the total number of offspring, i.e.:

$$T_c = \sum x l_x m_x / \sum l_x m_x$$

or

$$T_c = \sum x l_x m_x / R_0. \tag{4.16}$$

This is only approximately equal to the true generation time T, because it takes no account of the fact that some offspring may themselves develop and give birth during the reproductive life of the parent.

Thus Equations 4.15 and 4.16 allow us to calculate T_c, and thus an approximate value for r, from a cohort life table of a population with either overlapping generations or continuous breeding. In short, they give us the summary terms we require. A worked example is set out in Table 4.3, using data for the barnacle *Balanus glandula*. Note that the precise value of r, from Equation 4.14, is 0.085, compared with the approximation 0.080; whilst T, calculated from Equation 4.13, is 2.9 years compared with T_c = 3.1 years. The simpler and biologically transparent approximations are clearly satisfactory in this case. They show that since r was somewhat greater than zero, the population would have increased in size, albeit rather slowly, if the schedules had remained steady. Alternatively, we may say that, as judged by this cohort life table, the barnacle population had a good chance of continued existence.

Table 4.3 A cohort life table and a fecundity schedule for the barnacle *Balanus glandula* at Pile Point, San Juan Island, Washington, USA. The computations for R_0, T_c and the approximate value of r are explained in the text. Numbers marked with an asterisk were interpolated from the survivorship curve. *Source*: After Connell (1970).

Age (years) x	a_x	l_x	m_x	$l_x m_x$	$x l_x m_x$
0	1 000 000	1.000	0	0	
1	62	0.0000620	4600	0.285	0.285
2	34	0.0000340	8700	0.296	0.592
3	20	0.0000200	11 600	0.232	0.696
4	15.5*	0.0000155	12 700	0.197	0.788
5	11	0.000110	12 700	0.140	0.700
6	6.5*	0.0000065	12 700	0.082	0.492
7	2	0.0000020	12 700	0.025	0.175
8	2	0.0000020	12 700	0.025	0.200
				1.282	3.928

$$R_0 = 1.282; T_c = \frac{3.928}{1.282} = 3.1; r \approx \frac{\ln R_0}{T_c} = 0.08014.$$

4.8 Population projection models

4.8.1 Population projection matrices

A more general, more powerful, and therefore more useful method of analysing and interpreting the fecundity and survival schedules of a population with overlapping generations makes use of the *population projection matrix* (see Caswell, 2001 , for a full exposition). The word 'projection' in its title is important. Just like the simpler methods above, the idea is not to take the current state of a population and forecast what will happen to the population in the future, but to project forward to what would happen if the schedules remained the same. Caswell uses the analogy of the speedometer in a car: it provides us with an invaluable piece of information about the car's current state, but a reading of, say, 80 km h^{-1} is simply a projection, not a serious forecast that we will actually have travelled 80 km in one hour's time.

life cycle graphs

The population projection matrix acknowledges that most life cycles comprise a sequence of distinct classes with different rates of fecundity and survival: life cycle stages, perhaps, or size classes, rather than simply different ages. The resultant patterns can be summarised in a 'life cycle graph', though this is not a graph in the everyday sense but a flow diagram depicting the transitions from class to class over each step in time. Two examples are shown in Figure 4.15 (see also Caswell, 2001). The first (Figure 4.15a) indicates a straightforward sequence of classes where, over each time step, individuals in class i may: (i) survive and remain in that class (with probability p_i); (ii) survive and grow and/or develop into the next class (with probability g_i); and (iii) give birth to m_i newborn individuals into the youngest/smallest class. Figure 4.15b then goes on to show that a life cycle graph can also depict a more complex life cycle, for example with both sexual reproduction (here, from reproductive class 4 into 'seed' class 1) and vegetative growth of new modules (here, from 'mature module' class 3 to 'new module' class 2). Note that the notation here is slightly different from that in life tables like Table 4.2. There the focus was on age classes, and the passage of time inevitably meant the passing of individuals from one age class to the next: p values therefore referred to survival from one age class to the next. Here, by contrast (as in Table 4.1), an individual need not pass from one class to the next over a time step, and it is therefore necessary to distinguish survival within a class (p values here) from passage and survival into the next class (g values).

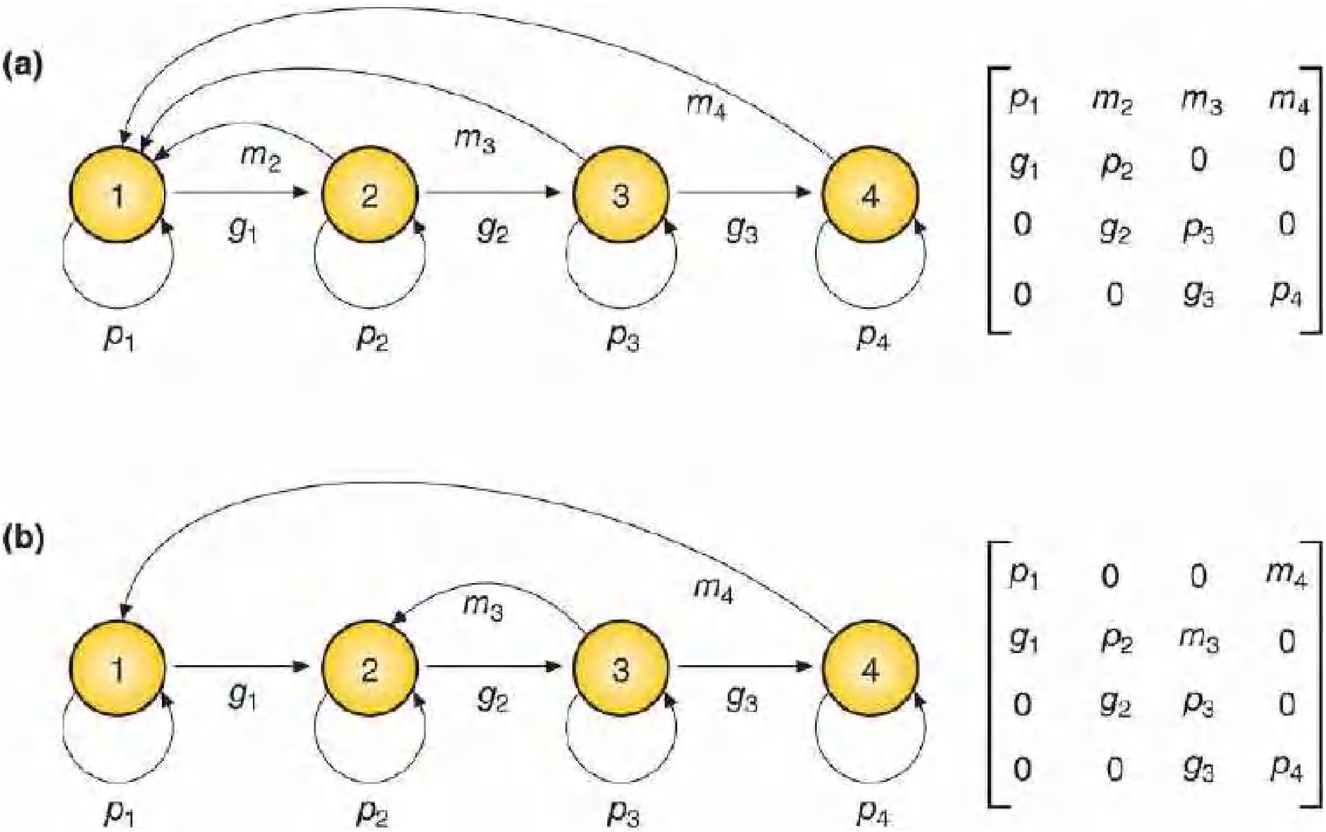


Figure 4.15 Life cycle graphs and population projection matrices for two different life cycles. The connection between the graphs and the matrices is explained in the text. (a) A life cycle with four successive classes. Over one time step, individuals may survive within the same class (with probability p_1), survive and pass to the next class (with probability g_1) or die, and individuals in classes 2, 3 and 4 may give birth to individuals in class 1 (with per capita fecundity m_1). (b) Another life cycle with four classes, but in this case only reproductive class 4 individuals can give birth to class 1 individuals, but class 3 individuals can 'give birth' (perhaps by vegetative growth) to further class 2 individuals.

the elements of the matrix

The information in a life cycle graph can be summarised in a population projection matrix. Such matrices are shown alongside the graphs in Figure 4.15. The convention is to contain the elements of a matrix within square brackets. In fact, a projection matrix is itself always 'square': it has the same number of columns as rows. The rows refer to the class number at the endpoint of a transition: the columns refer to the class number at the start. Thus, for instance, the matrix element in the third row of the second column describes the flow of individuals from the second class into the third class. More specifically, then, and using the life cycle in Figure 4.15a as an example, the elements in the main diagonal from top left to bottom right represent the probabilities of surviving and remaining in the same class (the *p*s), the elements in the remainder of the first row represent the fecundities of each subsequent class into the youngest class (the *m*s), while the *g*s, the probabilities of surviving and moving to the next class, appear in the subdiagonal below the main diagonal (from 1 to 2, from 2 to 3, etc.).

Summarising the information in this way is useful because, using standard rules of matrix manipulation, we can take the numbers in the different classes (n_1, n_2, etc.) at one point in time (t_1), expressed as a 'column vector' (simply a matrix comprising just one column), *pre*-multiply this vector by the projection matrix, and generate the numbers in the different classes one time step later (t_2). The mechanics of this – that is, where each element of the new column vector comes from – are as follows:

$$\begin{bmatrix} p_1 & m_2 & m_3 & m_4 \\ g_1 & p_2 & 0 & 0 \\ 0 & g_2 & p_3 & 0 \\ 0 & 0 & g_3 & p_4 \end{bmatrix} \times \begin{bmatrix} n_{1,t1} \\ n_{2,t1} \\ n_{3,t1} \\ n_{4,t1} \end{bmatrix} = \begin{bmatrix} n_{1,t2} \\ n_{2,t2} \\ n_{3,t2} \\ n_{4,t2} \end{bmatrix}$$

$$= \begin{bmatrix} (n_{1,t1} \times p_1) + (n_{2,t1} \times m_2) + (n_{3,t1} \times m_3) + (n_{4,t1} \times m_4) \\ (n_{1,t1} \times g_1) + (n_{2,t1} \times p_2) + (n_{3,t1} \times 0) + (n_{4,t1} \times 0) \\ (n_{1,t1} \times 0) + (n_{2,t1} \times g_2) + (n_{3,t1} \times p_3) + (n_{4,t1} \times 0) \\ (n_{1,t1} \times 0) + (n_{2,t1} \times 0) + (n_{3,t1} \times g_3) + (n_{4,t1} \times p_4) \end{bmatrix}$$

determining *R* from a matrix

Thus, the numbers in the first class, n_1, are the survivors from that class one time step previously plus those born into it from the other classes, and so on. Figure 4.16 shows this process repeated 20 times (i.e. for 20 time steps) with some hypothetical values in the projection matrix shown as an inset in the figure. It is apparent that there is an initial (transient) period in which the proportions in the different classes alter, some increasing and others decreasing, but that after about nine time steps, all classes grow at the same exponential rate (a straight line on a logarithmic scale, see Section 5.6), and so therefore does the whole population. The *R* value in this case is 1.25. Also, the proportions in the different classes are constant: the population has achieved a stable class structure with numbers in the ratios 51.5 : 14.7 : 3.8 : 1.

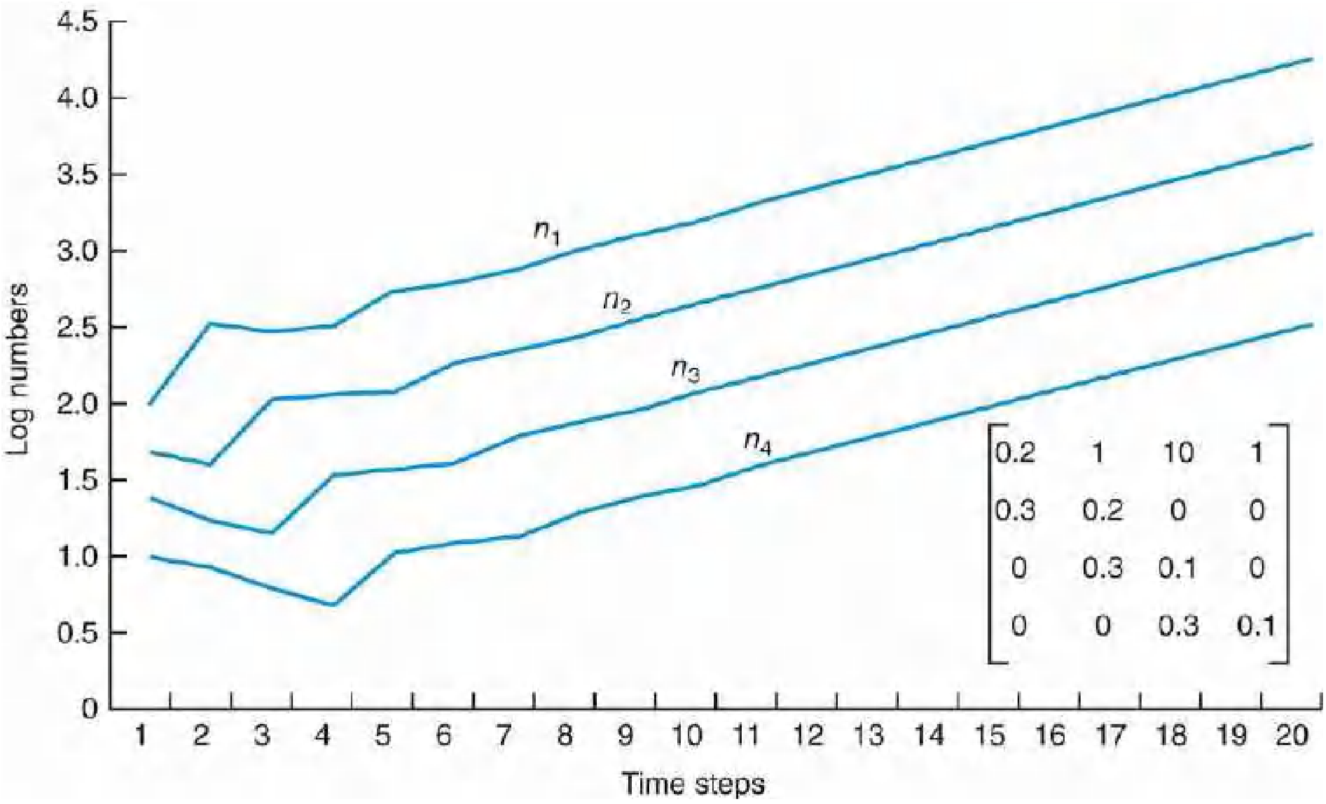


Figure 4.16 Populations with constant rates of survival and fecundity eventually reach a constant rate of growth and a stable age structure. A population growing according to the life cycle graph shown in Figure 4.15a, with parameter values as shown in the insert here. The starting conditions were 100 individuals in class 1 ($n_1 = 100$), 50 in class 2, 25 in class 4 and 10 in class 4. On a logarithmic (vertical) scale, exponential growth appears as a straight line. Thus, after about 10 time steps, the parallel lines show that all classes were growing at the same rate ($R = 1.25$) and that a stable class structure had been achieved.

Hence, a population projection matrix allows us to summarise a potentially complex array of survival, growth and reproductive processes, and characterise that population succinctly by determining the per capita rate of increase, *R*, implied by the matrix. But crucially, this 'asymptotic' *R* can be determined directly, without the need for a simulation, by application of the methods of matrix algebra (these are beyond our scope here, but see Caswell (2001)). Moreover, such algebraic analysis can also indicate whether a simple, stable class structure will indeed be achieved, and what that structure will be. It can also determine the importance of each of the different components of the matrix in generating the overall outcome, *R* – a topic to which we turn shortly. By convention, *R* in population projection matrices and related approaches (see below) is often referred to as λ. Here, for continuity with previous sections, we will continue to refer to the net reproductive rate as *R*.

integral projection models

Before we do so, however, we should acknowledge that the differences between individuals in a population, and hence the differences in their contribution to *R*, are not always best described in terms of age classes or stages. Often, demographic forces – birth rates, death rates and do on – vary with a continuous character, of which organism size is the most obvious example. In such cases, not a matrix model but an *integral projection model* (IPM) is appropriate. And where there is a mix of continuous and discrete characters, so-called *generalised* IPMs can be used (see Rees & Ellner (2009), Merow *et* al. (2014) and Rees *et al.* (2014) for accessible guides). An example of an IPM in action is shown in Figure 4.17, based on data from the females in a population of Soay sheep (*Ovis aries*) that have been intensively and extensively studied on the island of Hirta in the St Kilda archipelago in Scotland (Coulson, 2012). Here, rather than variations in survival, birth and so on being represented by successive elements in a projection matrix, field data are used to derive relationships between body weight and growth over the following year, survival to the next year, the production of offspring over the next year, and the size of those offspring (Figure 4.17a–d, respectively). As with the matrices, it is not necessary here to go into technical (mathematical) details, but it is easy to understand that fitting statistical models to the data in Figure 4.17 allows us, for example, to predict the probability of survival over the next year of a female, given her weight, in the same way as a value in a projection matrix allows us to predict the survival of that individual, given its age class. Doing this for each of the relationships allows us in turn to estimate R_O, the mean lifetime reproductive success (Equation 4.4) and especially R, the fundamental net reproductive rate (Equations 4.11–4.13). This, then, allows us to assess the current viability of the population, overall. (In this case, R was around 1.3 and hence the population was projected to increase from its current size.) It allows us, too, to predict the stable stage (here, size) distribution, which in this case has two peaks – one for lambs, and one for older individuals that have survived to adulthood but slowed or stopped in their rate of growth.

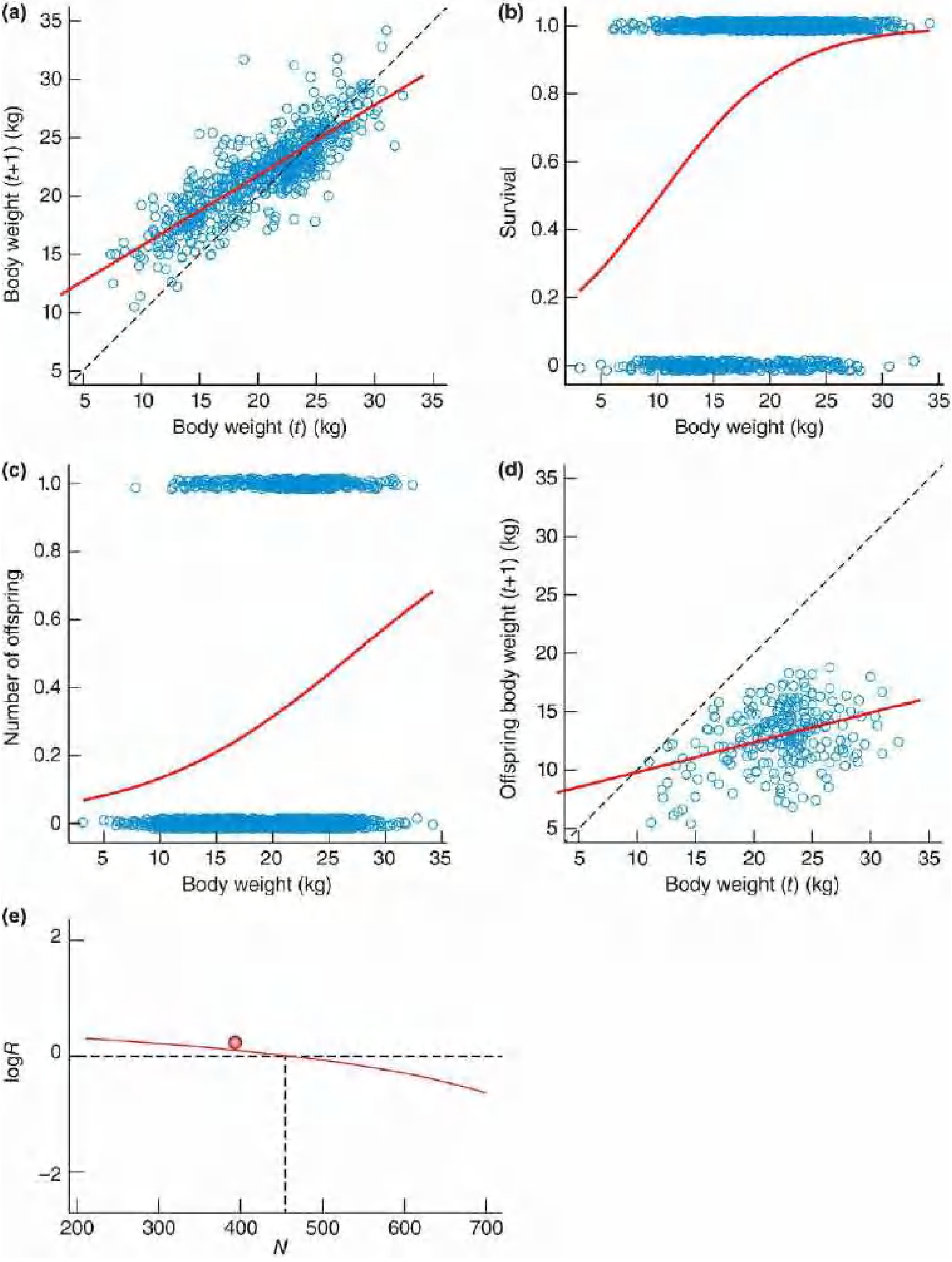

(a)
Body weight (t+1) (kg)
35
30
25
20
15
10
5
5
10
15
20
25
30
35
Body weight (t) (kg)
(b)
Survival
1.0
0.8
0.6
0.4
0.2
0
5
10
15
20
25
30
35
Body weight (kg)
(c)
Number of offspring
1.0
0.8
0.6
0.4
0.2
0
5
10
15
20
25
30
35
Body weight (kg)
(d)
Offspring body weight (t+1) (kg)
35
30
25
20
15
10
5
5
10
15
20
25
30
35
Body weight (t) (kg)
(e)
logR
2
0
−2
200
300
400
500
600
700
N

Figure 4.17 Elements and outcome of an integral projection model (IPM) for female Soay sheep (*Ovis aries*) on Hirta, St Kilda, Scotland. For details of the functions fitted to the data in each case, see the original text. (a) Growth rate: the relationship between body weight in successive years. (b) The effect of body weight on survival (whether or not a sheep is known to be alive). (c) The effect of body weight on offspring production. (Individuals either did or did not produce a lamb – twinning was rare and was ignored.) (d) The relationship between adult body weight and offspring body weight one year later. (e) The outcome, in terms of *R*, of the IPM that included the relationships in (a) to (d) – the red dot. Note that the IPM additionally included the effect of population size, *N*, on these processes, in which case *R* itself varies with *N*, declining as *N* increases (red line), equalling 1 ($\log R = 0$) when $N = 455$; see Section 5.4.2.

Source: After Coulson (2012).

What is more, the IPM approach, by projecting the future state of a population through the use of equations, is amenable to the inclusion in those equations of further factors that may also vary and affect survival, reproduction or growth. Probably the most important of these factors is the size of the population itself. The whole of Chapter 5 is devoted to intraspecific competition: the effects on individuals of being deprived of resources as a result of high local abundance and the consequent effects on populations. But even at this stage, without going into details, it makes sense to include a tendency for growth, survival and reproduction to decline as population abundance increases and resources become scarce. If these effects are incorporated into the IPM, we can see how this translates into the estimated net reproductive rate itself not being a fixed feature of the population, but declining with density (Figure 4.16e). We will refer back to this figure when we discuss intraspecific competition in detail in Chapter 5.

4.8.2 Life table response experiments

As we have noted, the overall value of *R*, calculated from a population projection matrix (or integral projection model), reflects the values of the various elements in that matrix, but their contribution to *R* is not equal. We are often interested in these relative contributions, because, for example, we may wish to increase the abundance of a threatened species (ensure *R* is as high as possible) or decrease the abundance of a pest (ensure *R* is as low as possible) and wish to know, therefore, which phases in the life cycle are the most important, since it is there that should be the focus of our efforts. In fact, there have been two distinct, though related approaches to this decomposition of *R*.

The term *life table response experiment* (LTRE) was initially used to describe studies in which the varied effects of a factor, for example a pollutant, on growth, survival and reproduction were combined to generate a meaningful overall response – the effect on *R* (Caswell, 1989). The key here is that the pollutant exerts its directly measurable effects on the growth, survival and reproduction of individuals, but we may be most interested in the overall, combined effect at the population level, that is, *R*. Subsequently, and now much more commonly, the term *LTRE analysis* has been used to describe retrospective analysis of populations subjected to different levels of a factor, with a view to determining the respective contributions of growth, survival and reproduction to overall differences in *R*. The contrast is between combination in the first case and decomposition in the second.

APPLICATION 4.3 Customised conservation of northern wheatears

Species threatened with extinction often persist as a series of small, fragmented populations. Conservation programmes must then tread a fine line between being focused on the particular needs of specific threatened fragments and retaining a commonality of approach that makes the programme as a whole affordable and practicable. With these ideas in mind, three small populations of the northern wheatear, *Oenanthe oenanthe*, one of the most rapidly declining breeding birds in Europe, were studied in the Netherlands (Van Oosten *et al*., 2015), where numbers have dropped by at least 80% since 1990. Of the three populations, numbers in the 1990s (females holding territories) increased from five to 30 at Aekingerzand (A) but decreased from 165 to 34 at Castricum (C) and fluctuated between 45 and 69 at Den Helder (D). Field data were therefore collected between 2007 and 2011 both to estimate the sizes of the breeding populations (Figure 4.18a) and to monitor the demographic processes (vital rates) determining the population sizes: fecundity, juvenile (first year) and adult survival, and in this case immigration.

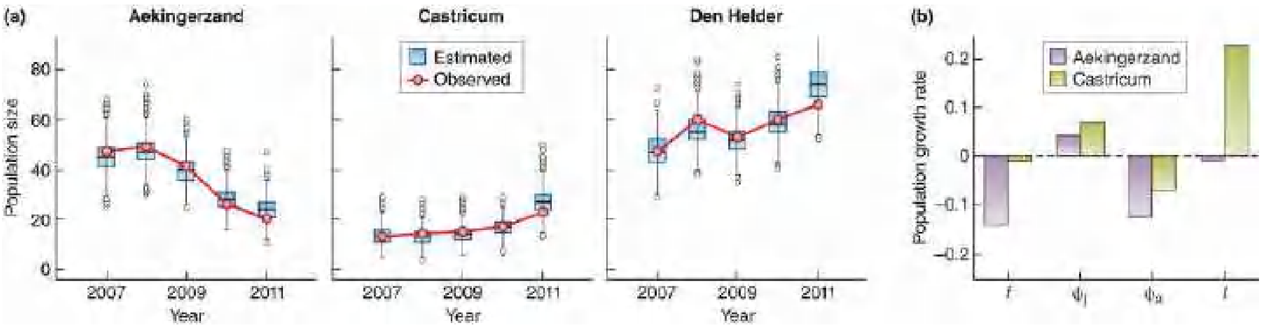


Figure 4.18 Analyses of life table response experiments (LTREs) can guide customised conservation. (a) Population sizes of northern wheatears, *Oenanthe oenanthe* (numbers of breeding territorial females) 2007–11 at three study sites in the Netherlands, as observed and as estimated from a matrix model integrating data on survival and fecundity. The variation in the estimates (box-and-whisker plots) arises because those estimates are taken regularly from chains generated by Monte Carlo Markov Chain (MCMC) procedures. The IPM does a good job at capturing the dynamics of the populations. (b) Contributions of the four vital rates (fecundity, *f*, juvenile and adult survival, ϕ_j and ϕ_a, and immigration, *I*) to the population growth rates observed in the Aekingerzand (purple) and Castricum (green) populations compared with that at Den Helder.

Source: After Van Oosten *et al*. (2015).

A matrix model integrating data on survival and fecundity was fitted to these demographic data and proved successful in recreating the observed population dynamics (Figure 4.18a), which in turn provides us with confidence that the IPM can usefully be used in an LTRE analysis. The results of this are shown in Figure 4.18b, comparing populations A and C to population D, as a benchmark, in terms of the contributions of the differences in vital rates to the overall differences in *R*. The relatively high rate of increase at site C (albeit from a low base) appears to be driven largely by immigration with some contribution from juvenile survival. The clear decrease at site A (*R* negative) resulted from low fecundity and adult survival.

The LTRE analysis, therefore, argues, most fundamentally, for customised approaches to be taken to local populations, since the forces driving their dynamics can be so different. Indeed, Van Oosten *et al*. (2015) conclude their study by explaining how they have responded to the analysis. To counteract the problems at site A, they introduced wire mesh covers to protect the nests and nesting adults, which led to an immediate more-than-doubling of fecundity and no predation by foxes, previously the main culprits. To support site C, so reliant on immigration, their strategy has been to protect site D, the main source of those immigrants. Fortunately, site D itself supports the most stable population of the three, and the strategy therefore is simply to protect the habitat there for the sake of that population and of others, like C, that rely on it.

4.8.3 Sensitivity and elasticity analysis

sensitivities and elasticities

In contrast to the retrospective LTRE analysis of the data contained in population projection matrices or IPMs, it is also possible to carry out *pros*pective *sensitivity* or *elasticity analyses* (Caswell, 2001). Without going into the algebraic details, the general principle is one of 'perturbing' the values of elements, or combinations of elements, in the matrix, and then noting the effects of those perturbations on aggregate properties such as *R*. The *sensitivity* of each element (i.e. each transition, birth or survival, in the overall life cycle) is the amount by which *R* would change for a given absolute change in the value of the matrix element, with the value of all the other elements held constant. Thus, sensitivities are highest for those processes that have the greatest power to influence *R*. However, whereas survival elements (*g*s and *p*s) are constrained to lie between 0 and 1, fecundities are not, and *R* therefore tends to be more sensitive to absolute changes in survival than to absolute changes of the same magnitude in fecundity. Moreover, *R* can be sensitive to an element in the matrix even if that element takes the value 0 (because sensitivities measure what *would* happen if there *was* an absolute change in its value). These shortcomings are overcome, though, by using the *elasticity* of each element to determine its contribution to *R*, since this measures the proportional change in *R* resulting from a proportional change in that element. Conveniently, too, with the matrix formulation, the elasticities sum to 1, so that the meaning of an elasticity of 0.5, say, is clear: the element concerned accounts for half the variation in *R*.

APPLICATION 4.4 Elasticity analysis and population management

elasticity analysis and the management of armadillo abundance

Elasticity analysis offers an especially direct route towards focused plans for the management of abundance. For example, an elasticity analysis has been applied to the population dynamics of nine-banded armadillos, *Dasypus novemcinctus*, in Mississippi, USA. The armadillo is a reservoir of infection of the causative agent of leprosy, *Mycobacterium leprae*, and indeed the only known non-human vertebrate host of the disease in the Americas (see Oli *et al.* (2017) who also carry out an elasticity analysis of leprosy dynamics within the armadillo populations). Globally, between 200 000 and 300 000 new human cases of leprosy are reported annually. Given the popular view of leprosy as a disease of a bygone era, a surprising number of these cases (around 200) are in southern USA, and these are increasingly being linked to infected armadillos. An understanding of the forces driving the population dynamics of the armadillos is important, therefore, because the risk of human infection increases with the abundance of infected armadillos, and hence with the abundance of the armadillos themselves. (The ecology of these 'zoonotic' infections, passed from wildlife to humans, is discussed in more detail in Section 12.3.2.) A life cycle graph and associated population projection matrix for armadillos is shown in Figure 4.19a. Three age classes are distinguished: juveniles (0–1 years old, prereproductive), yearlings (1–2 years) and adults (>2 years), though these adults may also transition into an infected (leprous) state that can also reproduce. Estimates for the various elements of the matrix, from field data, are shown in Figure 4.19b. The reproductive rates describe additions to the free-living juvenile class, since these, rather than newborns, are the youngest animals that can be trapped. However, the survival rate from birth to becoming trappable is unknown. The matrix model was therefore run for low, medium and high values for this survival rate, γ (0.5, 0.8 and 1.0). The elasticities of the various elements of the matrix are shown in Figure 4.19c.

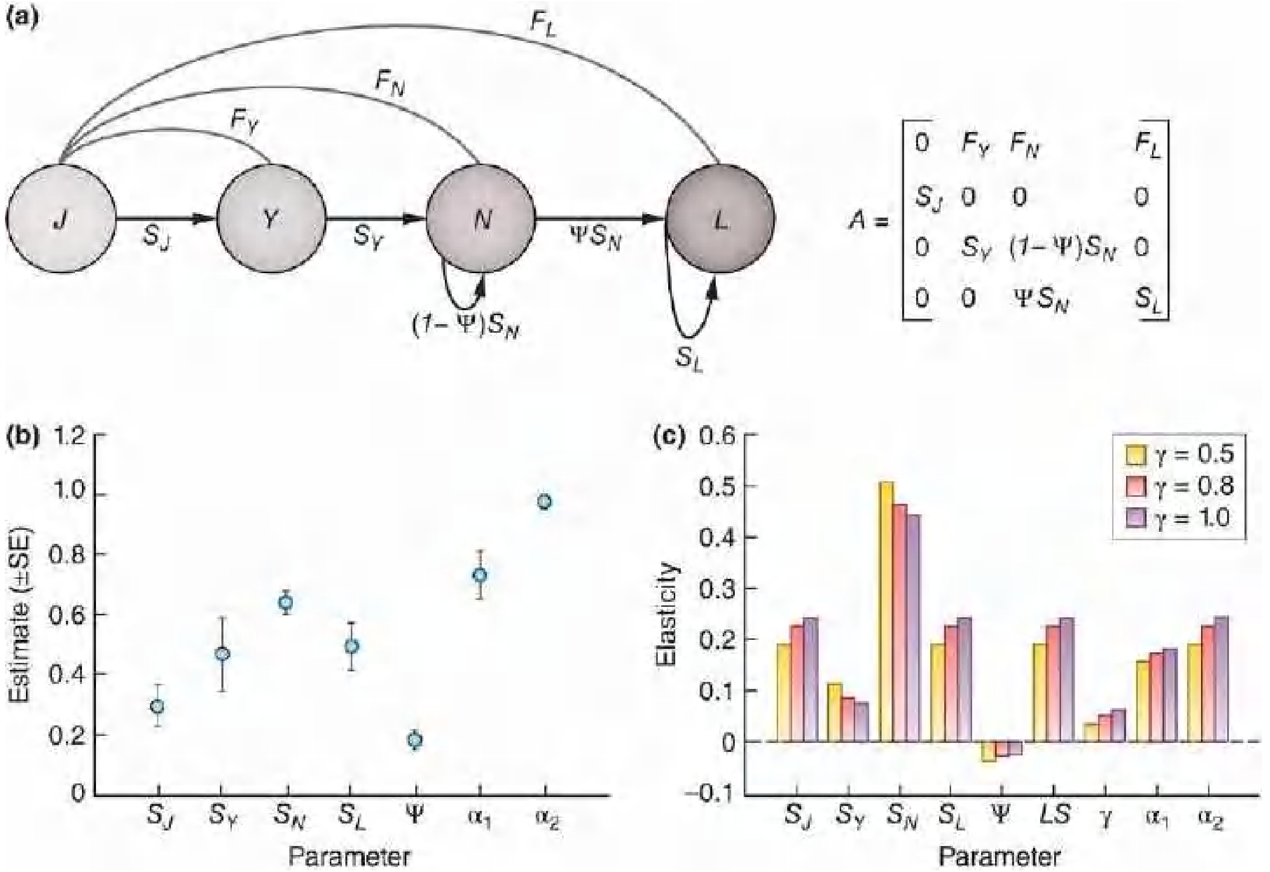


Figure 4.19 Elasticity analyses can guide the management of armadillo abundance. (a) Life cycle graph and population projection matrix for nine-banded armadillos, *Dasypus novemcinctus*, in Mississippi, USA, comprising fecundities, *F*, and survival rates, *S*, for juveniles, *J*, yearlings, *Y*, non-leprous adults, *N*, and leprous adults, *L*, and with ψ referring to the probability of non-leprous adults becoming leprous. (b) Estimates, with standard errors, of these parameters from field data, except that α_1 (yearlings) and α_2 (adults) are probabilities of reproduction that are combined with litter sizes to generate the fecundities. (c) The elasticities of the population growth rate, *R*, to these parameters, to litter size (*LS*) and to the (unknown) probability of surviving to a trappable age, γ, for three values of γ (0.5, 0.8, 1.0).

Source: After Oli *et al.* (2017).

Of those elasticities, it is encouraging, first, that the elasticity for the unknown survival rate, γ, is low, indicating that our conclusions are not strongly dependent on our assumptions about γ. Next, it is apparent from Figure 4.19b that infected adults had a reduced survival rate (down 14.5%), and it is for this reason that the elasticity values for the probability of transition of adults into the infected state were negative (Figure 4.19c). However, these elasticities were especially low, indicating that *R* for the armadillo population would not be greatly affected by the infection rate. Rather, the parameter with an elasticity value indicating the greatest influence on *R* (approaching 0.5) was the survival rate of adults.

The distribution of nine-banded armadillos is expanding northwards in the USA, and the incidence of leprosy in these populations is increasing drastically. The elasticity analysis suggests that leprosy itself will do little to halt the spread of armadillos. If their abundance is going to be controlled, adult survival is likely to be the most effective, as well as perhaps the most practical target.

elasticity analysis and thistle control

Elasticity analysis has been applied, too, to populations of the nodding thistle (*Carduus nutans*), a noxious weed that is prickly and unpalatable to most livestock, and that has expanded from its Eurasian origins to invade many parts of the world, including Australia and New Zealand. The question at issue in this case is why control measures for the thistle that are effective in one part of the world are not always effective elsewhere. The life cycle graph for the thistle has the same structure in the two countries (Figure 4.19), comprising four stages: a seed bank, and small, medium and large plants. In fact, ‘size’ is defined not literally but on the basis of their probability of flowering: <20%, 20–80%, and >80%, respectively, since the size–flowering relationship itself varies between the countries. Field data from sites in each country, summarised in the projection matrices in Figure 4.20, indicate that the detailed demography was also rather different in the two cases. In Australia, fecundity was relatively low compared with New Zealand, as indicated by transitions in the matrix from small, medium and large plants either into the seed bank, or directly into small plants, following germination (highlighted in bold in the matrices). Germination from the seed bank to small plants in Australia was also relatively low. On the other hand, the probabilities of surviving within a size-class and of surviving and growing into the next size-class were noticeably higher in Australia. This translates into values of $R = 1.2$ for the high survival, low fecundity Australian population, and $R = 2.2$ for the high fecundity, low survival New Zealand population, although despite this difference, the species is a highly invasive weed in both countries.

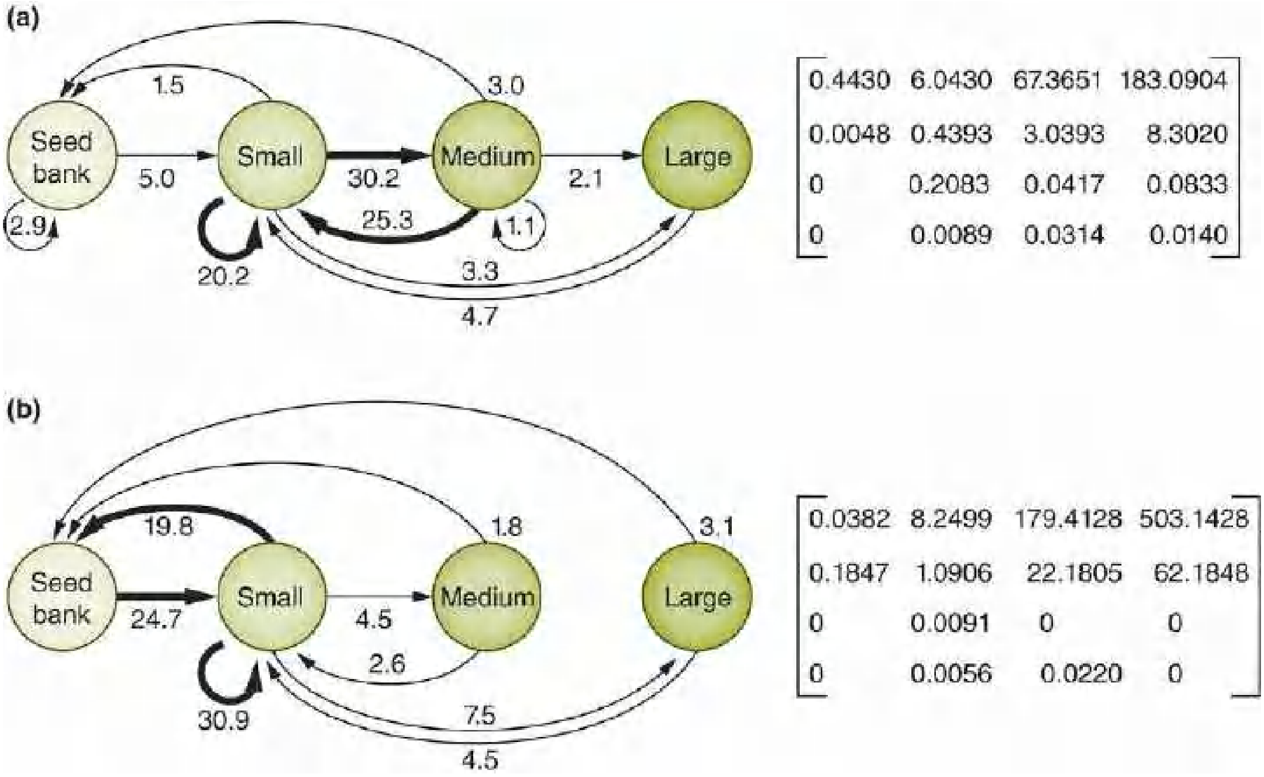


Figure 4.20 **Elasticity analysis can guide the management of thistle abundance.** (a) Life cycle graph and population projection matrix for the nodding thistle, *Carduus nutans*, in Australia, comprising a seed bank and small, medium and large plants. (b) The equivalent for a population in New Zealand. The arrows in the life cycle graphs are the transitions from year to year (survival, fecundity, growth and (for the seed bank) dormancy) and the numbers associated with them are the elasticities of R to these transitions, expressed as percentages of total elasticity. Dominant elasticities (≥20%) are indicated with bold arrows.

Source: After Shea *et al.* (2005).

These differences in demography led in turn to differences in the elasticities in the two cases (Figure 4.20). For the Australian population, the dominant transitions were the cycle from small and medium plants back to small plants via seed production and germination, the survival of small plants (hence two contributions to the bold arrow in Figure 4.20a from small back to small plants), and the growth of small into medium plants. For the New Zealand population, the dominant transitions again included the production of small plants by small plants via germinated seed, but also the addition of seeds to the seed bank by small plants and the germination of those seeds.

The vulnerabilities of the two populations to control measures are therefore also different. The main options are three species of insect: two beetles and a fly (biological control of pests and weeds is discussed fully in Chapter 15). First, the thistle receptacle weevil, *Rhinocyllus conicus*, reduces seed set of thistles by around 30–35% in both Australia and New Zealand. Its release has been the most effective measure in controlling thistles in many parts of the world, but not, seemingly, in New Zealand. Second, the receptacle gallfly, *Urophora solstitialis*, reduces seed production by around 70% in Australia, though no estimates are available for New Zealand. And last, the root-crown weevil, *Trichosirocalus horridus*, reduces plant growth by around 87% and thus affects both survival and fecundity. It is this species that has seemed to be most effective in Australia, and the elasticity analysis is entirely consistent with this, since growth and survivorship are relatively important there, compared with reproduction.

In New Zealand, by contrast, targeting seed production would seem from the elasticity analysis to be the most appropriate strategy, and the lack of success of *R. conicus* there is therefore likely to be a result not of inappropriate targeting but of the ineffectiveness of *R. conicus* (Shea & Kelly, 1998). This in turn suggests that the subsequent release of the gallfly should have been more successful, but sadly this seems not to have been the case either, perhaps because young gallfly larvae are themselves preyed upon by *R. conicus* larvae (Groenteman *et al.*, 2011). As these authors remark, ‘Thirty-five years into the biocontrol programme and three agents later, *C. nutans* is still a major weed in parts of New Zealand’. Elasticity analyses of population projection matrices can direct managers to a pest’s vulnerabilities, but they cannot conjure up effective biocontrol agents.

Chapter 5
Intraspecific Competition

5.1 Introduction

Organisms grow, reproduce and die (Chapter 4). They are affected by the conditions in which they live (Chapter 2) and by the resources that they obtain (Chapter 3). But no organism lives alone. Each, for at least part of its life, is part of a population of its own species.

a definition of competition

Competition can be defined as an interaction between individuals brought about by a shared requirement for a resource, leading to a reduction in the survivorship, growth and/or reproduction of at least some of the individuals concerned. Individuals of the same species have very similar requirements for survival, growth and reproduction, but their combined demand for a resource may exceed the immediate supply. The individuals then compete for the resource and at least some of them will become deprived. In this chapter we examine the nature of such intraspecific competition, its effects on the competing individuals and on populations of competing individuals.

Consider, initially, a simple hypothetical community: a thriving population of grasshoppers (all of one species) feeding on a field of grass (also of one species). To provide themselves with energy and material for growth and reproduction, grasshoppers must find and eat grass, but use energy in doing so. A grasshopper that finds itself at a spot where there is no grass because another grasshopper has eaten it must move on and expend more energy before it takes in food. The more grasshoppers there are, the more this will happen, increasing energy expenditure, decreasing the rate of food intake, and hence potentially decreasing its chances of survival and leaving it less energy for development and reproduction. Survival and reproduction determine a grasshopper's contribution to the next generation. Hence, the more intraspecific competitors for food a grasshopper experiences, the less its likely contribution will be.

As far as the grass itself is concerned, an isolated seedling in fertile soil may have a very high chance of surviving to reproductive maturity. It will probably exhibit extensive modular growth and probably, therefore, eventually produce many seeds. However, a seedling that is closely surrounded by neighbours (shading it with their leaves and depleting the water and nutrients of its soil with their roots) will be very unlikely to survive, and if it does, will almost certainly form few modules and set few seeds.

We see immediately that the ultimate effect of competition on an individual is a decreased contribution to the next generation compared with what would have happened had there been no competitors. Intraspecific competition typically leads to decreased rates of resource intake per individual, and thus to decreased rates of individual growth or development, or perhaps to decreases in the amounts of stored reserves or to increased risks of predation. These may lead, in turn, to decreases in survivorship and/or decreases in fecundity, which together determine an individual's reproductive output.

5.1.1 Exploitation and interference

exploitation

In many cases, competing individuals do not interact with one another directly. Rather, they deplete the resources that are available to each other. Grasshoppers may compete for food, but a grasshopper is not directly affected by other grasshoppers, but rather by the level to which they have reduced the food supply. Likewise, two grass plants may compete, and each may be adversely affected by the presence of close neighbours, but this is most likely to be because their *resource depletion zones* overlap – each may shade its neighbours from the incoming flow of radiation, and water or nutrients may be less accessible around the plants' roots than they would otherwise be. The data in Figure 5.1, for example, show the dynamics of the interaction between single-celled algal species, diatoms, and one of the resources they require, silicate. As diatom density increases over time, silicate concentration decreases until both reach a steady state in which there is less resource available for the many than there had been previously for the few. This type of competition – in which competitors interact only indirectly, through their shared resources – is termed *exploitation*.

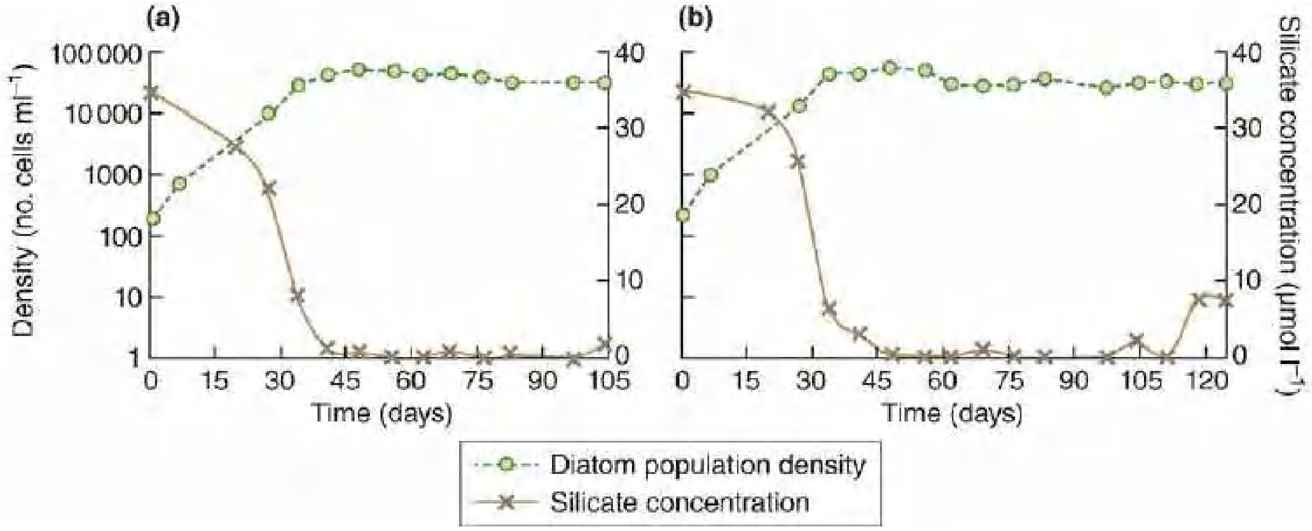


Figure 5.1 In exploitation competition, resource levels decline as population density increases. Dynamics over time of populations of the freshwater diatoms (a) *Cyclotella pseudostelligera* and (b) *Fragilaria crotonensis* and of the silicate which is one of their key resources. The diatoms consume silicate during growth and the populations of diatoms stabilise when the silicate has been reduced to a very low concentration.

Source: After Descamps-Julien & Gonzalez (2005).

interference

In other cases, competition takes the form of interference. Here individuals interact directly with each other, and one individual will actually prevent another from exploiting the resources within a portion of the habitat. For instance, this is seen amongst animals that defend territories (see

Section 5.8.4) and amongst the sessile animals and plants that live on rocky shores. The presence of a barnacle on a rock prevents any other barnacle from occupying that same position, even though the supply of food at that position may exceed the requirements of several barnacles. In such cases, space can be seen as a resource in limited supply. Another type of interference competition occurs when, for instance, two red deer stags fight for access to a harem of hinds. Either stag, alone, could readily mate with all the hinds, but they cannot both do so since matings are limited to the 'owner' of the harem. Thus, with exploitation, the intensity of competition is closely linked to the level of resource present and the level required, but with interference, intensity may be high even when the level of the real resource is not limiting.

In practice, many examples of competition probably include elements of both exploitation and interference. For instance, adult cave beetles, *Neapheanops tellkampfi*, in Great Onyx Cave, Kentucky, USA compete amongst themselves but with no other species and have only one type of food – cricket eggs, which they obtain by digging holes in the sandy floor of the cave. On the one hand, they suffer indirectly from exploitation: beetles reduce the density of their resource (cricket eggs) and then have markedly lower fecundity when food availability is low (Figure 5.2a). But they also suffer directly from interference: at higher beetle densities they fight more, forage less, dig fewer and shallower holes and eat far fewer eggs than could be accounted for by food depletion alone (Figure 5.2b).

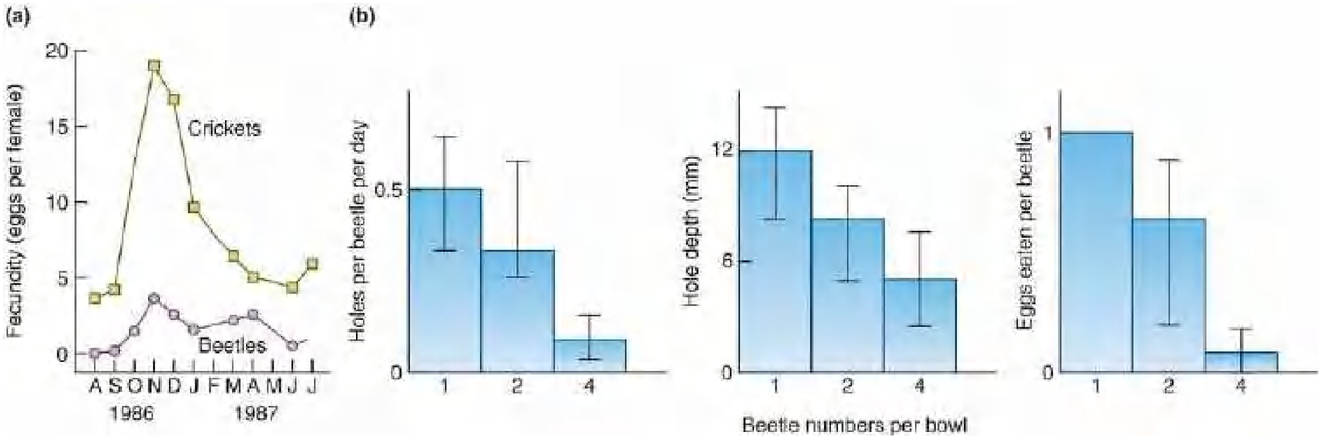


Figure 5.2 Competition may combine elements of both exploitation and interference. Intraspecific competition amongst cave beetles (*Neapheanops tellkampfi*). (a) Exploitation. Beetle fecundity is significantly correlated ($r = 0.86$) with cricket fecundity (itself a good measure of the availability of cricket eggs – the beetles' food). The beetles themselves reduce the density of cricket eggs. (b) Interference. As beetle density in experimental arenas with 10 cricket eggs increased from one to two to four, individual beetles dug fewer and shallower holes in search of their food, and ultimately ate much less ($P < 0.001$ in each case), in spite of the fact that 10 cricket eggs was sufficient to satiate them all. Means and standard deviations are given in each case.

Source: After Griffith & Poulson (1993).

5.2 Intraspecific competition, and density-dependent mortality, fecundity and growth

5.2.1 Density-dependent mortality and fecundity

under- and overcompensating density dependence

The likely effect of intraspecific competition on any individual is greater the more competitors there are. The effects of intraspecific competition are thus said to be *density dependent*. We see this in more detail in Figure 5.3a, which shows the pattern of mortality in the barnacle, *Semibalanus balanoides*, on a rocky shore in north Wales, UK, between their recruitment as settling larvae and their survival as established adults two years later. The same data have been expressed in two ways. From the upper panel it is clear that initially, as the number of recruits increased, this led to a corresponding increase in the number of surviving adults. But the number of survivors peaked at a recruit density of around 30 per cm^2 and thereafter declined. We can therefore say that at densities beyond this peak, the mortality rate showed *overcompensating density dependence* in that increases in initial numbers led to decreases in final numbers. By contrast, before the peak, the density dependence was *undercompensating* in that final numbers continued to rise as the number of recruits increased.

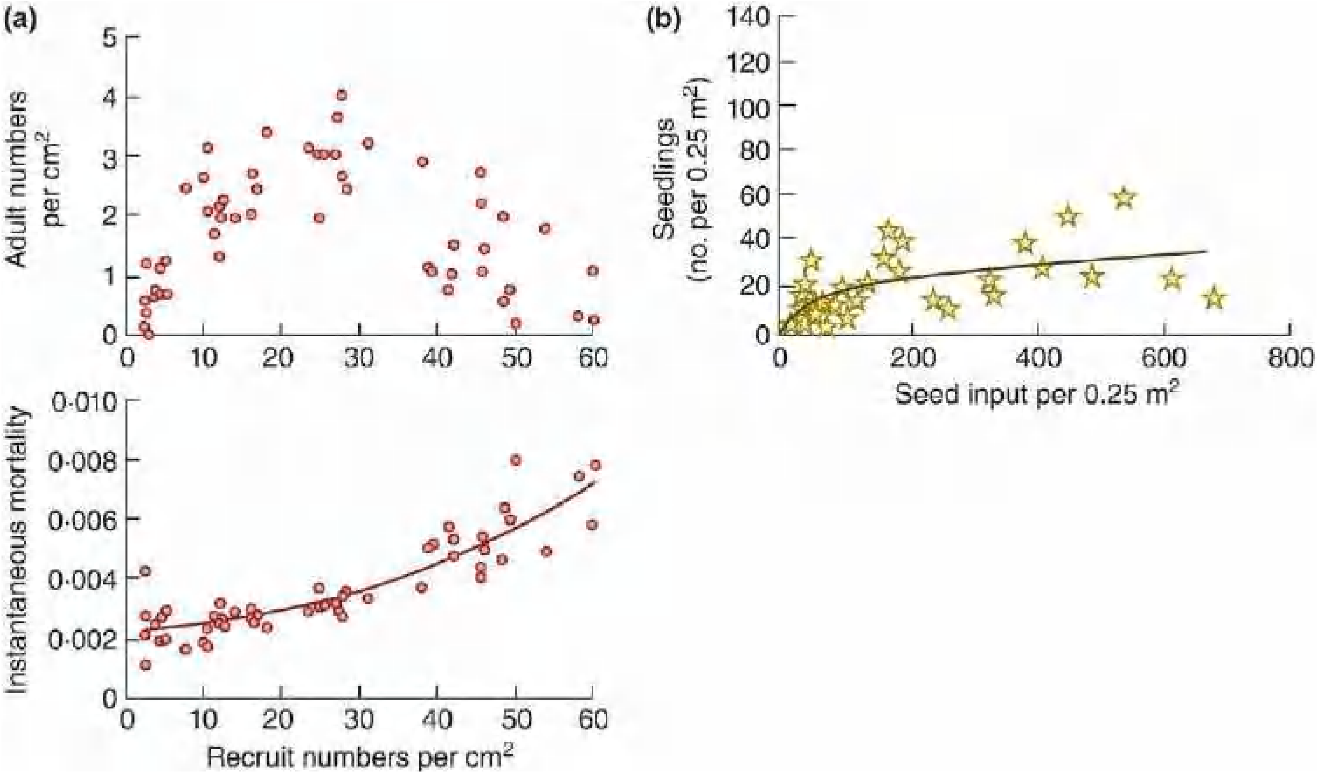


Figure 5.3 Density-dependent mortality. (a) Upper panel: the density of surviving adults of the barnacle, *Semibalanus balanoides*, in the UK as a function of the density of recruits two years earlier. Lower panel: the same data expressed as the relationship between the daily mortality rate and the density of recruits. (b) The density of surviving seedlings recruited to a population of the yellow star thistle, *Centaurea solstitialis*, in California as a function of the number of seeds present in the previous year.

Source: (a) After Jenkins *et al.* (2008). (b) After Swope & Parker (2010).

The lower panel plots the mortality rate against the initial number of recruits (mortality rate being calculated as $-ln(S/R)$, where S and R are the number of survivors and recruits, respectively, divided by 730 (two years) to give a daily rate – similar to the intrinsic rate of natural increase, r, calculated in Section 4.7.1). We can see that at the very lowest abundances of recruits, the relationship was flat. That is, the mortality rate stayed the same and was thus *density independent*. There was no evidence of intraspecific competition when initial abundances were low. As those abundances increased, the slope of the relationship became positive – there was density dependence and thus evidence of competition – and that slope became steeper as the density dependence moved from under- to overcompensation.

exactly compensating density dependence

A similar relationship is shown in Figure 5.3b, but this time for a plant, the yellow star thistle, *Centaurea solstitialis*, in California, USA, relating the density of seedlings to the initial number of seeds in the soil. This time, though, at the highest seed densities, the number of surviving seedlings levelled off. The density dependence was *exactly compensating*: as initial density increased the mortality rate rose to counteract it.

intraspecific competition and fecundity

The patterns of density-dependent fecundity that result from intraspecific competition are, in a sense, a mirror-image of those for mortality (Figure 5.4). Here, though, the per capita birth rate falls as intraspecific competition intensifies. At low enough densities, the birth rate may be density independent (Figure 5.4a, lower densities). But as density increases, and the effects of intraspecific competition become apparent, birth rate initially shows undercompensating density dependence (Figure 5.4a, higher densities), and may then show exactly compensating density dependence (Figure 5.4b, throughout; Figure 5.4c, lower densities) or overcompensating density dependence (Figure 5.4c, higher densities).

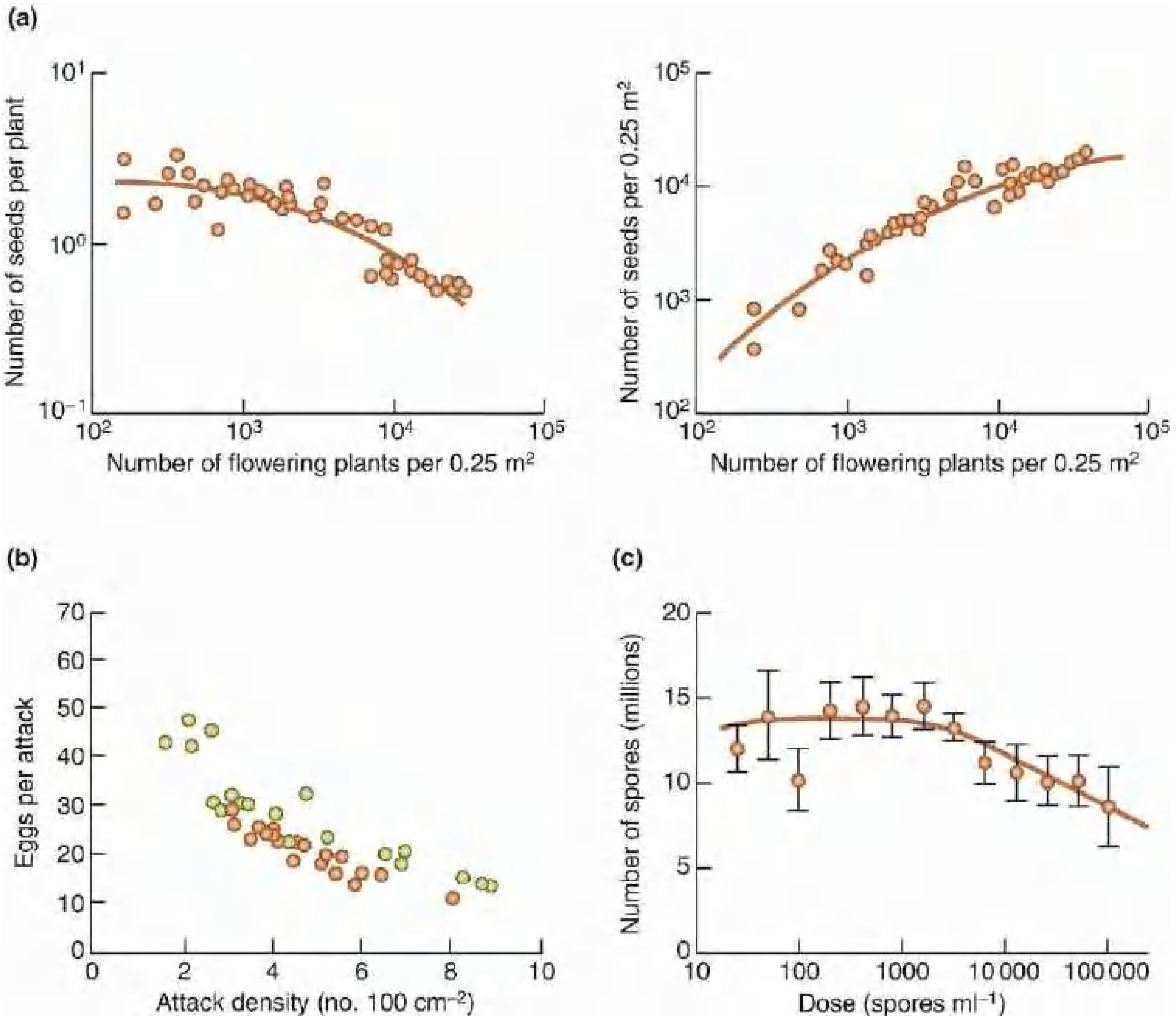


Figure 5.4 **Density-dependent fecundity.** (a) The fecundity (seeds per plant) of the annual dune plant *Vulpia fasciculata* is constant at the lowest densities (density independence, left). However, at higher densities, fecundity declines but in an undercompensating fashion, such that the total number of seeds continues to rise (right). (b) Fecundity in the southern pine beetle, *Dendroctonus frontalis*, in East Texas, USA (the number of eggs laid each time a beetle 'attacks' a tree) declines with increasing density of these attacks in a way that compensates more or less exactly for the density increases: the total number of eggs produced (eggs per attack × attack density) was roughly 100 per 100 cm^2, irrespective of attack density over the range observed (🟠, 1992; 🟢, 1993). (c) When the planktonic crustacean *Daphnia magna* was infected with varying numbers of spores of the bacterium *Pasteuria ramosa*, the total number of spores produced per host in the next generation was independent of density (exactly compensating) at the lower densities, but declined with increasing density (overcompensating) at the higher densities. Standard errors are shown.

Source: (a) After Watkinson & Harper (1978). (b) After Reeve *et al.* (1998). (c) After Ebert *et al.* (2000).

5.2.2 Intraspecific competition and density-dependent growth

Intraspecific competition can also have a profound effect on individual growth, which is often a forerunner of the effects on survival and fecundity that we have already seen: smaller individuals are less likely to survive and less fecund. We see an example of an effect on butterfly larvae in Figure 5.5.

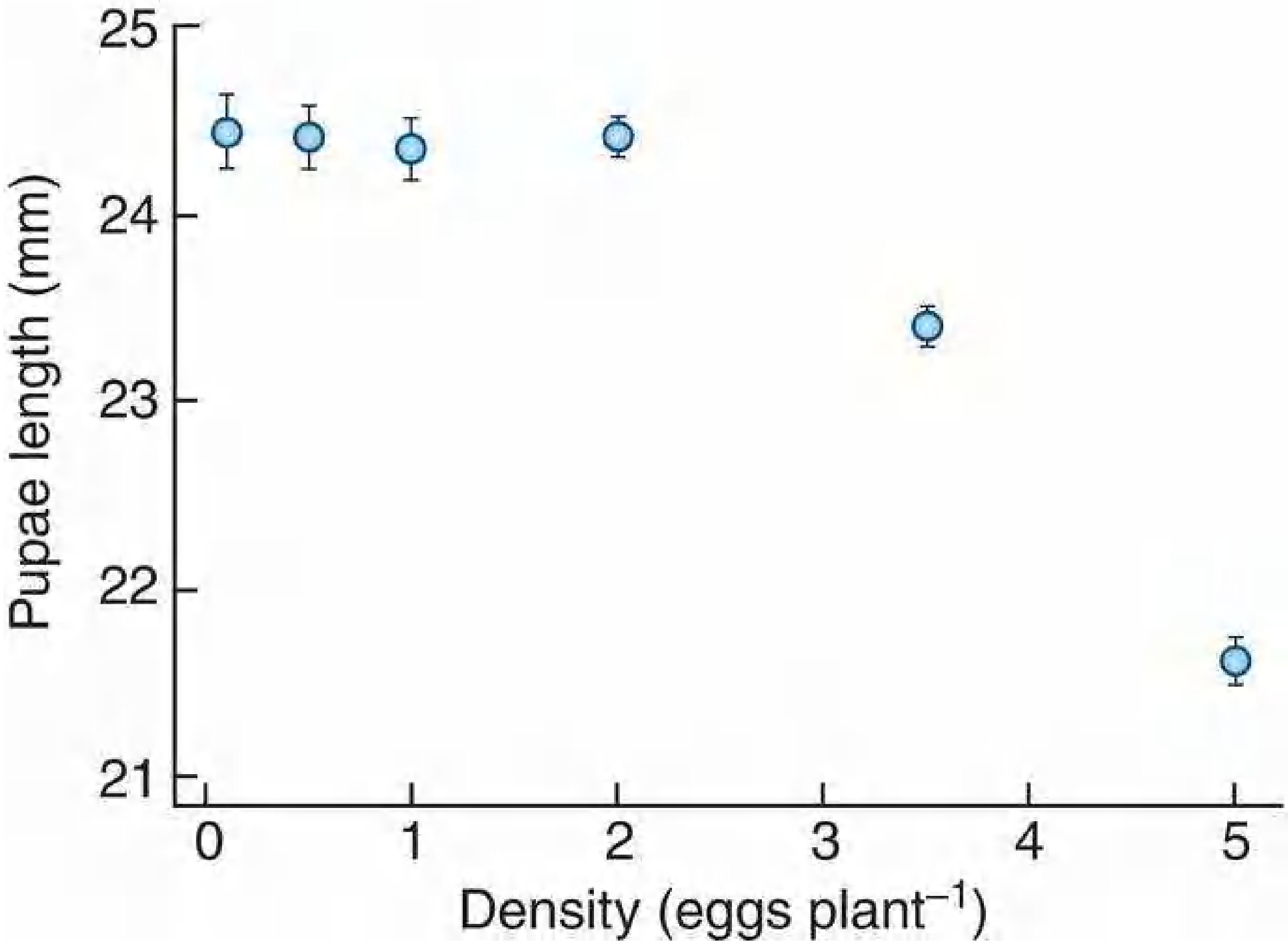
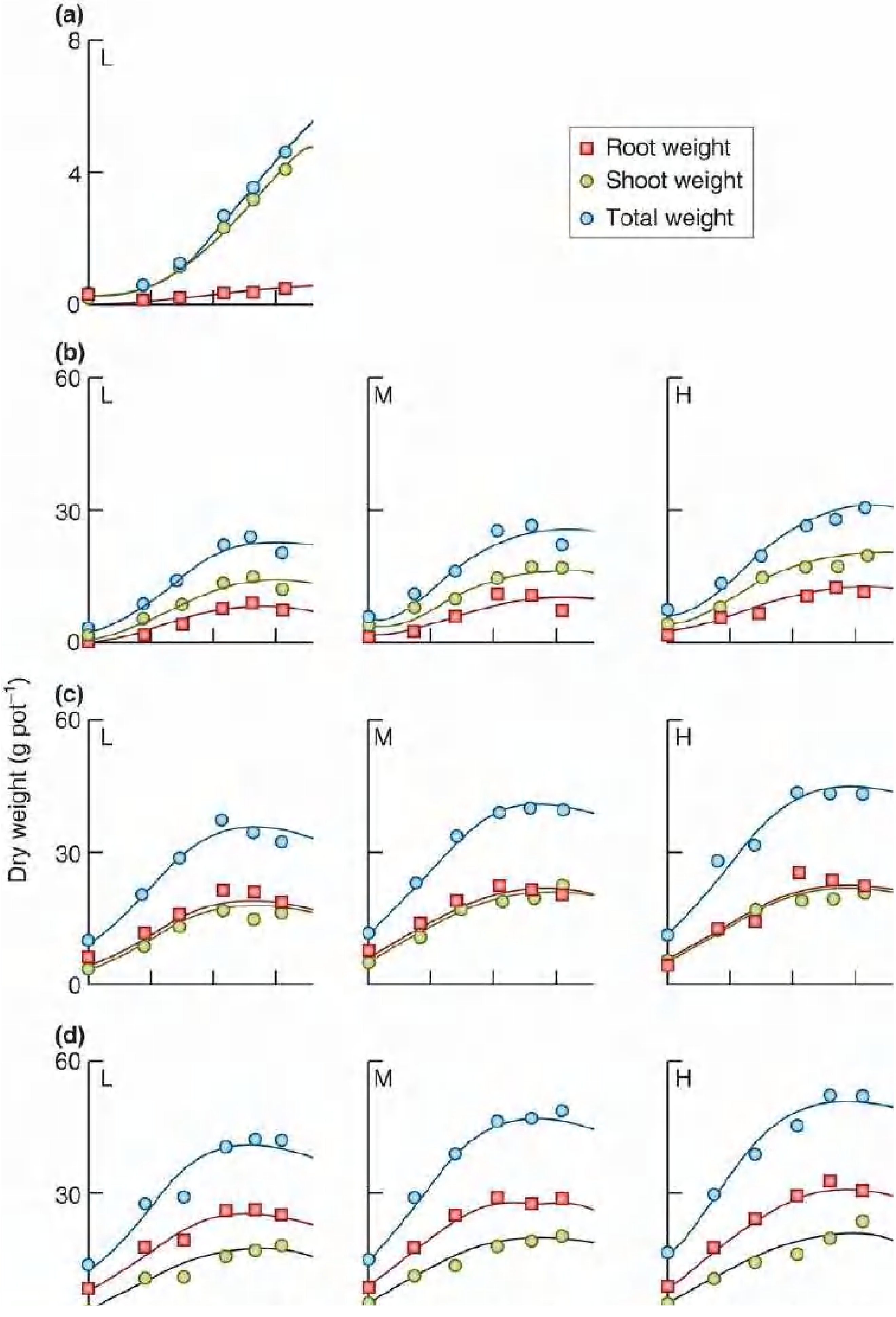


Figure 5.5 **Density-dependent growth.** The mean length of monarch butterfly pupae, *Danaus plexippus*, reared at different densities. Growth was density independent at the lowest densities but declined at densities above around two eggs per plant. Bars are SEs.

Source: After Flockhart *et al*. (2012).

constant final yield

Such effects are particularly marked in modular organisms. For example, when carrot seeds (*Daucus carota*) were sown at a range of densities, the yield per pot at the first harvest (29 days) increased with the density of seeds sown (Figure 5.6). After 62 days, however, and even more after 76 and 90 days, yield no longer reflected the numbers sown. Rather it was the same over a wide range of initial densities, especially at higher densities where competition was most intense. Plant ecologists have referred to this pattern as reflecting the 'law of constant final yield' (see Weiner and Freckleton (2010) for a review), though it is debatable whether ecology really has any universal 'laws' in the same way as, say, physics does. Individuals suffer density-dependent reductions in growth rate, and thus in individual plant size, which tend to compensate exactly for increases in density (hence the constant final yield). This suggests, of course, that there are limited resources available for plant growth, especially at high densities, which is borne out in Figure 5.6 by the higher (constant) yields at higher nutrient levels.

(a)
8
L
4
0
Root weight
Shoot weight
Total weight
(b)
60
L
M
H
30
0
Dry weight (g pot^{-1})
(c)
60
L
M
H
30
0
(d)
60
L
M
H
30

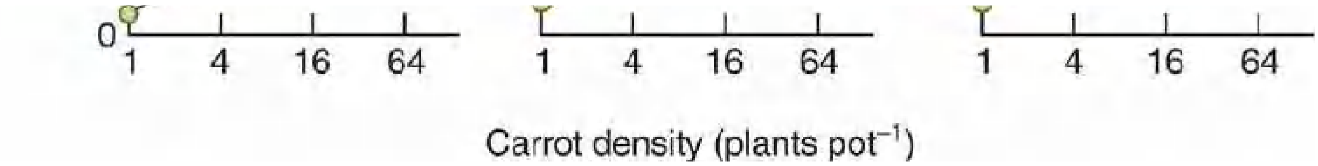


Figure 5.6 Plants sown at a range of densities often grow to achieve a constant final yield. The relationship between yield per pot and sowing density in carrots (*Daucus carota*) at four harvests ((a) 29 days after sowing, (b) 62 days, (c) 76 days, and (d) 90 days) and at three nutrient levels (low, medium and high: L, M and H, respectively), given to pots weekly after the first harvest. Points are means of three replicates, with the exception of the lowest density and the first harvest (both nine replicates). The curves were fitted in line with theoretical yield–density relationships, the details of which are unimportant in this context.

Source: After Li *et al*. (1996).

Yield is density (d) multiplied by mean weight per plant $(\overline{w})$. Thus, if yield is constant (c):

$$d\overline{w} = c, \tag{5.1}$$

and so:

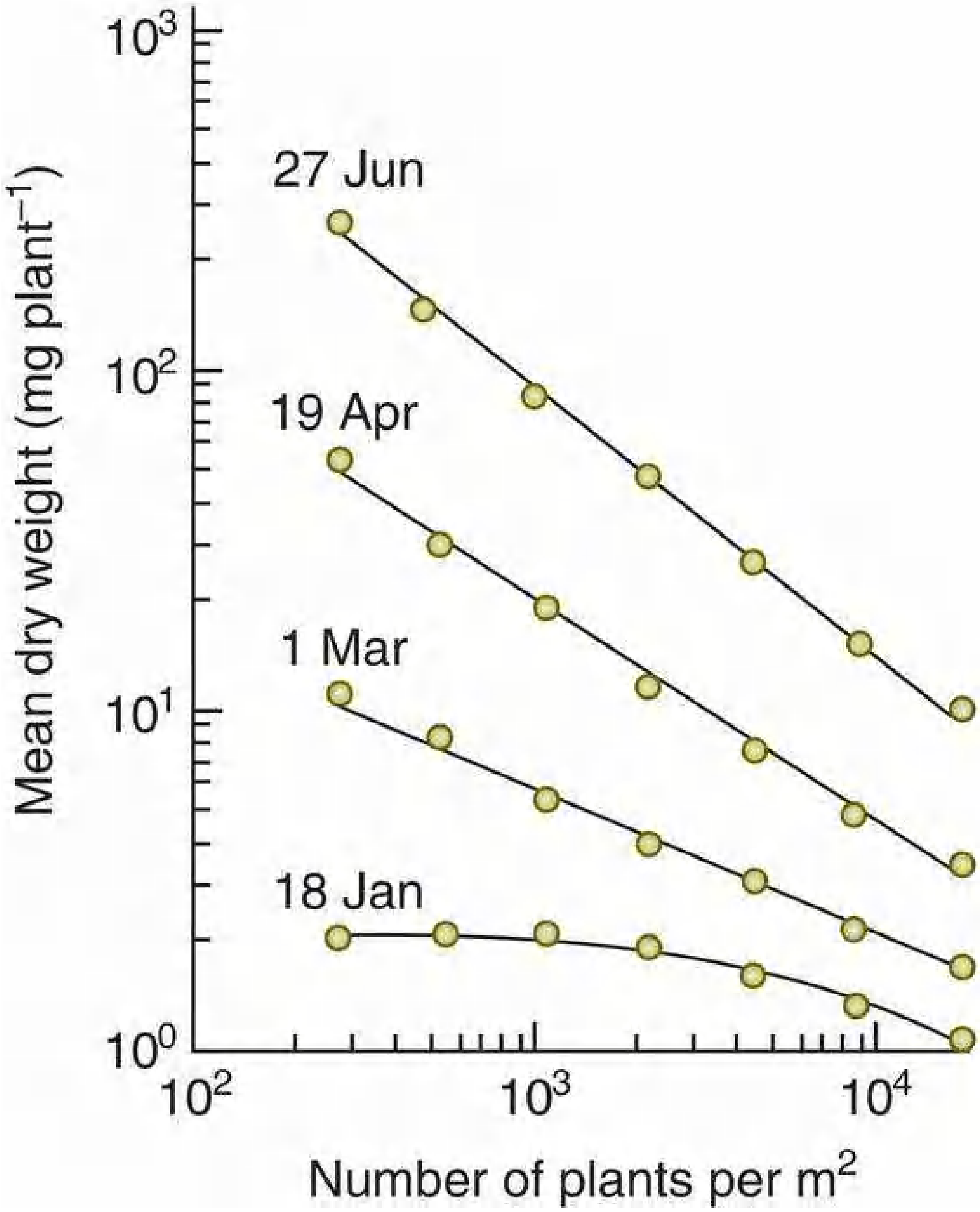
$$\log d + \log \overline{w} = \log c \tag{5.2}$$

and:

$$\log \overline{w} = \log c - 1 \cdot \log d. \tag{5.3}$$

Thus, a plot of log mean weight against log density should have a slope of –1.

Data on the effects of density on the growth of the grass *Vulpia fasciculata* are shown in Figure 5.7, and the slope of the curve towards the end of the experiment does indeed approach a value of –1. Here too, as with the carrot plants, individual plant weight at the first harvest was reduced only at very high densities – but as the plants became larger, they interfered with each other at successively lower densities.



[Figure 5.7](#) The 'constant final yield' of plants illustrated by a line of slope −1 when log mean weight is plotted against log density in the dune annual, *Vulpia fasciculata*. On 18 January, particularly at low densities, growth and hence mean dry weight were roughly independent of density. But by 27 June, density-dependent reductions in growth compensated exactly for variations in density, leading to a constant yield.

Source: After Watkinson ([1984](#)).

constant yield and modularity

The constancy of the final yield is a result, to a large extent, of the modularity of plants. This was clear when perennial rye grass (*Lolium perenne*) was sown at a 30-fold range of densities ([Figure 5.8](#)). After 180 days some genets had died, but the range of final tiller (module) densities was far narrower than that of genets (individuals). The constancy arose largely through effects on the number of modules per genet rather than the number of genets themselves.

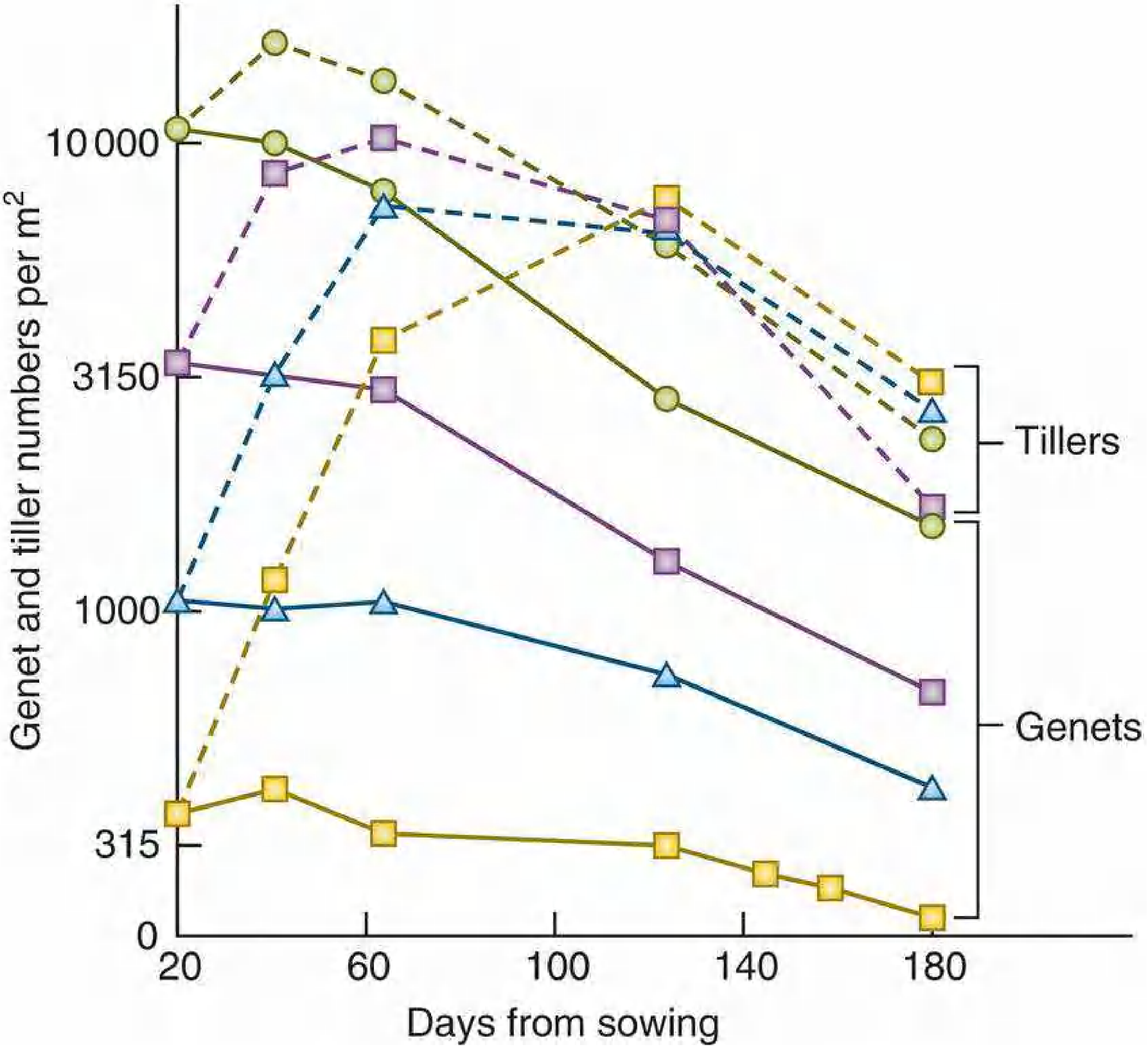
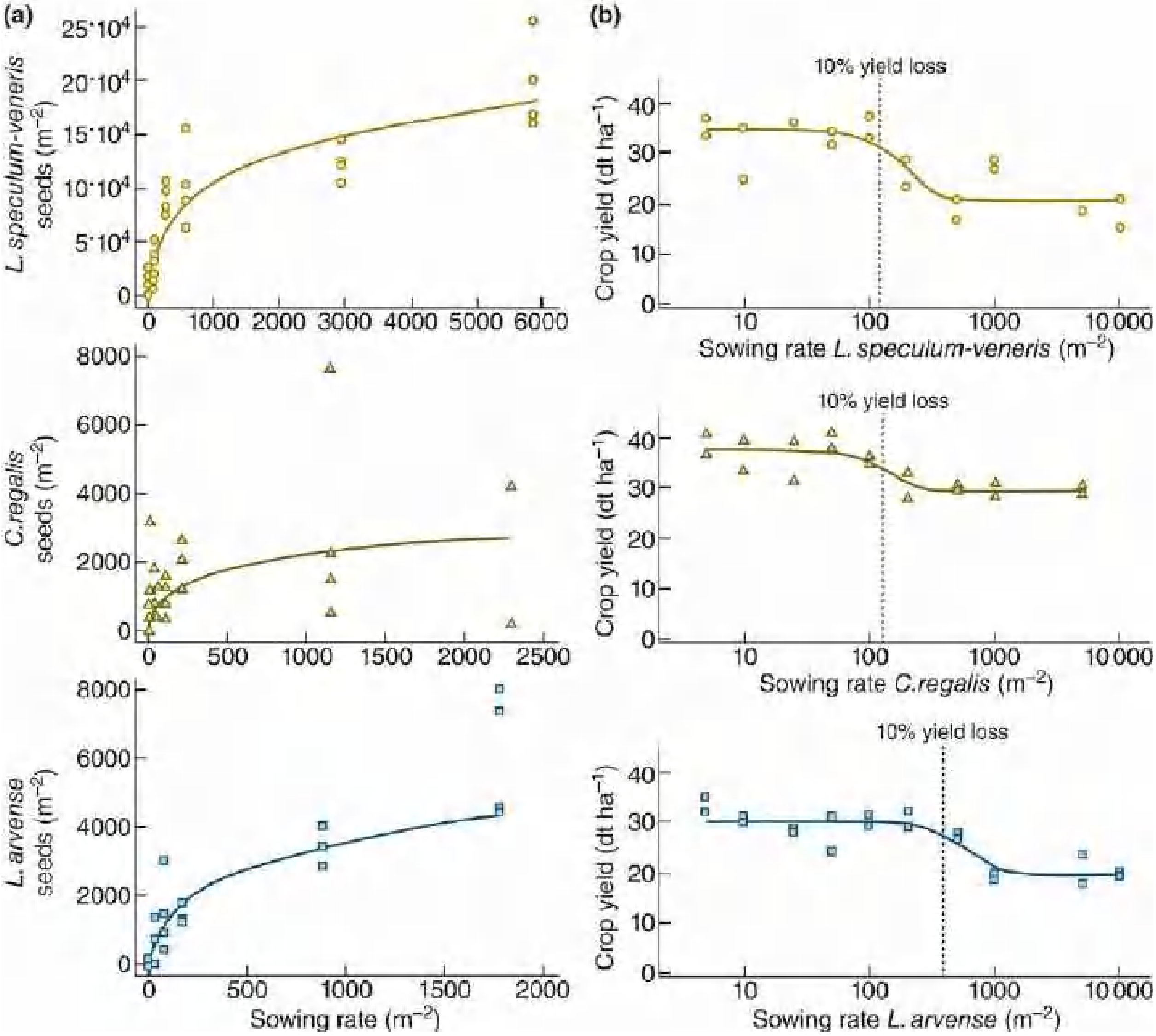


Figure 5.8 Intraspecific competition in plants often regulates the number of modules. When populations of rye grass (*Lolium perenne*) were sown at a range of densities, the range of final tiller (i.e. module) densities was far narrower than that of genets.

Source: After Kays & Harper (1974).

APPLICATION 5.1 Optimal sowing rates for conservation

The general tendency for final yield to saturate – not continue to increase in line with initial density – is something that underlies many activities throughout agriculture, horticulture and even animal husbandry. Generally, we wish to avoid investing unnecessarily in high stocking rates that will not translate ultimately into higher yields. A more unusual example of its application, though, is provided by a study seeking to conserve three species of plant facing extinction in large parts of Europe that may also sometimes be weeds, reducing the productivity of crops with which they may coexist (Lang *et al.*, 2016). The species concerned were European Venus' looking glass, *Legousia speculum-veneris*, forking larkspur, *Consolida regalis*, and field gromwell, *Lithospermum arvense*, which can all act as weeds in fields of winter rye, *Secale cereale*. How might we combine the conservation of these species with avoidance of crop loss? It is clear in Figure 5.9a that as the sowing densities of the plants was increased, significant levelling off was apparent at sowing densities above a few hundreds per square metre (though a strictly 'constant' yield was not reached). We can also see that below these kinds of levels, the yield of the rye was not affected (Figure 5.9b), whereas with further increases in weed seed density, crop yield showed a significant drop (though at densities above around 1000 weed seeds per square metre it again levelled off). We will not discuss *inter*specific competition, between the weeds and the crop, in detail until Chapter 8. Nonetheless, we can understand that very low sowing rates of the three plants would be ineffective in establishing viable populations of these threatened species. However, a levelling off of the yield of new seeds was well underway before sowing densities were reached that affected crop yield. Lang and her co-workers were therefore able to recommend 'optimal' sowing rates of 100 seeds per m^2 for *C. regalis* and *L. arvense* and 50 seeds per m^2 for *L. speculum-veneris*, combining successful establishment, the avoidance of wasted seeds, and negligible crop loss.



[**Figure 5.9**] **Optimal sowing rates can conserve rare weedy plants without threatening crop yields.** (a) In experimental plots near Munich, Germany, seeds of three plant species, *Legousia speculum-veneris* (top), *Consolida regalis* (middle) and *Lithospermum arvense* (bottom), were sown at a range of densities amongst a growing crop plant, rye, *Secale cereale*, and the yield of seeds of the three plants monitored at the end of the growing season. (b) Results for the same plots in terms of the biomass yield of the rye (dry tons per hectare).

Source: After Lang *et al.* (2016).

5.2.3 Density or crowding?

Of course, the intensity of intraspecific competition experienced by an individual is not really determined by the density of the population as a whole. The effect on an individual is determined, rather, by the extent to which it is crowded or inhibited by its immediate neighbours. Even in a population of mobile animals, individuals are unlikely to move around enough to interact with every other member of the population.

three meanings of density

One way of emphasising this is by noting that there's more than one meaning of 'density' (see Lewontin & Levins, 1989 , where details of calculations and terms can be found). Consider a population of insects, distributed over a population of plants (patches of resource) on which they feed. The density would usually be calculated simply as the number of insects (let us say 1000) divided by the number of plants (say 100), i.e. 10 insects per plant. However, this is actually the 'resource-weighted density', but it only gives an accurate measure of the intensity of competition suffered by the insects (the extent to which they are crowded) if there are exactly 10 insects on every plant.

Suppose, instead, that 10 of the plants support 91 insects each, and the remaining 90 support just one insect. The resource-weighted density would still be 10 insects per plant. But to determine average density *experienced* by the insects, we note that 910 of them experienced a density of 91 insects per plant and 90 experienced just one per plant, giving us an overall average of 82.9 insects per plant (910 × 91, plus 90 × 1, divided by 1000) This is the 'organism-weighted density', and it clearly gives a much more satisfactory measure of the intensity of competition the insects are likely to suffer. Clearly, the normal practice of calculating the resource-weighted density and calling it 'the density' can be misleading.

neighbours

The difficulties of relying on density to characterise the potential intensity of intraspecific competition are particularly acute with sessile, modular organisms, because, being sessile, they compete almost entirely only with their immediate neighbours, and being modular, competition is directed most at the modules that are closest to those neighbours. We can see this, for example, when seedlings of three species of fir tree, Douglas fir (*Pseudotsuga menziesii*), grand fir (*Abies grandis*) and noble fir (*A. procera*), were grown together in conspecific pairs or grown alone. In all three, after two years, the diameter growth of branches facing their competitor was significantly reduced in trees grown in pairs compared with the growth of branches of trees grown alone (Figure 5.10a). But these effects of competition were expressed almost entirely by those branches facing their neighbour (Figure 5.10b). Those on the other side of the tree had growth rates similar to those of the isolated plants. The individual plants, and their component modules, were affected by the proximity of their competitors rather than density *per se*.

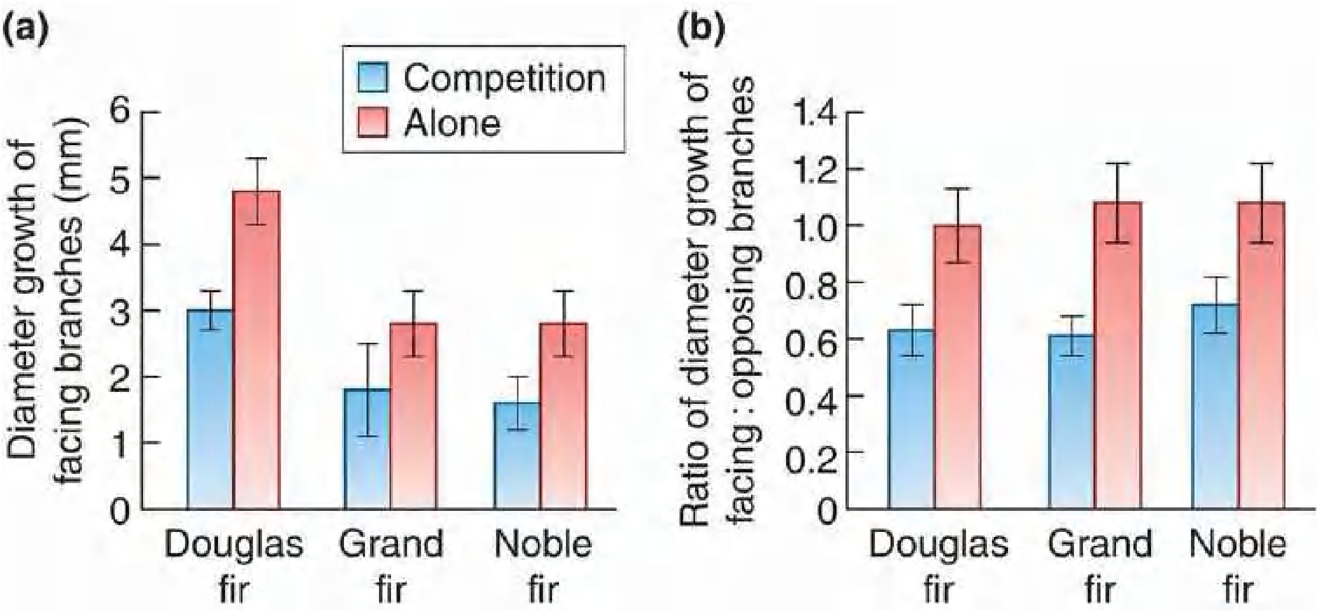


Figure 5.10 **When modular organisms compete, the modules closest to neighbours are the most affected.** (a) The growth of 'facing' branches of three species of fir tree after two years of growth either in pairs (where the branches faced the competitor) or in isolation (where 'facing' branches grew in the same direction but with no competitor). (b) The ratio of the growth of facing branches (from (a)) to 'opposing' branches on the opposite side of the tree seedlings. In isolation, the ratios were not significantly different from 1; but in competition, the ratios indicate that the growth of opposing branches was similar to that of all branches in the isolated plants. Bars are 95% CIs.

Source: From data in Devine & Harrington (2011).

density: a convenient expression of crowding

Thus, whether mobile or sessile, different individuals meet or suffer from different numbers of competitors. Density, especially resource-weighted density, is an abstraction that applies to the population as a whole but need not apply to any of the individuals within it. Nonetheless, density may often be the most convenient way of expressing the degree to which individuals are crowded – and it is certainly the way it has usually been expressed.

5.3 Quantifying intraspecific competition

We have seen that there are general patterns in the action of intraspecific competition, and now we quantify these general patterns more exactly, utilising k values (see Chapter 4) to summarise the effects of intraspecific competition on mortality, and then extending this to fecundity and growth.

use of k values

A k value was defined by the formula:

$$k = \log(\text{initial density}) - \log(\text{final density}), \tag{5.4}$$

or, equivalently:

$$k = \log(\text{initial density}/\text{final density}). \tag{5.5}$$

For present purposes, ‘initial density’ may be denoted by *B*, standing for ‘numbers before the action of intraspecific competition’, while ‘final density’ may be denoted by *A*, standing for ‘numbers after the action of intraspecific competition’. Thus:

$$k = \log(B/A). \tag{5.6}$$

Note that *k* increases as mortality rate increases.

plots of *k* against log density

Some examples of the effects of intraspecific competition on mortality are shown in Figure 5.11, in which *k* is plotted against log *B*. In Figure 5.11a and b, *k* is constant at the lowest densities. This is an indication of density independence: the proportion surviving is not correlated with initial density. At higher densities, *k* increases with initial density; this indicates density dependence. The way in which *k* varies with the logarithm of density indicates the precise nature of the density dependence. Thus, Figure 5.11a and b describe, respectively, situations in which there is under- and exact compensation at higher densities. The exact compensation in Figure 5.11b is indicated by the slope of the curve (denoted by *b*) taking a constant value of 1 (the mathematically inclined will see that this follows from the fact that with exact compensation, *A* is constant). The undercompensation that preceded this at lower densities, and which is seen in Figure 5.11a even at higher densities, is reflected in *b* being less than 1. Of course, *b* does not always vary over the observed range of densities. Figure 5.11c is an example in which density-dependence is undercompensating throughout the range, whereas in Figure 5.11d the mortality shows overcompensation ($b > 1$) throughout the range.

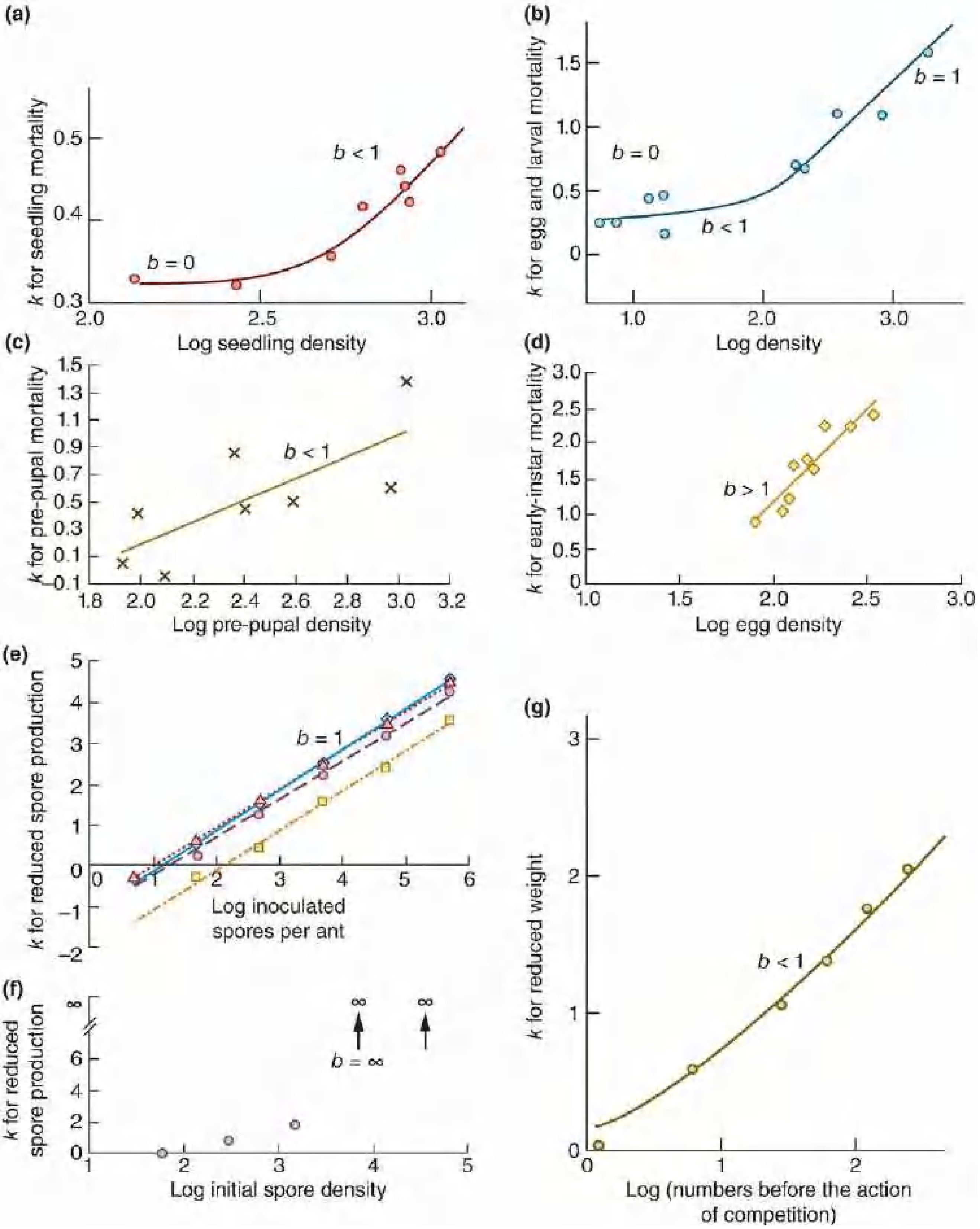

(a)
k for seedling mortality
0.3
0.4
0.5
b = 0
b < 1
2.0
2.5
3.0
Log seedling density
(b)
k for egg and larval mortality
0
0.5
1.0
1.5
b = 0
b < 1
b = 1
1.0
2.0
3.0
Log density
(c)
k for pre-pupal mortality
–0.1
0.1
0.3
0.5
0.7
0.9
1.1
1.3
1.5
b < 1
1.8
2.0
2.2
2.4
2.6
2.8
3.0
3.2
Log pre-pupal density
(d)
k for early-instar mortality
0
0.5
1.0
1.5
2.0
2.5
3.0
b > 1
1.0
1.5
2.0
2.5
3.0
Log egg density
(e)
k for reduced spore production
–2
–1
0
1
2
3
4
5
b = 1
0
1
2
3
4
5
6
Log inoculated spores per ant
(f)
k for reduced spore production
0
2
4
6
∞
∞
∞
b = ∞
1
2
3
4
5
Log initial spore density
(g)
k for reduced weight
0
1
2
3
b < 1
0
1
2
Log (numbers before the action of competition)

Figure 5.11 The use of *k* values for describing patterns of density-dependent mortality, fecundity and growth. (a) Seedling mortality in the dune annual, *Androsace septentrionalis*, in Poland. (b) Egg mortality and larval competition in the almond moth, *Ephestia cautella*. (c) Pre-pupal mortality in the Argentine stem weevil, *Listronotus bonariensis*, in New Zealand. (d) Early instar mortality in the European corn borer, *Ostrinia nubilalis*, in New York state. (e) The reduction in spore production (fecundity) of the fungus *Metarhizium anisopliae* growing on the leaf-cutting ant, *Acromyrmex echinatior*, for three different isolates of the fungus and for a mixture of the three. (f) The reduction in spore production (fecundity) of the fungus *Metschnikowiella bicuspidata* growing on the planktonic crustacean *Daphnia magna*. (g) Reduced growth in the shepherd's purse plant, *Capsella bursa-pastoris*.

Source: (a) After Symonides (1979). (b) After Benson (1973). (c) After Goldson *et al*. (2011). (d) After Kuhar *et al*. (2002). (e) After Hughes *et al*. (2004). (f) After Ebert *et al*. (2000). (g) After Palmblad (1968).

scramble and contest

Exact compensation (*b* = 1) is often referred to as contest competition, because there are a constant number of winners (survivors) in the competitive process. The term was initially proposed by Nicholson (1954), who contrasted it with what he called scramble competition. Scramble is the most extreme form of overcompensating density dependence, in which all competing individuals are so adversely affected that none of them survive, i.e. *A* = 0. We saw an example of this in Figure 5.3a. This would be indicated by a *b* value of infinity (a vertical line). More common is simple overcompensation as in Figure 5.11d.

For fecundity, we think of *B* as the 'total number of offspring that would have been produced had there been no intraspecific competition', i.e. if each reproducing individual had produced as many offspring as it would have done in a competition-free environment. *A* is then the total number of offspring actually produced. (In practice, *B* is usually estimated from the population experiencing the least competition – not necessarily competition-free.) For growth, we think of *B* as the total biomass, or total number of modules, that would have been produced had all individuals grown as if they were in a competition-free situation. *A* is then the total biomass or total number of modules actually produced. Examples are shown in Figure 5.11e–g. The patterns are essentially similar to those in Figure 5.11aÂ–d. Each falls somewhere on the continuum ranging between density independence and pure scramble, and their position along that continuum is immediately apparent: exactly compensating density dependence for fecundity in Figure 5.11e, a *b* value rising to infinity for the reproduction in Figure 5.11f, and density dependence remaining undercompensating for the growth in Figure 5.11g.

5.4 Intraspecific competition and the regulation of population size

We have seen that there are typical patterns in the effects of intraspecific competition on birth and death, and these are summarised in Figure 5.12.

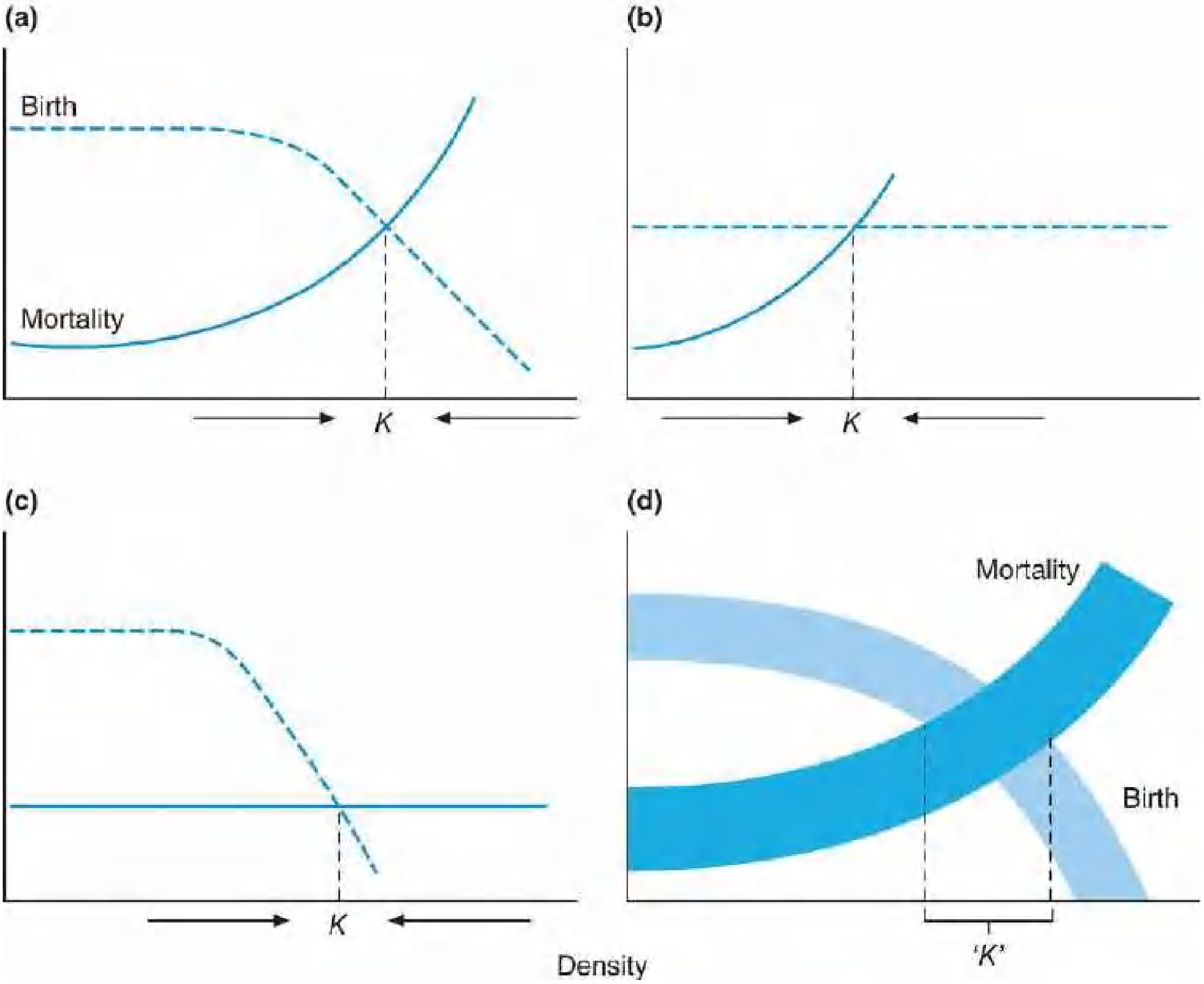


Figure 5.12 Density-dependent birth and mortality rates lead to the regulation of population size. When both are density dependent (a), or when either of them is (b, c), their two curves cross. The density at which they do so is called the carrying capacity (*K*). Below this the population increases, above it the population decreases: *K* is a stable equilibrium. However, these figures are caricatures. The situation is closer to that shown in (d), where mortality rate broadly increases, and birth rate broadly decreases, with density. It is possible, therefore, for the two rates to balance not at just one density, but over a broad range of densities, and it is towards this broad range that other densities tend to move.

5.4.1 Carrying capacities

Figures 5.12a–c reiterate the fact that as density increases, the per capita birth rate eventually falls and the per capita death rate eventually rises. There must, therefore, be a density at which these curves cross. At densities below this point, the birth rate exceeds the death rate and the population increases in size. At densities above the crossover point, the death rate exceeds the birth rate and the population declines. At the crossover density itself, the two rates are equal and there is no net change in population size. This density therefore represents a stable equilibrium, in that all other densities will tend to approach it. In other words, intraspecific competition, by acting on birth rates and death rates, can regulate populations at a stable density at which the birth rate equals the death rate. This density is known as the carrying capacity of the population and is usually denoted by *K* (Figure 5.12). It is called a carrying capacity because it represents the population size that the resources of the environment can just maintain ('carry') without a tendency to either increase or decrease.

real populations lack simple carrying capacities

However, while hypothetical populations caricatured by line drawings like Figures 5.12a–c can be characterised by a simple carrying capacity, this is not true of any natural population. There are unpredictable environmental fluctuations, individuals are affected by a whole wealth of factors of which intraspecific competition is only one, and resources not only affect density but respond to density as well. Hence, the situation is likely to be closer to that depicted in Figure 5.12d. Intraspecific competition does not hold natural populations to a predictable and unchanging level (the carrying capacity), but it may act upon a very wide range of starting densities and bring them to a much narrower range of final densities, and it therefore tends to keep density within certain limits. It is in this sense that intraspecific competition may be said typically to be capable of regulating population size.

In fact, the concept of a population settling at a stable carrying capacity, even in caricatured populations, is relevant only to situations in which density dependence is not strongly overcompensating. Where there is overcompensation, cycles or even chaotic changes in population size may be the result. We return to this point later (see Section 5.6.5).

5.4.2 Net recruitment curves

peak recruitment occurs at intermediate densities

An alternative general view of intraspecific competition is shown in Figure 5.13a, which deals with numbers rather than rates. The difference there between the births curve and the deaths curve is ‘net recruitment’, the net number of additions expected in the population during the appropriate stage or over one interval of time. Because of the shapes of the curves, the net number of additions is small at the lowest densities, increases as density rises, declines again as the carrying capacity is approached and is then negative (deaths exceed births) when the initial density exceeds K (Figure 5.13b). Thus, total recruitment into a population is small when there are few individuals available to give birth, and small when intraspecific competition is intense. It reaches a peak, i.e. the population increases in size most rapidly, at some intermediate density.

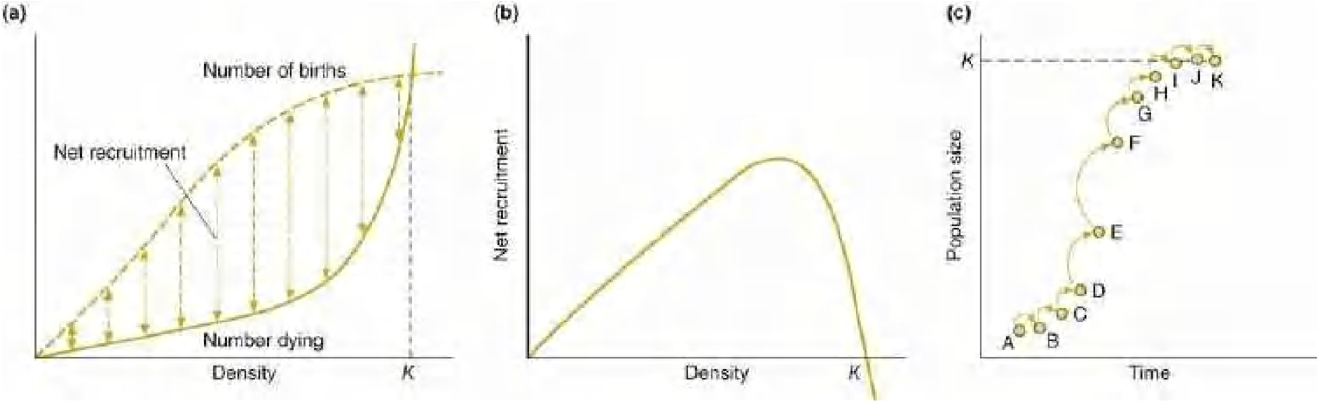
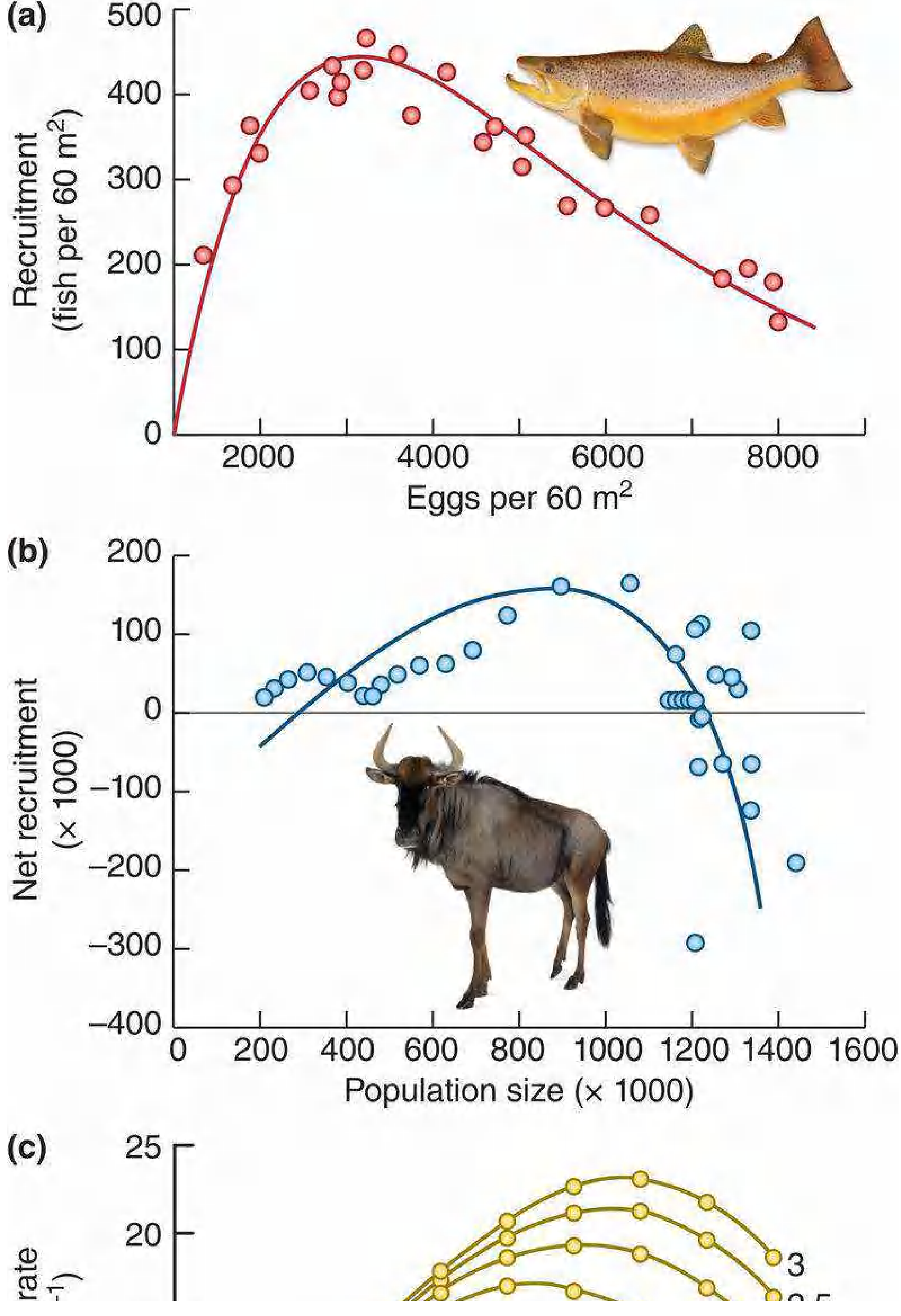
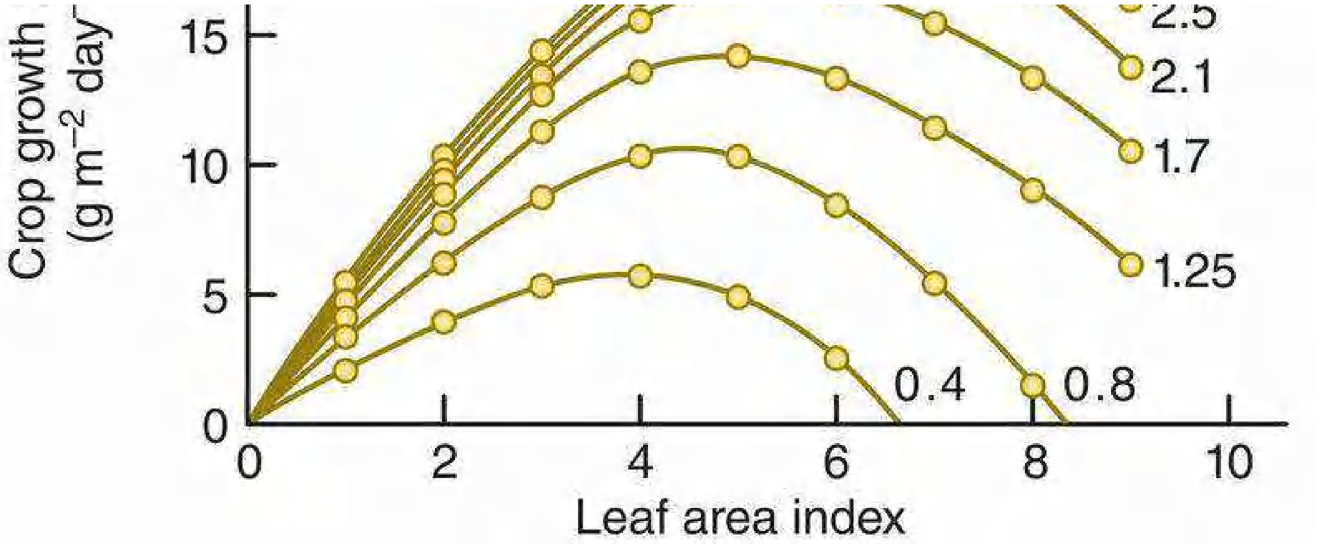


Figure 5.13 **Intraspecific competition typically generates n-shaped net recruitment curves and S-shaped growth curves.** (a) Density-dependent effects on the numbers dying and the number of births in a population: net recruitment is 'births minus deaths'. Hence, as shown in (b), the density-dependent effect of intraspecific competition on net recruitment is a domed or 'n'-shaped curve. (c) A population increasing in size under the influence of the relationships in (a) and (b). Each arrow represents the change in size of the population over one interval of time. Change (i.e. net recruitment) is small when density is low (i.e. at small population sizes: A to B, B to C) and is small close to the carrying capacity (I to J, J to K), but is large at intermediate densities (E to F). The result is an 'S'-shaped or sigmoidal pattern of population increase, approaching the carrying capacity.

The precise nature of the relationship between a population's net rate of recruitment and its density varies with the detailed biology of the species concerned (e.g. the trout, wildebeest and clover plants in Figure 5.14a–c). Moreover, because recruitment is affected by a whole multiplicity of factors, the data points rarely fall exactly on any single curve. Yet, in each case in Figure 5.14, a domed curve is apparent. This reflects the general nature of density-dependent birth and death whenever there is intraspecific competition. Note also that one of these (Figure 5.14c) is modular: it describes the relationship between the leaf area index (LAI) of a plant population (the total leaf area being borne per unit area of ground) and the population's growth rate (modular birth minus modular death). The growth rate is low when there are few leaves, peaks at an intermediate LAI, and is then low again at a high LAI, where there is much mutual shading and competition and many leaves may be consuming more in respiration than they contribute through photosynthesis. We return to these net recruitment curves in Section 15.3 when we look in detail at how natural populations (fisheries, forests) may be exploited – pushed from right to left along their net recruitment curves – in order to optimise the sustainable harvest we can take from them.


(a)
500
400
300
200
100
0
Recruitment (fish per 60 m^2)
2000
4000
6000
8000
Eggs per 60 m^2
(b)
200
100
0
−100
−200
−300
−400
Net recruitment (× 1000)
0
200
400
600
800
1000
1200
1400
1600
Population size (× 1000)
(c)
25
20
rate
3
2.5



[Figure 5.14](#) Some dome-shaped net-recruitment curves. (a) Six-month old brown trout, *Salmo trutta*, in Black Brows Beck, UK, between 1967 and 1989. (b) Wildebeest, *Connochaetes taurinus*, in the Serengeti, Tanzania 1959–95. (c) The relationship between crop growth rate of subterranean clover, *Trifolium subterraneum*, and population size (leaf area index) at various intensities of radiation (0.4–3 kJ cm^{-2} day^{-1}).

Source: (a) After Myers (2001), following Elliott (1994). (b) After Mduma *et al.* (1999). (c) After Black (1963).

back to an integral projection model of Soay sheep

Finally here, we return to Figure 4.17, where integral projection models were used to combine density-dependent patterns of growth, survival and fecundity for Soay sheep in Scotland in order to estimate how a population's net reproductive rate will itself vary with abundance. We saw, as we have now come to expect, that as abundance increased the net reproductive rate, R, declined (Figure 4.17e). What is also apparent now is that ln R (= r) was equal to zero (R = 1) at a population size of around 455 sheep, which is therefore the predicted carrying capacity of the population, and an equilibrium, with positive values of ln R at population sizes smaller than this, but negative values at population sizes larger.

5.4.3 Sigmoidal growth curves

Curves of the type shown in Figure 5.13a and b may also be used to suggest the pattern by which a population might increase from an initially very small size (e.g. when a species colonises a previously unoccupied area). This is illustrated in Figure 5.13c. Imagine a small population, well below the carrying capacity of its environment (point A). Because the population is small, it increases in size only slightly during one time interval, and only reaches point B. Now, however, being larger, it increases in size more rapidly during the next time interval (to point C), and even more during the next (to point D). This process continues until the population passes beyond the peak of its net recruitment curve (Figure 5.13b). Thereafter, the population increases in size less and less with each time interval until the population reaches its carrying capacity (K) and ceases completely to increase in size. The population might therefore be expected to follow an S-shaped or 'sigmoidal' curve as it rises from a low density to its carrying capacity. This is a consequence of the hump in its recruitment rate curve, which is itself a consequence of intraspecific competition.

Of course, Figure 5.13c, like the rest of Figure 5.13, is a gross simplification. It assumes, apart from anything else, that changes in population size are affected only by intraspecific competition.

Nevertheless, something akin to sigmoidal population growth can be perceived in many natural and experimental situations (Figure 5.15).

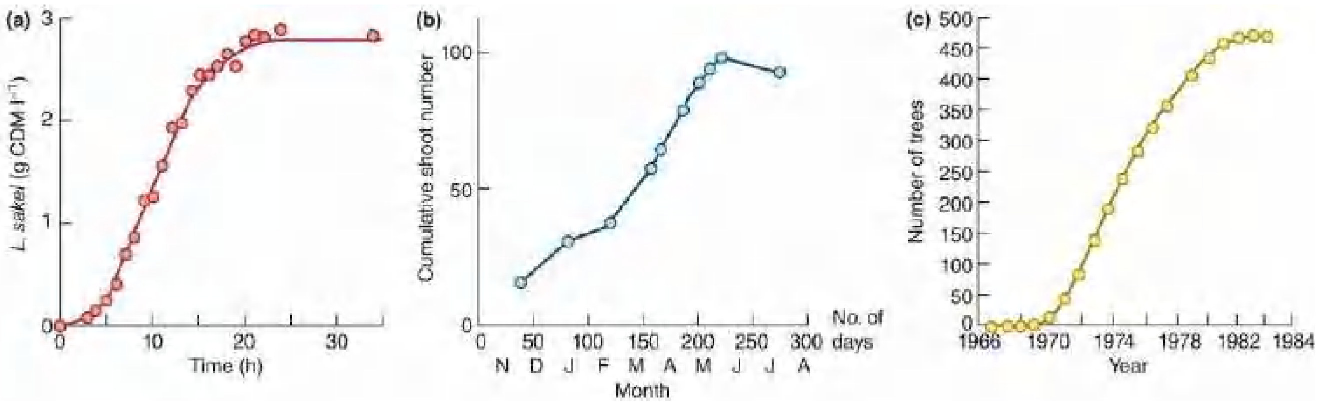


Figure 5.15 Real examples of S-shaped population increase. (a) The bacterium *Lactobacillus sakei* (measured as grams of 'cell dry mass' or CDM per litre) grown in nutrient broth. (b) The population of shoots of the annual plant *Juncus gerardii* in a salt marsh habitat on the west coast of France. (c) The population of the willow tree (*Salix cinerea*) in an area of land after myxomatosis had effectively prevented rabbit grazing.

Source: (a) After Leroy & de Vuyst (2001). (b) After Bouzille *et al.* (1997). (c) After Alliende & Harper (1989).

Intraspecific competition will be obvious in certain cases (such as overgrowth competition between sessile organisms on a rocky shore), but this will not be true of every population examined. Individuals are also affected by predators, parasites and prey, competitors from other species, and the many facets of their physical and chemical environment. Any of these may outweigh or obscure the effects of intraspecific competition; or the effect of these other factors at one stage may reduce the density to well below the carrying capacity for all subsequent stages. Nevertheless, intraspecific competition probably affects most populations at least sometimes during at least one stage of their life cycle.

APPLICATION 5.2 Human population growth and a global carrying capacity

The root of many of the environmental problems facing us is a large and growing global human population. More people means more competition: for renewable resources like fish and forests, for food production from agriculture, but also for energy, and non-renewable resources like oil and minerals. During the first half of the 20th century, the global population increased by 40%, from 1.8 to 2.5 billion people. But since then the population has almost tripled to over seven billion. What's more, the percentage of people living in cities has also grown steadily (Figure 5.16). By 2010, the number of city dwellers had equalled the number in rural environments for the first time in history, and the United Nations predicts that this trend will continue, with two-thirds of the global population living in cities by 2050. The human population is growing ever larger and getting disproportionately crowded. We have seen through this chapter that the normal consequence, when populations grow, is that competition eventually slows that rate of growth and ultimately stops it, and that the overall size of the population settles, if not at a fixed carrying capacity, then within some regulated band. Is this what we've seen with the human population?

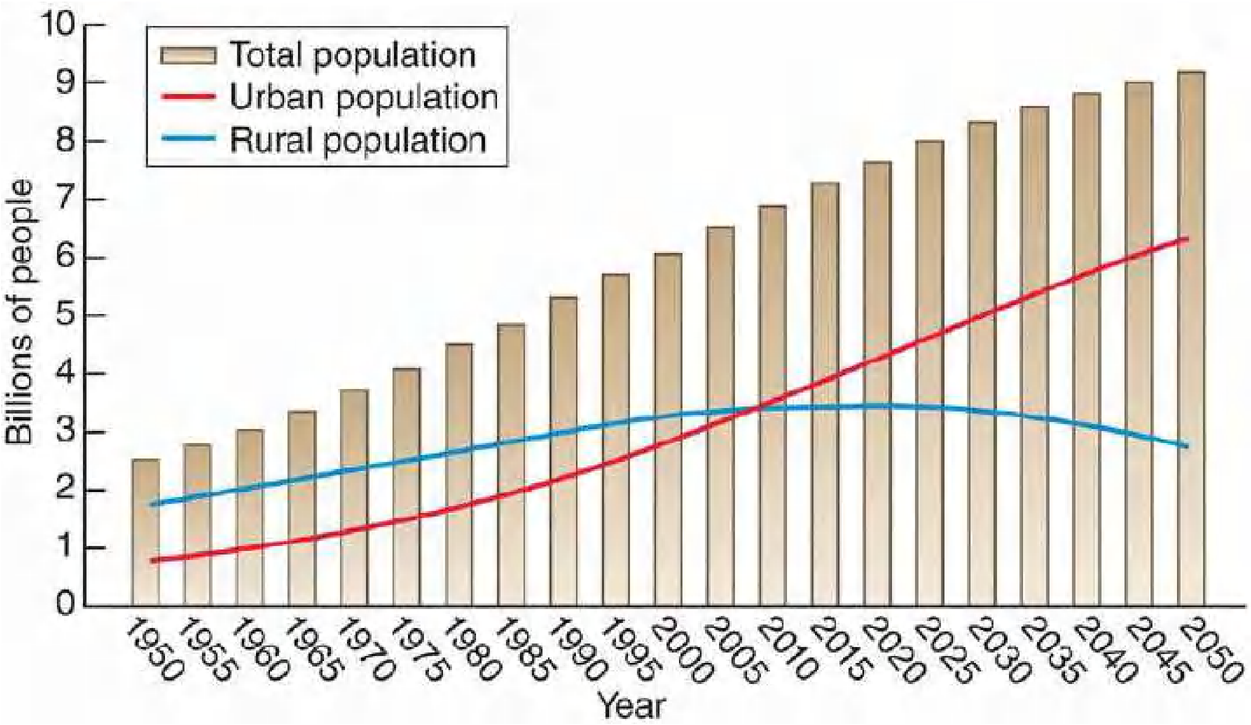


Figure 5.16 The global urban population has overtaken its rural counterpart and will probably run away from it. The sizes of the total, rural and urban populations of the world from 1950 to 2010, and projections from the United Nations up to 2050.

Source: After UNEP (2014).

population growth up to the present

With exponential growth (Section 5.6), the population as a whole grows at an accelerating rate simply because the growth rate is a product of a constant individual rate of increase and the accelerating number of individuals. For thousands of years, the growth of the global human population appears indeed to have been exponential (Figure 5.17). But the growth rate was slow, despite a jump around 10 000 years ago at the dawn of agriculture. More recently, however, with growing urbanisation and industrialisation, growth, far from slowing, accelerated to become faster than exponential for several centuries because the *per capita* rate itself increased. Only very recently has the rate slowed again.

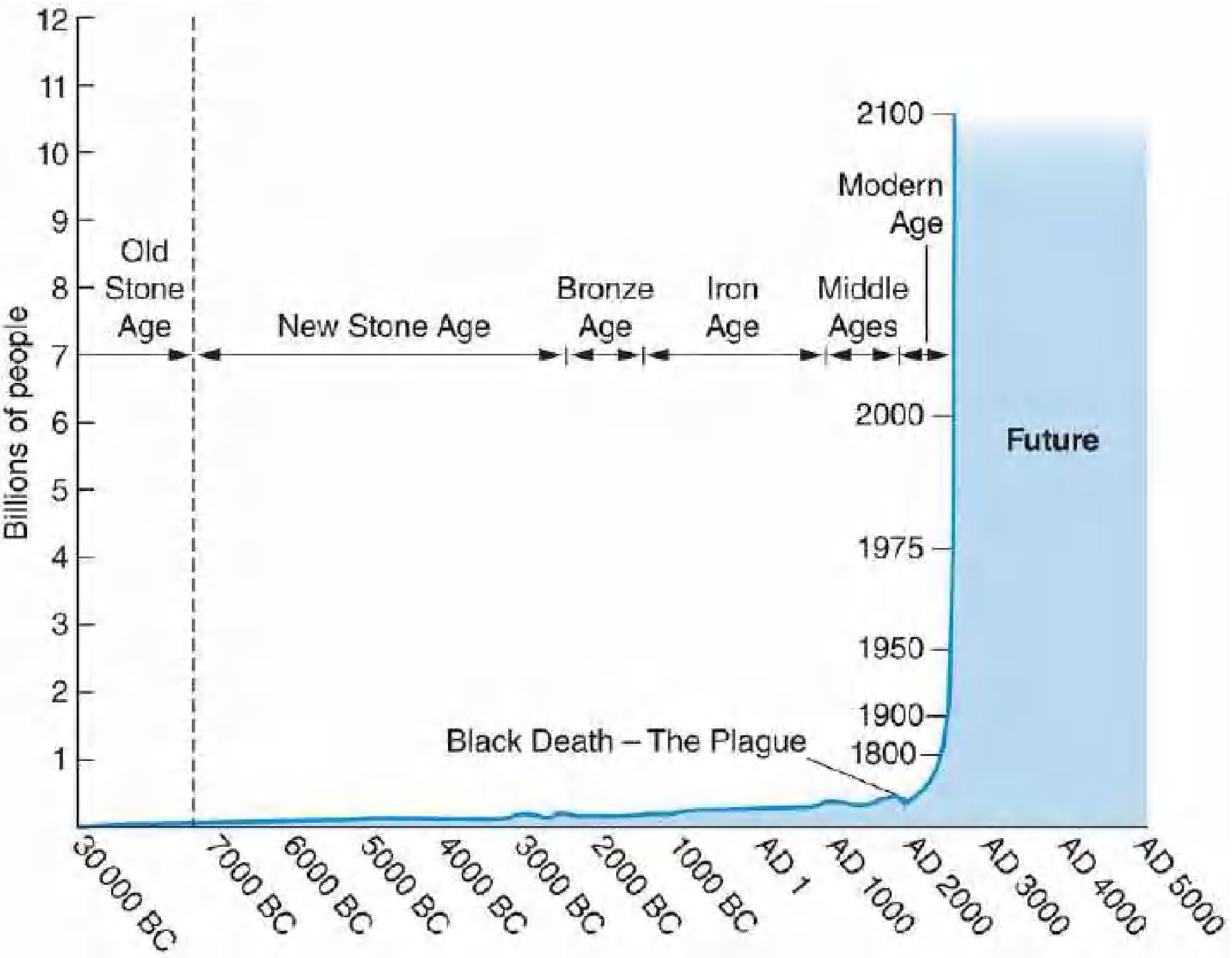


Figure 5.17 The global human population grew slowly for millennia but has recently shown faster than exponential growth. The estimated size of the global human population over the past 30 000 years and projected into the future.

Source: After Population Reference Bureau (2006).

Are these modest indications of a slow-down a sign that competition is intensifying? If so, this is far from being the whole story. We humans already appropriate a high proportion of the global plant production for our own uses (discussed further in Application 20.2), and average food consumption per person has not been falling, as it would with intensifying competition, but rising. It has increased steadily over the past 50 years, from 2360 calories per day in the mid-1960s to 2940 calories today (WHO, 2013). Both figures exceed the 2250 calories per day estimated by the US National Institutes of Health to be sufficient for a moderately active adult. Of course, hunger and malnutrition remain major problems in many areas, with perhaps one billion people receiving insufficient food. Yet even in developing countries, average consumption has increased from 2054 calories per day in the 1960s to 2850 today. Hunger results not from inadequate global food production but from unequal distribution.

demographic transitions

In fact, the slow-down in population growth seems to have less to do with a direct effect of resource shortages than with a change in individuals' social conditions and decision-making. In particular, we have seen in human populations in many parts of the world a *demographic transition*– a switch from a combination of high birth and death rates to one of low birth and death rates. Indeed, we can distinguish three categories of human population: those that passed through this demographic transition before 1945 ('early') (Figure 5.18), those that have passed through one since 1945 ('late'), and those that have not yet passed through the transition. The pattern is as follows. Initially, both the birth rate and the death rate are high, but the former is only slightly greater than the latter, so the overall rate of population increase is only moderate or small. As we saw in Figure 5.17, this, broadly, was the case for the global human population until around 300 years ago. Next, the death rate declines while the birth rate remains high, so the population growth rate increases, giving us the more-than-exponential rate we also saw in Figure 5.17. Next, however, the birth rate also declines until it is similar to or perhaps even lower than the death rate. Hence, the population growth rate eventually declines again and may even become negative, though with a far larger population than before the transition began.

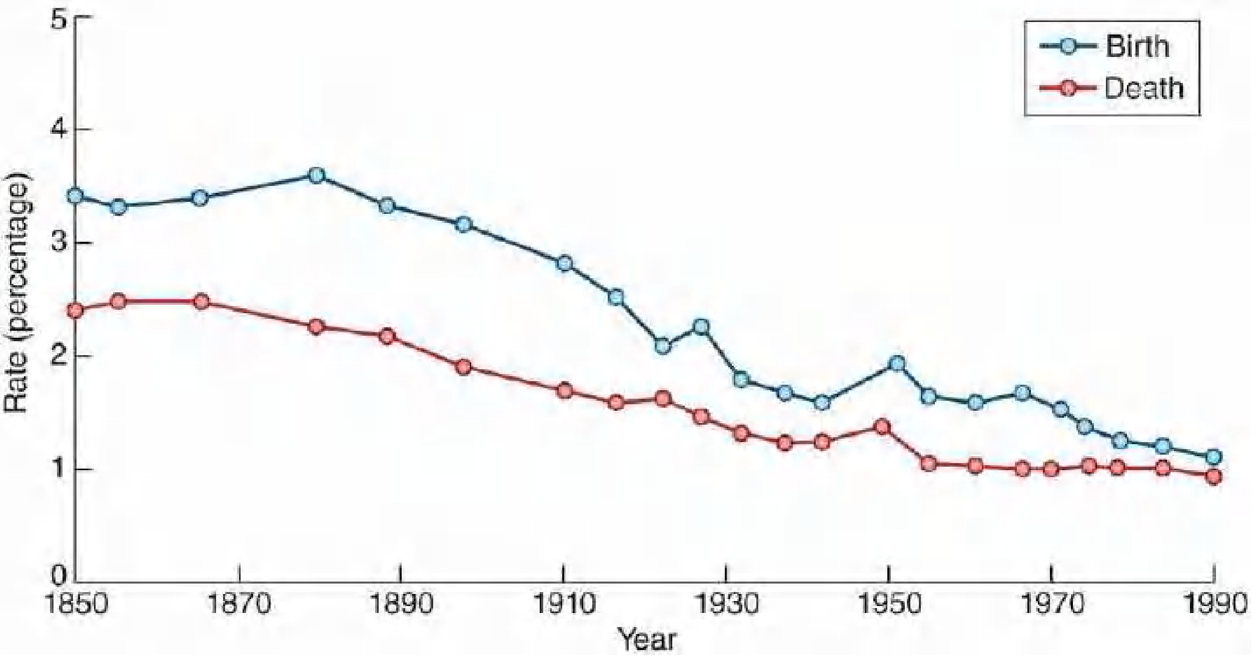


Figure 5.18 The birth and death rates in Europe since 1850. The annual net rate of population growth is given by the gap between the two. Death rates declined in the late 19th century, followed decades later by a decline in birth rates, leading ultimately to a narrowing of the gap between the two.

Source: After Cohen (1995).

The generally accepted explanation, though probably not the whole story, is that the transition is an inevitable consequence of industrialisation, education, and general modernisation, leading first, through medical advances, to the drop in death rates, and then, through the choices people make (such as delaying having children) to the drop in birth rates. Certainly, when we consider the populations of the different regions of the world together, there has been a dramatic decline from the peak population growth rate of about 2.1% per year in 1965–70 to around 1.1–1.2% per year today (Figure 5.19a).

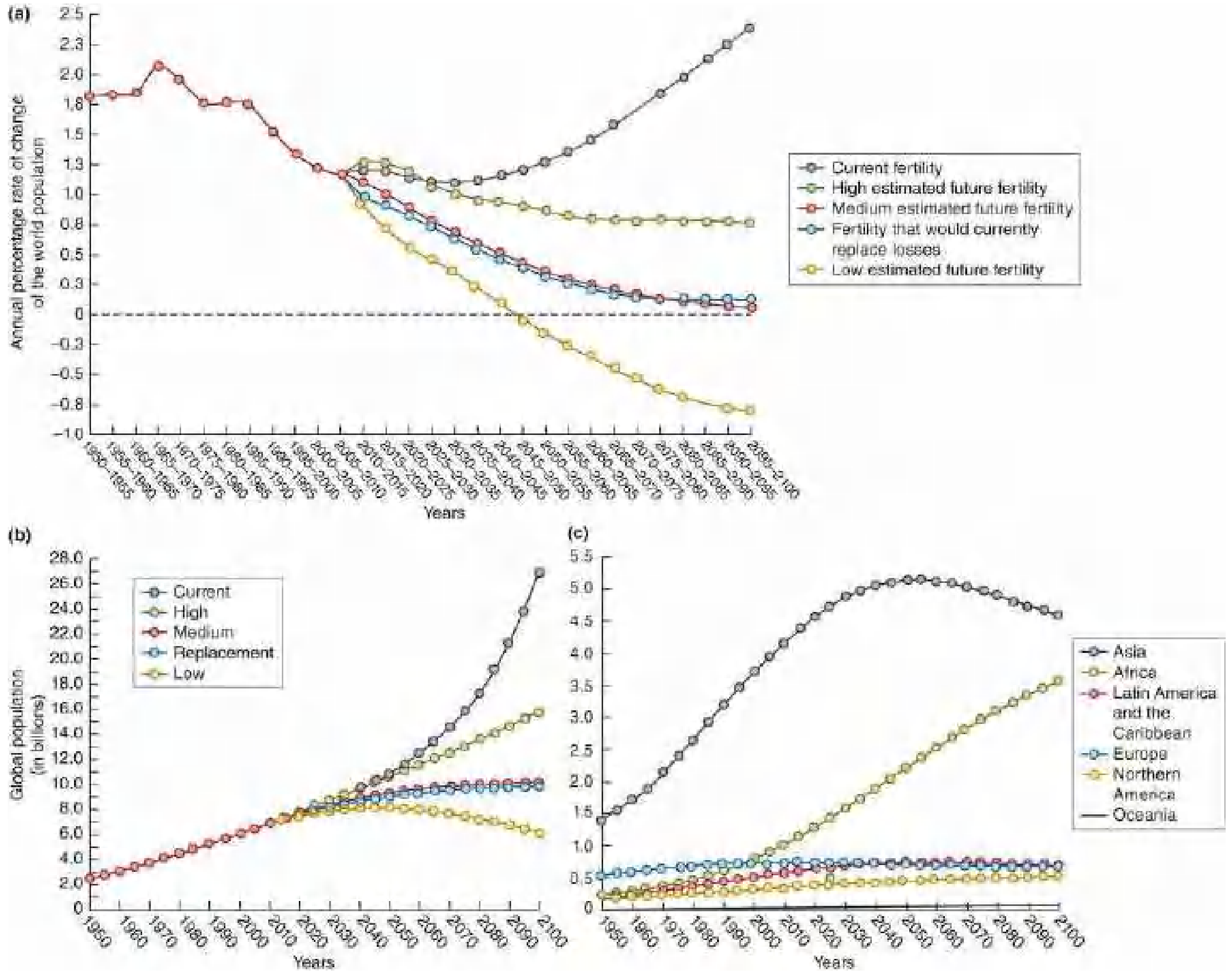


Figure 5.19 **What happens to the global human population size depends on future fertility patterns.** (a) The average annual percentage rate of change of the world population observed from 1950 to 2010, and projected forward to 2100 on the basis of various assumptions about future fertility rates. (b) The estimated size of the world's population from 1950 and 2010 and projected forward to 2100 on the basis of various assumptions regarding fertility rates. (c) The estimated size of the populations of the world's main regions from 1950 and 2010 and projected forward to 2100 assuming 'medium' fertility rates.

Source: After United Nations (2011).

a global carrying capacity?

It seems clear, then, that the rate of human population growth is slowing not simply as a result of intraspecific competition, but as a result of the choices people make. Nonetheless, if current trends continue, we might hope that the size of the global human population could level off and approach what, in terms of intraspecific competition, we would call a global carrying capacity. This in turn raises the question of what a reasonable global carrying capacity would be. Estimates have been proposed over the last 300 years or so. They vary to an astonishing degree. Even those suggested since 1970 span three orders of magnitude – from 1 to 1000 billion. To illustrate the difficulty of arriving at a good estimate, we can look at a few examples (see Cohen, 1995 , 2005 for further details).

In 1679, van Leeuwenhoek estimated the inhabited area of the Earth as 13 385 times larger than his home nation of Holland, whose population then was about one million people. He assumed all this area could be populated as densely as Holland, yielding an upper limit of roughly 13.4 billion. In 1967, De Wit asked how many people could live on Earth if photosynthesis was the limiting factor (but neither water nor minerals were limiting) and suggested 1000 billion, though if people wanted to eat meat or have a reasonable amount of living space the estimate would be lower. By contrast, Hulett in 1970 assumed levels of affluence and consumption in the USA were optimal for the whole world, and he included requirements not only for food but also for renewable resources like wood and non-renewable resources like steel and aluminium. He suggested a limit of no more than 1 billion. Kates and others made similar assumptions using global rather than US averages. They estimated a global carrying capacity of 5.9 billion people subsisting on a basic diet (principally vegetarian), 3.9 billion on an 'improved' diet (about 15% of calories from animal products), or 2.9 billion on a diet with 25% of calories from animal products.

As Cohen (2005) has pointed out, most estimates have relied heavily on a single dimension – biologically productive land area, water, energy, food and so on – when in reality the impact of one factor depends on the value of others. Thus, for example, if water is scarce and energy is abundant, water can be desalinated and transported to where it is in short supply, a solution that is not available if energy is expensive. And as the examples above make clear, there is a difference between the number the Earth can support (the concept of a carrying capacity we normally apply to other organisms) and the number it can support at an acceptable standard of living. It is unlikely that many of us would choose to live crushed up against an environmental ceiling or wish it on our descendants.

what is the 'human population problem'?

Our difficulties in defining a global carrying capacity raise a deeper difficulty. What is 'the human population problem'? It may be simply that the present size of the global human population is unsustainably high – greater than the (presently unknown) carrying capacity. Or it may be not the size of the population but its distribution over the Earth that is unsustainable. Crowding as much as population size is the problem. As we have seen, the fraction of the population concentrated in urban environments has risen from around 3% in 1800 to more than 50% today. Each agricultural worker today has to feed her- or himself plus one city dweller. By 2050 that will have risen to each worker feeding two urbanites (Cohen, 2005). Or perhaps it is not the size but the age distribution of the global population that is unsustainable. In developed regions, the percentage of the population over 65 rose from 7.6% in 1950 to 12.1% in 1990. This proportion is now increasing faster still, as the large cohort born after World War II passes 65. Or finally it may not be that resources are limited but that their uneven distribution is unsustainable. Competition may be unbearably intense for some, while for others density-independence prevails. In 1992, the 830 million people of the world's richest countries enjoyed an average income equivalent to US$22 000 per annum. The 2.6 billion people in the middle-income countries received $1600. But the two billion in the poorest countries got just $400. These averages themselves hide other enormous inequalities.

Of course, the human population problem, just like the problem in any crowded population, is not simply one of intense intraspecific competition for limited resources. Individuals in poor condition may be more vulnerable to predation and parasitism, and the spread of parasites may itself be enhanced. We return to the ways in which the abundances of populations are determined by the combination of forces acting on them in Chapter 14.

inescapable momentum

Finally we can ask what would happen if it were possible to bring demographic transition to all countries of the world so birth rates equalled death rates and population growth was zero. Would the population problem be solved? The answer is no, for at least two important reasons. We saw in Chapter 4 that the net reproductive rate of a population is a reflection of age-related patterns of survival and birth, but these patterns also give rise to different age structures within the population. If birth rates are high but survival rates low ('pre-transition'), there will be many young and relatively few old individuals in the population. But if birth rates are low and survival rates high – the ideal to which we might aspire post-transition – relatively few young, productive individuals must support the many who are old, unproductive, and dependent: an aspect of the problem that we noted above.

In addition, even if our understanding was so sophisticated and our power so complete that we could establish equal birth and death rates tomorrow, would the human population stop growing? The answer, once again, is 'No.' Population growth has its own momentum, and even with birth rate matched to death rate, it would take many years to establish a stable age structure, while considerable growth continued in the meantime. According to projections by the United Nations, even with low fertility the world's population will grow from slightly more than seven billion today to more than eight billion by 2050 (Figure 5.19b). There are many more babies in the world now than 25 years ago, so even if birth rate *per capita* drops considerably now, there will still be many more births in 25 years' time than now, and these children, in turn, will continue the momentum before an approximately stable age structure is eventually established. As Figure 5.19c shows, it is the populations in the developing regions of the world, dominated by young individuals, that will provide most of the momentum for further population growth.

5.5 Mathematical models: introduction

The desire to formulate general rules in ecology often finds its expression in the construction of mathematical or graphical models. It may seem surprising that ecologists interested in the natural living world should spend time reconstructing it in an artificial mathematical form. But there are several good reasons why this should be done. The first is that models can crystallise, or at least bring together in terms of a few parameters, the important, shared properties of a wealth of unique examples. This simply makes it easier for ecologists to think about the problem or process under consideration, by forcing us to try to extract the essentials from complex systems. Thus, a model can provide a 'common language' in which each unique example can be expressed; and if each can be expressed in a common language, then their properties relative to one another, and relative perhaps to some ideal standard, will be more apparent.

Even more important, models can shed light on the real world that they mimic. They can exhibit properties that the system being modelled had not previously been known to possess, and they can make it clear how the behaviour of a population, for example, depends on the properties of the individuals that comprise it. Models do this because they force us to make explicit any assumptions that we choose to make, and mathematical methods are designed precisely to allow a set of assumptions to be followed through to their natural conclusions. We can ask, for example, 'If only juveniles migrate, what would this do to the dynamics of their populations?'. As a consequence, models often suggest what would be the most profitable experiments to carry out or observations to make – 'Since juvenile migration rates appear to be so important, these should be measured in each of our study populations'.

These ideas are more familiar, perhaps, in other contexts, but are equally valid in ecology. Other than in their imaginations, Newton never laid hands on a perfectly frictionless body, and Boyle never saw an ideal gas. But Newton's Laws of Motion and Boyle's Law have been of immeasurable value to us for centuries.

These reasons for constructing models are also criteria by which any model should be judged. Indeed, a model is only useful (i.e. worth constructing) if it does perform one or more of these functions. Of course, in order to perform them a model must adequately describe real situations and real sets of data, and this ‘ability to describe’ or ‘ability to mimic’ is itself a further criterion by which a model can be judged. However, the crucial word is ‘adequate’. The only perfect description of the real world is the real world itself. A model is an adequate description, ultimately, as long as it performs a useful function.

In the present case, some simple models of intraspecific competition will be described. They will be built up from a very elementary starting point, and their properties (i.e. their ability to satisfy the criteria described above) will then be examined. Initially, a model will be constructed for a population with discrete breeding seasons.

5.6 A model with discrete breeding seasons

5.6.1 Basic equations

In Section 4.7 we developed a simple model for species with discrete breeding seasons, in which the population size at time t, N_t, altered in size under the influence of a fundamental net reproductive rate, R. This model can be summarised in two equations:

$$N_{t+1} = N_t R \tag{5.7}$$

and:

$$N_t = N_0 R^t. \tag{5.8}$$

no competition: exponential growth

The model, however, describes a population in which there is no competition. R is constant, and if $R > 1$, the population will continue to increase in size indefinitely (‘exponential growth’, shown in Figure 5.20a).

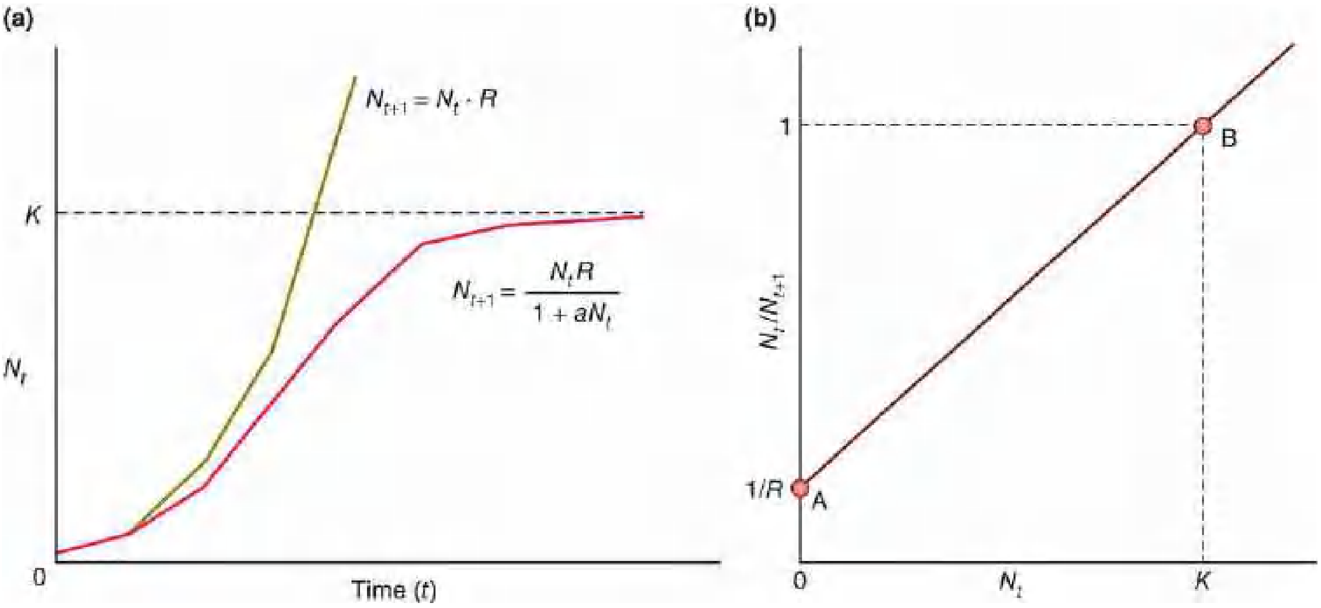


Figure 5.20 **Mathematical models of population increase.** (a) In populations with discrete generations population increases with time: exponential increase (left) and sigmoidal increase (right). (b) The simplest, straight-line way in which the inverse of generation increase (N_t/N_{t+1}) might rise with density (N_t) offers a way of adding competition to exponential increase. For further explanation, see text.

incorporating competition

The first step is therefore to modify the equations by making the net reproductive rate subject to intraspecific competition. This is done in Figure 5.20b, which has three components. At point A, the population size is very small (N_t is virtually zero). Competition is therefore negligible, and the net reproductive rate is adequately defined by an unmodified R. Thus, Equation 5.7 is still appropriate, or, rearranging the equation:

$$N_t/N_{t+1} = 1/R. \tag{5.9}$$

At point B, by contrast, the population size (N_t) is very much larger and there is a significant amount of intraspecific competition, such that the net reproductive rate has been so modified by competition that the population can collectively do no better than replace itself each generation, because ‘births’ equal ‘deaths’. In other words, N_{t+1} is simply the same as N_t, and N_t/N_{t+1} equals 1. The population size at which this occurs is, by definition, the carrying capacity, K (see Figure 5.13).

The third component of Figure 5.20b is the straight line joining point A to point B and extending beyond it. This describes the progressive modification of the net reproductive rate as population size increases; but its straightness issimply an assumption made for the sake of expediency, since all straight lines are of the simple form: $y =$ (slope)x + (intercept). In Figure 5.21b, N_t/N_{t+1} is measured on the y-axis, N_t on the x-axis, the intercept is $1/R$ and the slope, based on the segment between points A and B, is $(1 - 1/R)/K$. Thus:

$$\frac{N_t}{N_{t+1}} = \frac{1 - \frac{1}{R}}{K} \cdot N_t + \frac{1}{R} \tag{5.10}$$

or, rearranging:

$$N_{t+1} = \frac{N_t R}{1 + \frac{(R-1)N_t}{K}}. \tag{5.11}$$

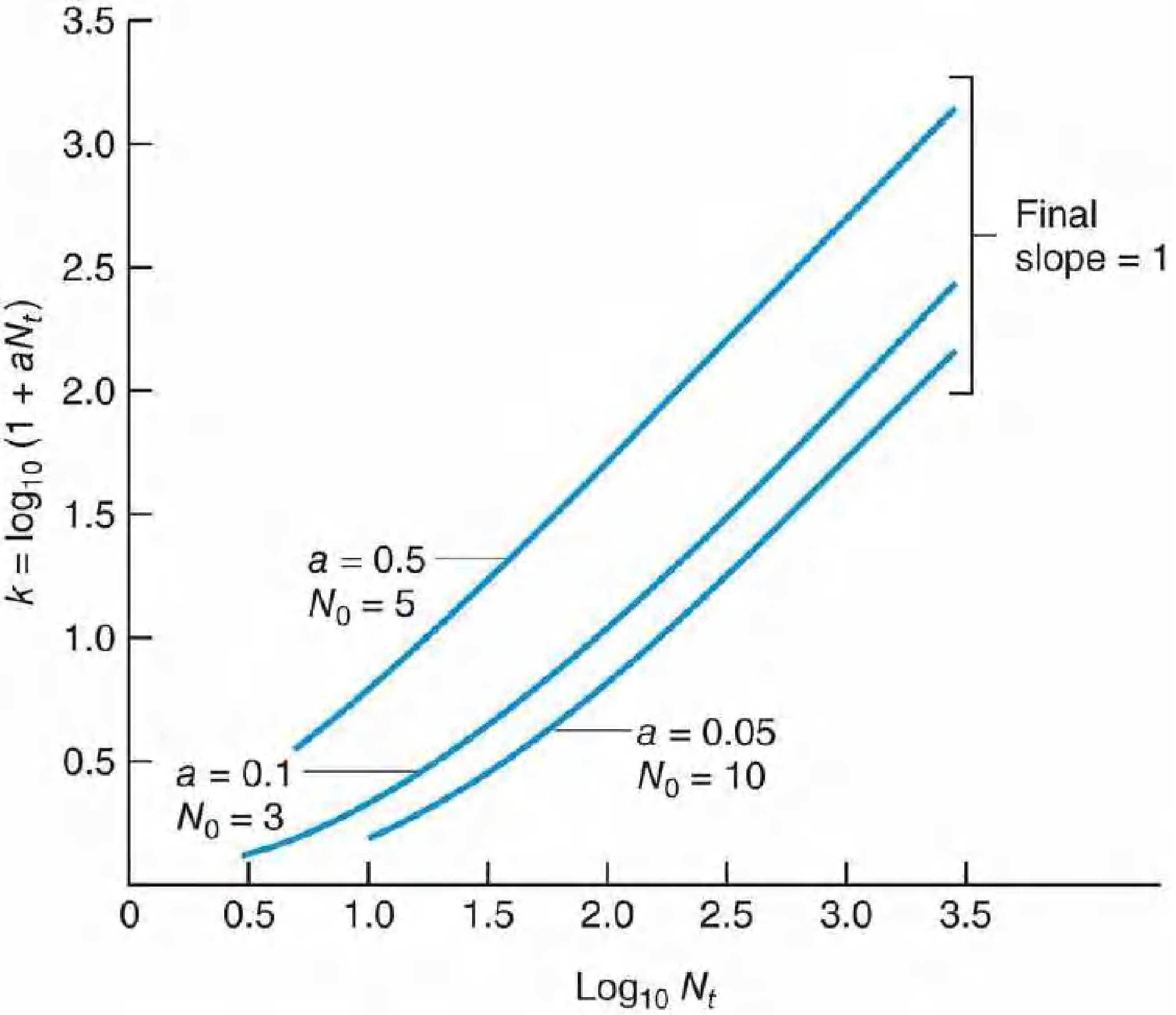


Figure 5.21 The intraspecific competition inherent in Equation 5.13. The final slope of k against $\log_{10} N_t$ is unity (exact compensation), irrespective of the starting density N_0 or the constant a (= $(R - 1)/K$).

a simple model of intraspecific competition

For further simplicity, $(R - 1)/K$ may be denoted by a giving:

$$N_{t+1} = \frac{N_t R}{(1 + aN_t)}. \tag{5.12}$$

This is a model of population increase limited by intraspecific competition. Its essence lies in the fact that the unrealistically constant R in Equation 5.7 has been replaced by an actual net reproductive rate, $R/(1 + aN_t)$, which decreases as population size (N_t) increases.

which comes first – *a* or *K*?

We, like many others, derived Equation 5.12 as if the behaviour of a population is jointly determined by *R* and *K*, the per capita rate of increase and the population's carrying capacity – *a* is then simply a particular combination of these. An alternative point of view is that *a* is meaningful in its own right, measuring the per capita susceptibility to crowding: the larger the value of *a*, the greater the effect of density on the actual rate of increase in the population (Kuno, 1991). Now the behaviour of a population is seen as being jointly determined by two properties of the individuals within it – their intrinsic per capita rate of increase and their susceptibility to crowding, *R* and *a*. The carrying capacity of the population ($K = (R - 1)/a$) is then simply an outcome of these properties. The great advantage of this viewpoint is that it places individuals and populations in a more realistic biological perspective. Individuals come first: individual birth rates, death rates and susceptibilities to crowding are subject to natural selection and evolve. Populations simply follow: a population's carrying capacity is just one of many features that reflect the values these individual properties take.

properties of the simplest model

The properties of the model in Equation 5.12 may be seen in Figure 5.20a (which shows a model population increasing in size over time in conformity with Equation 5.12) and Figure 5.20b (from which the model was derived). The population in Figure 5.20a describes an S-shaped curve over time. As we saw earlier, this is a desirable quality of a model of intraspecific competition. But there are many other models that would also generate such a curve. The advantage of Equation 5.12 is its simplicity.

The behaviour of the model in the vicinity of the carrying capacity can best be seen by reference to Figure 5.20b. At population sizes that are less than *K* the population will increase in size; at population sizes that are greater than *K* the population size will decline; and at *K* itself the population neither increases nor decreases. The carrying capacity is therefore a stable equilibrium for the population, and the model exhibits the regulatory properties classically characteristic of intraspecific competition.

5.6.2 What type of competition?

But what type or range of competition is this model able to describe? We can answer this question by tracing the relationship between *k* values and log *N* (as in Section 5.3). Each generation, the potential number of individuals produced (i.e. the number that would be produced if there were no competition) is $N_t R$. The actual number produced (i.e. the number that survive the effects of competition) is $N_t R/(1 + aN_t)$.

We have seen that:

$$k = \log\,(\text{number produced}) - \log\,(\text{number surviving}). \tag{5.13}$$

Thus, in the present case:

$$k = \log N_t R - \log N_t R/(1 + aN_t), \tag{5.14}$$

or, simplifying:

$$k = \log\,(1 + aN_t). \tag{5.15}$$

Figure 5.21 shows a number of plots of k against $\log_{10} N_t$ with a variety of values of a inserted into the model. In every case, the slope of the graph approaches and then attains a value of 1. In other words, the density dependence always begins by undercompensating and then compensates perfectly at higher values of N_t. The model is therefore limited in the type of competition it can produce. So far, we can only say that this type of competition leads to very tightly controlled regulation of populations.

5.6.3 Time lags

One simple modification that we can make is to relax the assumption that populations respond instantaneously to changes in their own density, i.e. that present density determines the amount of resource available to a population and this in turn determines the net reproductive rate within the population. Suppose instead that the amount of resource available is determined by the density one time interval previously. For example, suppose that the amount of grass in a field in spring (the resource available to cattle) might be determined by the level of grazing (and hence, the density of cattle) in the previous year. In such a case, the reproductive rate itself will be dependent on the density one time interval ago. Thus, since in Equations 5.7 and 5.12:

$$N_{t+1} = N_t \times \text{reproductive rate}, \tag{5.16}$$

Equation 5.12 may be modified to:

$$N_{t+1} = \frac{N_t R}{1 + aN_{t-1}}. \tag{5.17}$$

time lags provoke population fluctuations

Now there is a time lag in the population's response to its own density, caused by a time lag in the response of its resources. The behaviour of the modified model is as follows:

$$R < 1.33 : \text{direct approach to a stable equilibrium}$$

$$R > 1.33 : \text{damped oscillations towards that equilibrium.}$$

In comparison, the original Equation 5.12, without a time lag, gave rise to a direct approach to its equilibrium for all values of R. The time lag has provoked the fluctuations in the model, and it can be assumed to have similar, destabilising effects on real populations.

5.6.4 Incorporating a range of competition

A further simple modification of Equation 5.12 allows us to incorporate a range of types of competition, as follows (Maynard Smith & Slatkin, 1973 ; Bellows, 1981):

$$N_{t+1} = \frac{N_t R}{1 + (aN_t)^b}. \tag{5.18}$$

We can see how this works from Figure 5.22, which plots k against $\log N_t$, as in Figure 5.17, but now k is $\log_{10}[1 + (aN_t)^b]$. The slope of the curve, instead of approaching 1 as it did previously, now approaches the value taken by b in Equation 5.18. Thus, by the choice of appropriate values, the model can portray undercompensation ($b < 1$), perfect compensation ($b = 1$), scramble-like overcompensation ($b > 1$) or even density independence ($b = 0$). This model has the generality that Equation 5.12 lacks, with the value of b determining the type of density dependence that is being incorporated.

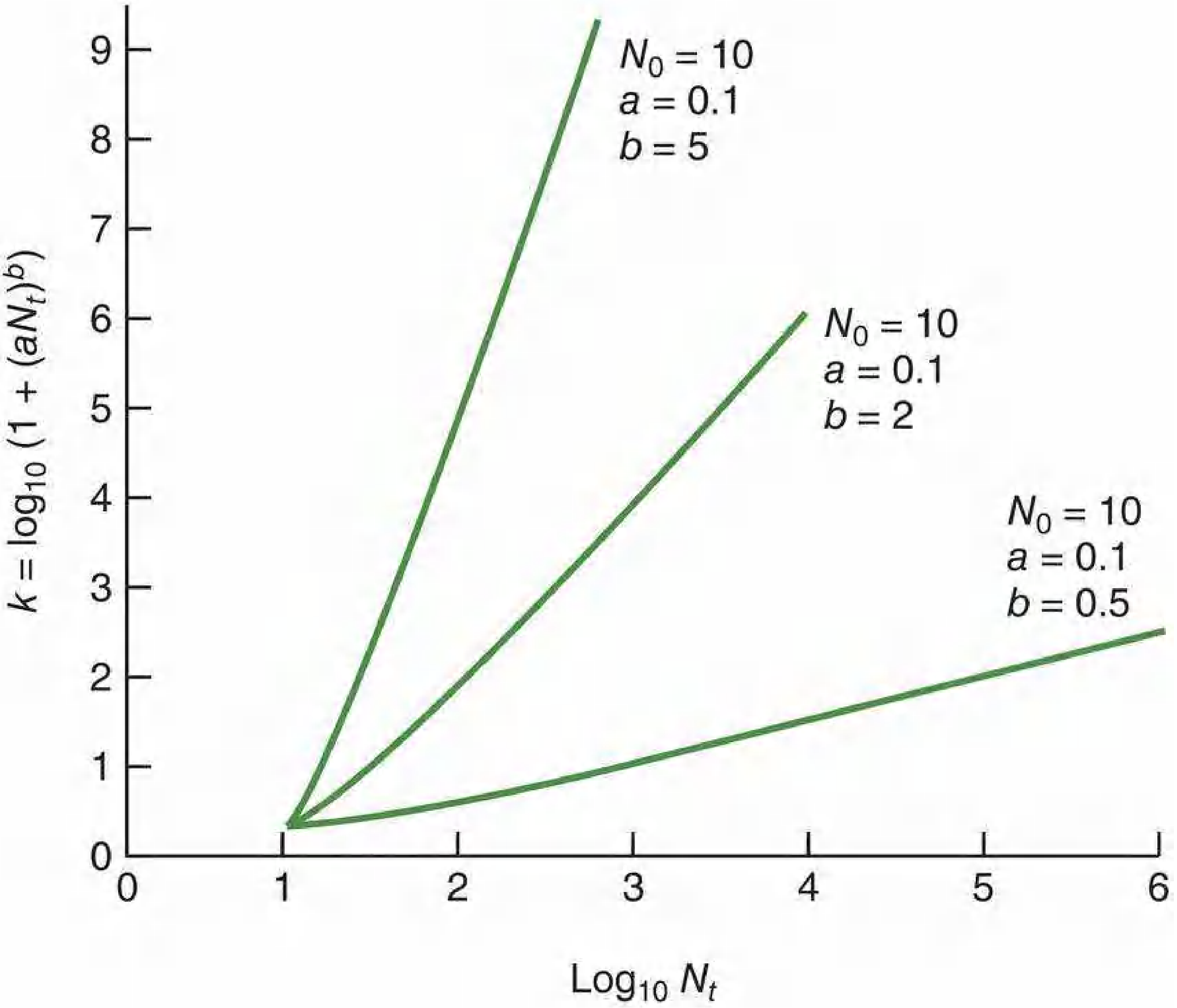


Figure 5.22 The intraspecific competition inherent in Equation 5.19. The final slope is equal to the value of *b* in the equation.

dynamic patterns: *R* and *b*

Equation 5.18 also shares with other good models an ability to throw fresh light on the real world. Analysing the population dynamics generated by the equation, we can draw guarded conclusions about the dynamics of natural populations. The mathematical method by which this and similar equations can be examined is described by May (1975a), but the results of the analysis (Figure 5.23) can be appreciated without dwelling on the analysis itself. Figure 5.23a sets out the conditions under which we get the various patterns of population growth and dynamics that Equation 5.18 can generate. Figure 5.23b shows what these patterns are. Note first that the pattern of dynamics depends on two things: (i) *b*, the precise type of competition or density dependence; and (ii) *R*, the effective net reproductive rate (taking density-independent mortality into account). By contrast, *a* determines not the pattern of fluctuation, but only the level about which any fluctuations occur.

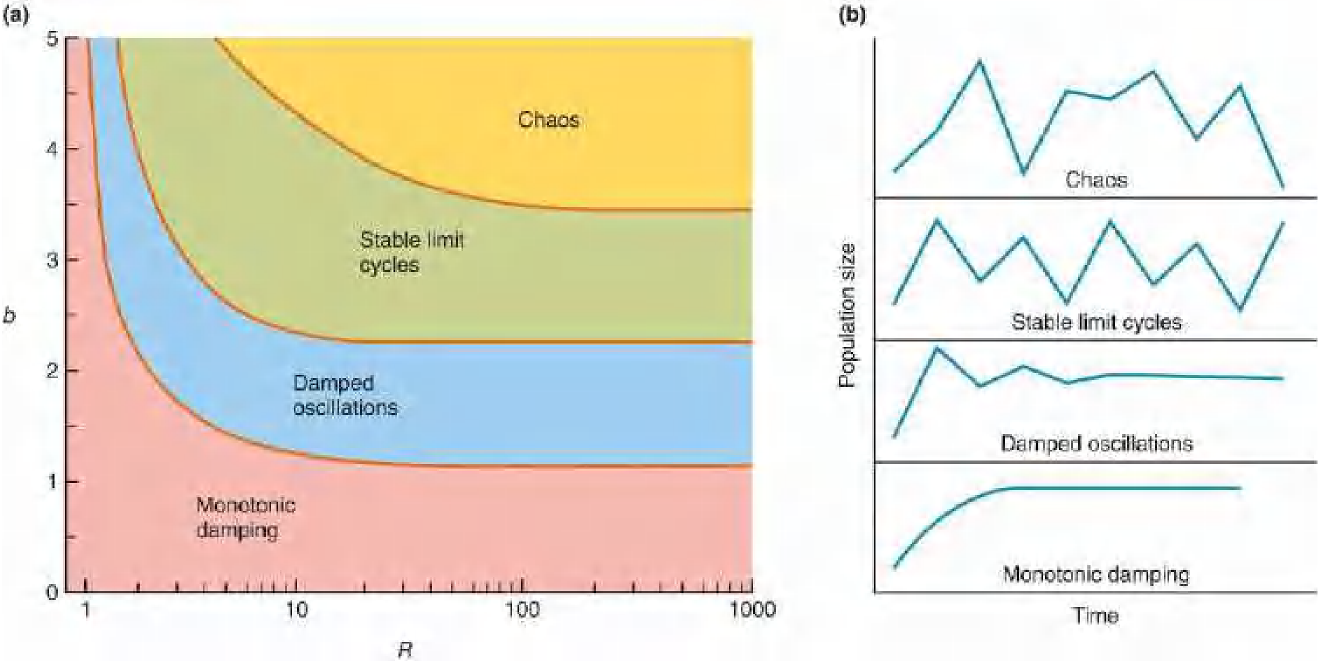


Figure 5.23 The range of population fluctuations generated by Equation5.19. (a) Reflecting the various possible combinations of *b* and *R*. (b) The patterns of those fluctuations.

Source: After May (1975a) and Bellows (1981).

As Figure 5.23a shows, low values of *b* and/or *R* lead to populations that approach their equilibrium size without fluctuating at all ('monotonic damping'). This has already been hinted at in Figure 5.20a. There, a population behaving in conformity with Equation 5.12 approached equilibrium directly, irrespective of the value of *R*. Equation 5.12 is a special case of Equation 5.18 in which $b = 1$ (perfect compensation). Figure 5.23a confirms that for $b = 1$, monotonic damping is the rule whatever the effective net reproductive rate.

As the values of *b* and/or *R* increase, the behaviour of the population changes first to damped oscillations gradually approaching equilibrium, and then to 'stable limit cycles' in which the population fluctuates around an equilibrium level, revisiting the same two, four or even more points time and time again. Finally, with large values of *b* and *R*, the population fluctuates in an apparently irregular and chaotic fashion.

5.6.5 Chaos

Thus, a model built around a density-dependent, supposedly regulatory process (intraspecific competition) can lead to a very wide range of population dynamics. If a model population has even a moderate fundamental net reproductive rate (and the ability to leave 100 (= *R*) offspring in the next generation in a competition-free environment is not unreasonable), and if it has a density-dependent reaction which even moderately overcompensates, then far from being stable, it may fluctuate widely in numbers without the action of any extrinsic factor. The biological significance of this is the strong suggestion that even in an environment that is wholly constant and predictable, the intrinsic qualities of a population and the individuals within it may, by themselves, give rise to population dynamics with large and perhaps even chaotic fluctuations. The consequences of intraspecific competition are clearly not limited to 'tightly controlled regulation'.

Two things are therefore clear. Firstly, time lags, high reproductive rates and overcompensating density dependence are capable (either alone or in combination) of producing all types of fluctuations in population density, without invoking any extrinsic cause. Secondly, and equally important, this has been made apparent by the analysis of mathematical models.

key characteristics of chaotic dynamics

In fact, the recognition that even simple ecological systems may contain the seeds of chaos led to chaos itself becoming a topic of interest amongst ecologists (Schaffer & Kot, 1986 ; Hastings *et al.*, 1993 ; Perry *et al.*, 2000). A detailed exposition of the nature of chaos is not appropriate here, but a few key points should be understood.

Firstly, the term 'chaos' may itself be misleading if it is taken to imply a fluctuation with absolutely no discernable pattern. Chaotic dynamics do not consist of a sequence of random numbers. On the contrary, there are tests (although they are not always easy to put into practice) designed to distinguish chaotic from random and other types of fluctuations. And since these patterns emerge from deterministic models (i.e. models with no random forces (stochasticity) incorporated), the term 'deterministic chaos' has been popularly used to describe them.

Secondly, fluctuations in chaotic ecological systems occur between definable upper and lower densities. Thus, in the model of intraspecific competition that we have discussed, the idea of 'regulation' has not been lost altogether, even in the chaotic region.

Third, however, unlike the behaviour of truly regulated systems, two similar population trajectories in a chaotic system will not tend to converge on ('be attracted to') the same equilibrium density or the same limit cycle (both of them 'simple' attractors). Rather, the behaviour of a chaotic system is governed by a '*strange* attractor'. Initially very similar trajectories become exponentially less and less like one another over time: chaotic systems exhibit 'extreme sensitivity to initial conditions'.

Hence, and finally, the long-term future behaviour of a chaotic system is effectively impossible to predict, and prediction becomes increasingly inaccurate as one moves further into the future. Even if we appear to have seen the system in a particular state before – and know precisely what happened subsequently last time – tiny (perhaps immeasurable) initial differences will be magnified progressively, and past experience will become of increasingly little value.

beyond chaos

This altered perspective opened up by an appreciation of chaos led initially to an excited optimism, but as Bjørnstad (2015) explains, ecologists grew steadily more sceptical. There were occasional demonstrations of apparent chaos in artificial laboratory environments (Costantino *et al.*, 1997), but the unlikelihood of natural systems being devoid of random forces and having very few interacting elements, combined with technical difficulties in detecting the signature of chaos in real datasets (Bjørnstad & Grenfell, 2001), led to questions of how often, if ever, ecological systems are chaotic, and whether, practically, we could detect chaos in nature even if it existed. The fashion for studying chaos has largely passed. Nonetheless, the altered perspective – that unpredictability may be inherent in ecological systems without the involvement of major perturbations from outside the system – has been far more enduring. Ecology must aim to become a predictive science. Chaotic systems, if they exist, would set us some of the sternest challenges in prediction. But unanticipated shifts in dynamics from one pattern to another, and the transient dynamics that may link them, set similar challenges that continue to be taken up. We return to these in Chapters 14 and 17.

5.6.6 Stochastic models

The models described so far have all been 'deterministic' models, that is, once the parameter values of the model have been specified (for example, Equation 5.7 with $N_t = 10$, $R = 1.1$) the outcome is definite, or 'determined'. Whenever that model is run with those values, the outcome is the same: after one time step, for example, there will be 11 individuals in the population. But the real world is not like that. The most that we could say of any population would be that, over each time step, there is a certain probability that there will be no births, a probability that there will be one birth, a probability that there will be no deaths, and so on – such that typically, or on average, 10 individuals will become 11 individuals over the course of a time step. The actual outcome, though, would reflect the consequences of those probabilities playing out: sometimes 11 individuals, but sometimes 10, or 12, or more rarely 9, or 13, etc. *Stochastic population models* incorporate these probabilistic processes. They are therefore more realistic, but also more unwieldy, more difficult to analyse, and for the non-specialist, more difficult to understand. Along related lines, *individual-based models* deal with these stochastic processes by explicitly acknowledging each individual in a population and giving those individuals their own chances of being born, dying, and in more complex models, moving or growing, and so on (Black & McKane, 2012).

We deal mostly with deterministic models throughout this book – because of their relative simplicity, and because for larger populations, the dynamics of deterministic and stochastic (or individual-based) models are hard to distinguish. Note, nonetheless, that for smaller populations, those models can behave very differently. Figure 5.24a, for example, shows a population reflecting a deterministic model like that in Equation 5.12, with an initial size of 3 and a carrying capacity of 25, and therefore exhibiting S-shaped growth as we saw in Figure 5.20a. Figure 5.24a also shows, however, three runs of an equivalent stochastic model. Two of those follow a similar path to their deterministic counterpart, albeit a more irregular one. But the third, far from approaching its notional carrying capacity, falls to extinction. The population, putting it colloquially, had a run of bad luck, but of course, such things do happen, whereas the population in the deterministic model is incapable of going extinct, or indeed of doing anything other than approaching its carrying capacity smoothly, and staying there.

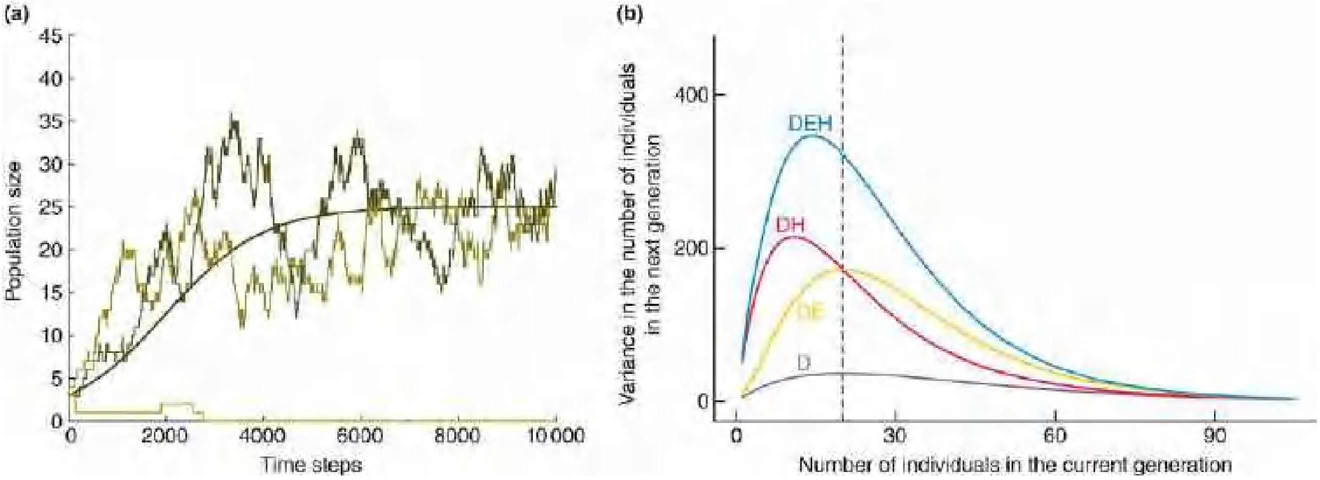


Figure 5.24 Populations in stochastic models may have a high chance of going extinct even when their deterministic counterparts are incapable of doing so. (a) The smooth line is the output of a deterministic model of population growth regulated by intraspecific competition, initiated with a population size of 3 and with a carrying capacity of 25. The irregular lines are outputs of three runs of an equivalent stochastic model. (b) The variance in the number of individuals in the next generation as a function of the number in the current generation – an established proxy for the estimated time to population extinction – for various stochastic models of population growth regulated by intraspecific competition, as indicated. D has stochastic birth and death only. DE and DEH also have environmental stochasticity. DH and DEH also have individual heterogeneity in birth and death probabilities. The vertical dashed line marks the carrying capacity, 20. The greater the variance, the shorter the expected time to extinction.

Source: (a) After Allen & Allen (2003). (b) After Melbourne & Hastings (2008).

stochastic models of population extinction

These differences between the dynamics of deterministic and stochastic models are arguably most important when we are focused on small populations that we wish to conserve (reviewed by Ovaskainen & Meerson (2010)). Deterministic models of regulated population growth, taken at face value, would tell us that a population was safe from extinction, however small it was, as long as it had a finite carrying capacity and positive value of *R*. Believing this would be dangerously complacent. By contrast, stochastic models, going back to early studies (e.g. Leigh, 1981) have estimated expected times to extinction and found these to increase exponentially with carrying capacity. This reflects the reality than even in a world of stochastic birth and death rates, large populations have a negligible chance of going extinct, but for small populations, extinction within a given time window may be close to inevitable. Moreover, if we relax the assumption that the environment itself is stable, and acknowledge instead that there is likely also to be environmental stochasticity as well as, for example, heterogeneity amongst individuals in their birth and death probabilities, then these further compound the effects of demographic stochasticity, greatly reducing the expected times to extinction (Figure 5.24b). We deal more fully with the conservation of small populations in Section 15.4.2, but it is clear that in order to judge the threats that they face, we shall often need to acknowledge all of the stochasticities affecting them, and to incorporate these into models that are themselves stochastic.

5.7 Continuous breeding: the logistic equation

The model derived and discussed in Section 5.6 was appropriate for populations that have discrete breeding seasons and that can therefore be described by equations growing in discrete steps, i.e. by 'difference' equations. Such models are not appropriate, however, for those populations in which birth and death are continuous. These are best described by models of continuous growth, or 'differential' equations, which we consider next.

r, the intrinsic rate of natural increase

The net rate of increase of such a population will be denoted by d*N*/d*t* (referred to in speech as 'd*N* by d*t*'). This represents the 'speed' at which a population increases in size, *N*, as time, *t*, progresses. The increase in size of the whole population is the sum of the contributions of the various individuals within it. Thus, the average rate of increase per individual, or the 'per capita rate of increase' is given by d*N*/d*t*(1/*N*). But we have already seen in Section 4.7 that in the absence of competition, this is the definition of the 'intrinsic rate of natural increase', *r*. Thus:

$$\frac{dN}{dt}\left(\frac{1}{N}\right) = r \qquad (5.19)$$

and:

$$\frac{dN}{dt} = rN. \qquad (5.20)$$

A population increasing in size under the influence of Equation 5.20, with $r > 0$, is shown in Figure 5.25. Not surprisingly, there is unlimited, 'exponential' increase. In fact, Equation 5.20 is the continuous form of the exponential difference Equation 5.8, and as discussed in Section 4.7, *r* is simply $\log_e R$. (Mathematically adept readers will see that Equation 5.20 can be obtained by differentiating Equation 5.8.) *R* and *r* are clearly measures of the same commodity: 'birth plus survival' or 'birth minus death'. The difference between *R* and *r* is merely a change of currency.

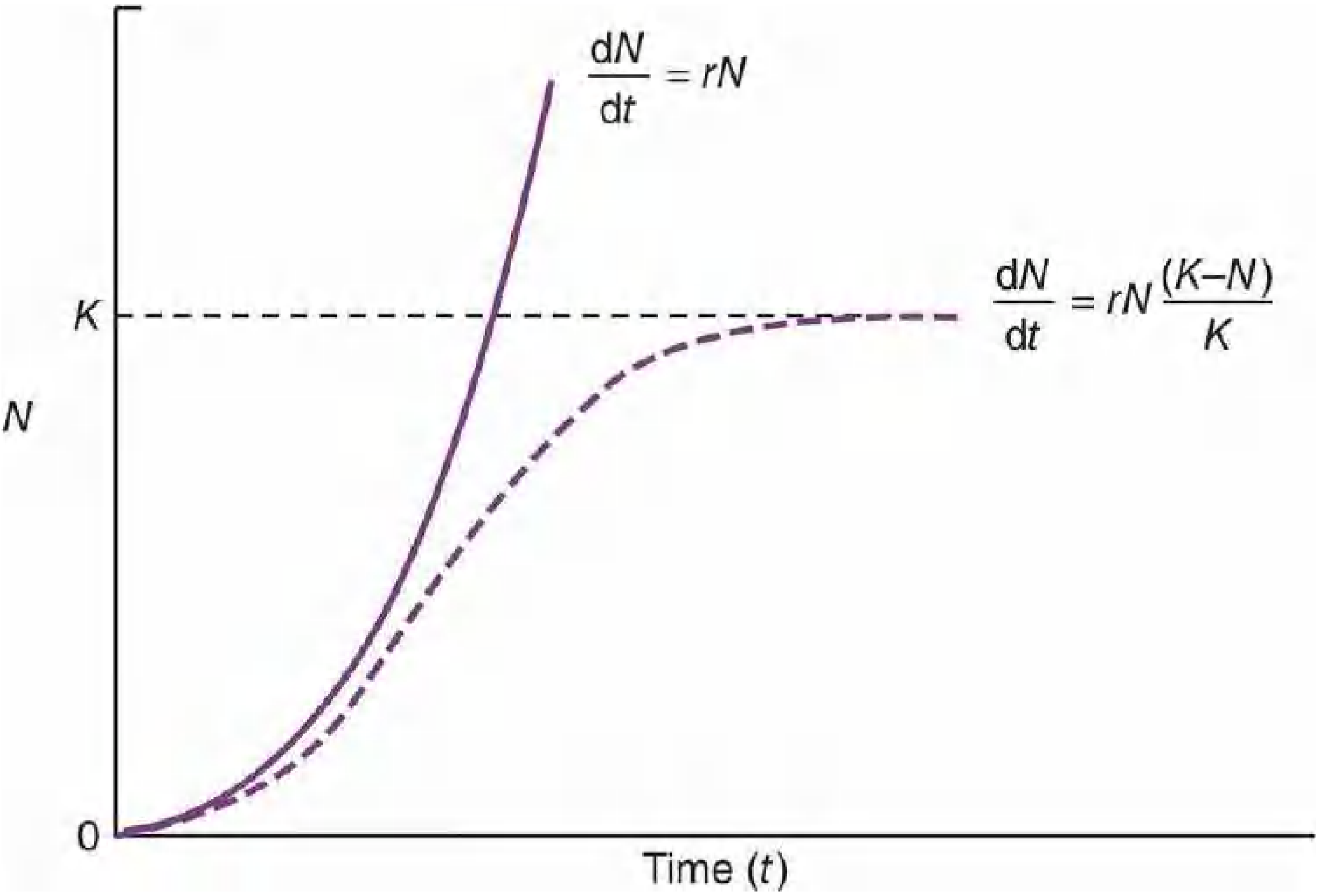


Figure 5.25 Exponential (solid line) and sigmoidal (dashed line) increase in density (N) with time for models of continuous breeding. The equation giving sigmoidal increase is the logistic equation.

the logistic equation

Intraspecific competition can be added to Equation 5.20 by a method exactly equivalent to the one used in Figure 5.18b, giving rise to:

$$\frac{dN}{dt} = rN\left(\frac{K-N}{K}\right). \tag{5.21}$$

This is known as the 'logistic' equation, and a population increasing in size under its influence is also shown in Figure 5.25.

The logistic equation is the continuous equivalent of Equation 5.12, and it therefore has all the essential characteristics of Equation 5.12, and all of its shortcomings. It describes a sigmoidal growth curve approaching a stable carrying capacity, but it is only one of many reasonable equations that do this. Its major advantage is its simplicity. Moreover, while it was possible to incorporate a range of competitive intensities into Equation 5.12, this is by no means easy with the logistic equation. The logistic is therefore doomed to be a model of perfectly compensating density dependence. Nevertheless, in spite of these limitations, the equation will be an integral component of models in Chapters 8 and 10, and it has played a central role in the development of ecology.

5.8 Individual differences: asymmetric competition

5.8.1 Size inequalities

Until now, we have focused on what happens to the whole population or the average individual within it. Different individuals, however, may respond to intraspecific competition in very different ways. For example, when larval salamanders (*Ambystoma tigrinum nebulosum*) were competed amongst one another in groups, the sizes of the largest surviving larvae were no different from those reared alone ($P = 0.42$) but the smallest larvae were much smaller ($P < 0.0001$) (Ziemba & Collins, 1999). Similarly, the overwinter survival of red deer calves, *Cervus elaphus*, in the resource-limited population on the island of Rhum, Scotland, declined sharply as the population became more crowded, but those that were smallest at birth were by far the most likely to die (Clutton-Brock *et al.*, 1987). The effects of competition are far from being the same for every individual. Weak competitors may make only a small contribution to the next generation, or no contribution at all. Strong competitors may have their contribution only negligibly affected.

What effect does this have at the population level? Figure 5.26 shows the results of a classic experiment in which flax (*Linum usitatissimum*) was sown at three densities, and harvested at three stages of development, recording the weight of each plant individually. This made it possible to monitor the effects of increasing amounts of competition not only as a result of variations in initial density, but also as a result of plant growth (between the first and the last harvests). When intraspecific competition was at its least intense (at the lowest sowing density after only two weeks' growth) the individual plant weights were distributed symmetrically about the mean. When competition was at its most intense, however, the distribution was strongly skewed to the left: there were many very small individuals and a few large ones. As the intensity of competition gradually increased, the degree of skewness increased as well. Something very similar seems to be happening with the data in Figure 5.27 for cod (*Gadus morhua*) living off the coast of Norway. At higher densities (and presumably greater intensities of competition) size decreased but the skewness in the distribution of sizes increased.

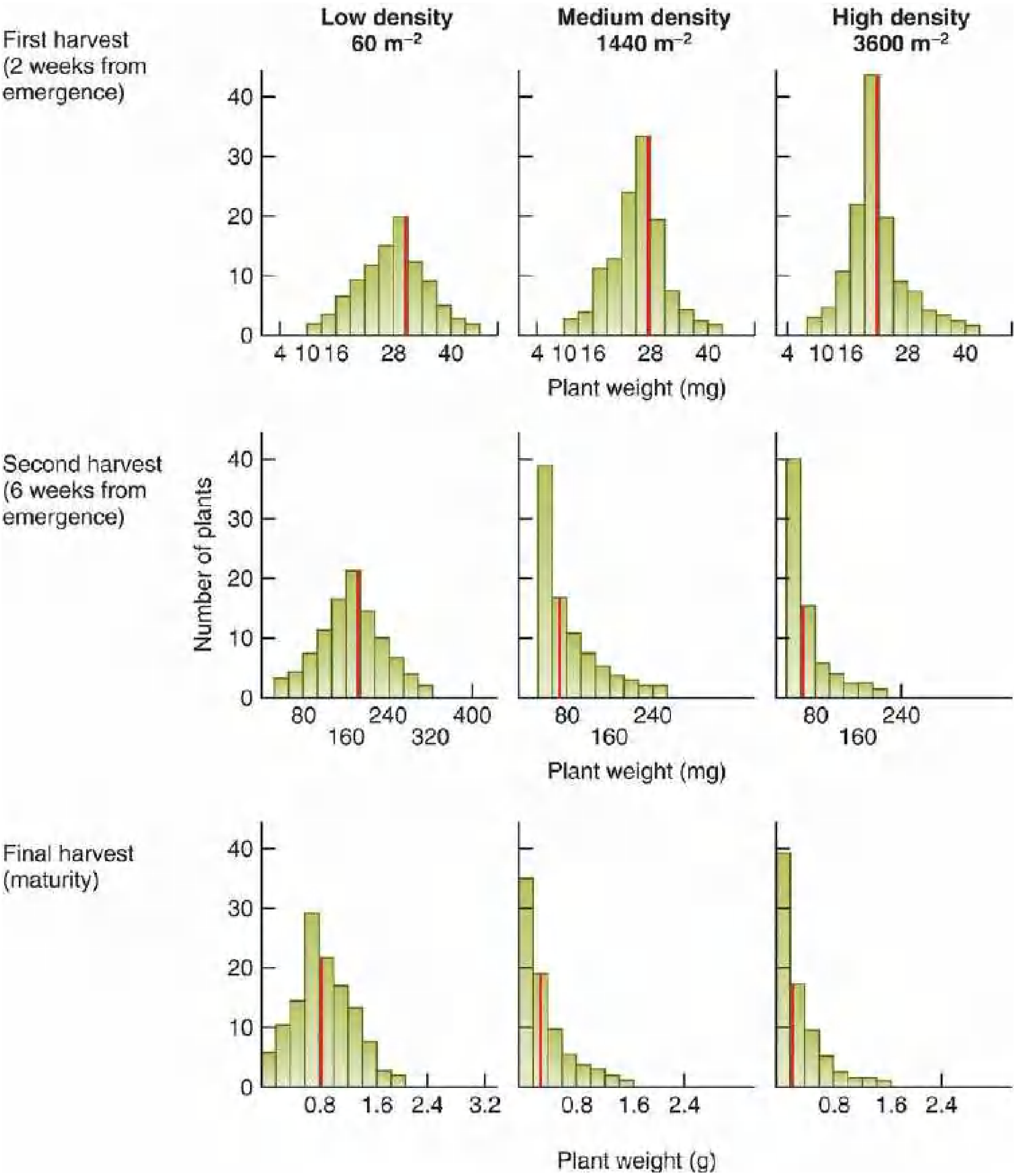


Figure 5.26 Intraspecific competition increases the skewing in the distribution of plant weights. Frequency distributions of individual plant weights in populations of flax (*Linum usitatissimum*), sown at three densities and harvested at three ages. The red bar is the mean weight.

Source: After Obeid *et al*. (1967).

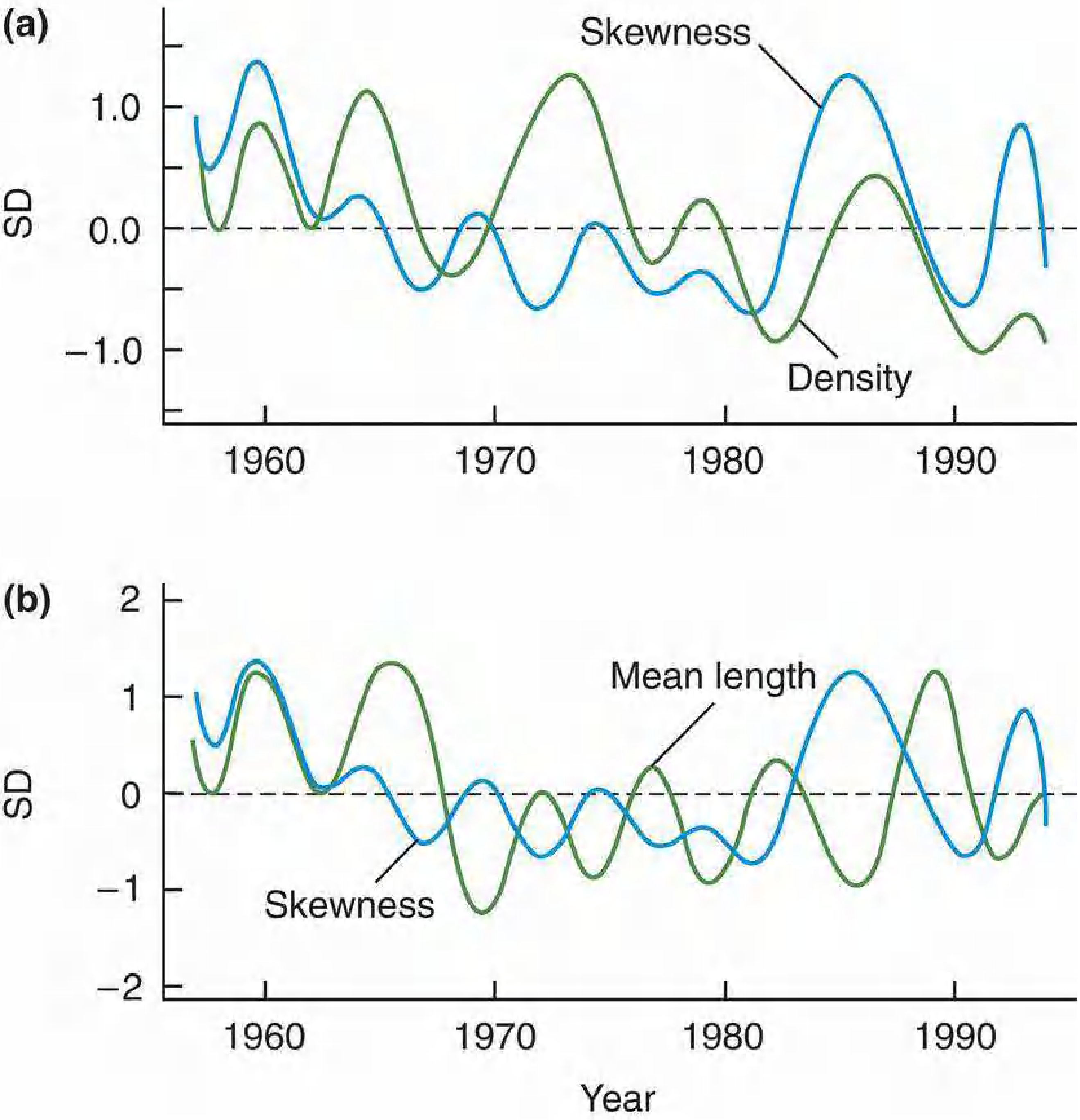


Figure 5.27 Intraspecific competition increases the skewing in the distribution of cod lengths but decreases mean length. (a) Values of density and of skewness in the frequency distribution of lengths, both expressed as standard deviations from mean values, for the years 1957–94 for cod (*Gadus morhua*) from the Skagerrak, off the coast of Norway. (b) Similar patterns for skewness and mean length. Despite marked fluctuations from year to year, much of it the result of variations in weather, when density rose, that is, when competition was most intense, skewness tended also to rise ($r = 0.58$, $P < 0.01$) but mean length tended to decline ($r = -0.45$, $P < 0.05$).

Source: After Lekve *et al.* (2002).

the inadequacy of the average

Skewness aside, we can say more generally that increased competition tends to increase the degree of size inequality within a population, at least initially (Weiner, 1990). Characterising a population by an arbitrary 'average' individual can therefore be very misleading under such circumstances, and can divert attention from the fact that intraspecific competition is a force affecting individuals, even though its effects may often be detected in whole populations.

5.8.2 The generation and dilution of size inequalities

The processes by which these size inequalities may be generated, and also ultimately diminished, are illustrated by a study of competition in two separate species of large brown seaweed, *Laminaria digitata* and *Fucus serratus*, off the Isle of Man, UK, reared at a range of densities, from 650 to 5156 plants per m^2 (Figure 5.28). We can see, firstly, that for all populations there was, for a large part of the experiment, a marked tendency for size inequality to increase (Figure 5.28a). The Gini coefficient (a measure of inequality originally developed by economists to capture inequalities in wealth), applied to frond (leaf) lengths, increased steadily as the plants grew, that is, as mean mass increased and competition intensified. However, for many of the populations, as the plants grew further still, the Gini coefficient declined again.

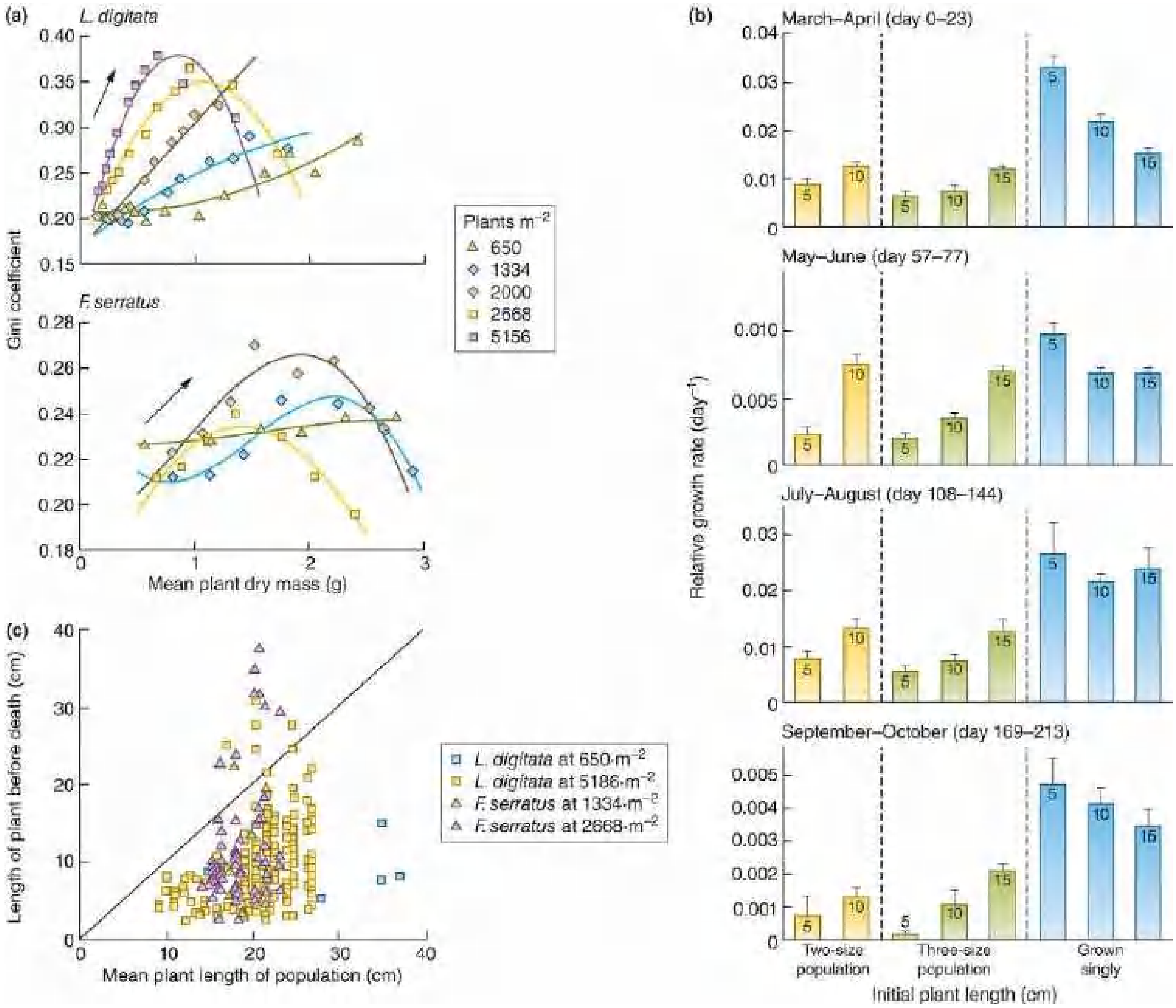


Figure 5.28 Size inequalities first increase then decrease in competing populations of seaweeds. (a) The effect of density on the relationship between inequality of plant frond lengths (measured by Gini coefficients) and mean plant mass in populations of *Laminaria digitata* (above) and *Fucus serratus* (below) at the range of densities indicated. Data are means ($n = 4$). Curved lines are fitted by polynomial functions to indicate the general pattern of change between variables. Arrows indicate the direction of the time sequence. (b) The relative growth rates of *L. digitata* plants grown singly and in populations consisting of two or three sizes of plants over four selected time periods. Data are means of $n = 6$ singly grown plants and $n = 3$ population means for portions of populations. Bars are numbered to indicate initial plant size (in cm). Bars are SEs. (c) The frond length of *L. digitata* and *F. serratus* plants at the sample time before death, plotted against the population mean frond length at that time. Two densities for each species are plotted to demonstrate the generality of the relationships. The diagonal line represents the condition where the length of the dying plants equals the population mean plant length.

Source: After Creed *et al.* (1998).

The process driving the initial increases in inequality is illustrated in Figure 5.28b. This shows data on the relative growth rates of *L. digitata* plants of different initial lengths (5 cm, 10 cm and 15 cm). Relative growth rates measure changes in mass relative to initial mass, which makes sense since without this correction, large plants would almost inevitably grow faster than small ones. The plants were combined in populations of different size-composition but at a high density (7619 plants m^{-2}) where competition would have been intense. Plants were also grown in isolation. When grown singly, the smallest plants grew fastest, at least initially. However, as soon as they were combined, either with two or three size-classes together, the growth-rate differentials were reversed, with the smallest plants growing least and the largest most. Indeed, in some cases, the growth of the larger plants was barely affected by competition. Size inequality increased, therefore, because plants that were smaller initially were more affected by their neighbours. Small initial differences were transformed by competition into much larger differences over time. Competition amongst the seaweeds was therefore asymmetric: there was a hierarchy. It seems that the larger plants pre-empted or 'captured' space and resources, and subsequently were little affected by intraspecific competition, while the smaller plants were being starved of these resources. Berntson and Wayne (2000) provide a rare empirical confirmation of such size-uptake relationships for developing stands of birch seedlings (*Betula alleghaniensis*).

On the other hand, later in the main experiment (Figure 5.28a) size inequality *de*creased in many cases. The reason is apparent in Figure 5.28c, which shows that the plants that died in the populations, ultimately, were the smallest ones. Thus, while size-dependent growth-differentials increased the size inequalities, size-dependent survival-differentials decreased them again by cutting off the tail of the size distribution. The modular nature of plants makes this separation of processes especially likely. There can be an extended period over which the smaller, weaker competitors are stunted in their growth but do not die, and only later, after the inequalities have been exaggerated, do differentials in survival reverse this. For most (unitary) animals, stunting is not an option and the weakest competitors die far sooner. As we have seen, however, patterns in animals like those seen in plants, while perhaps rarer, are certainly not unknown.

roots and shoots: the strength and asymmetry of competition

If competition is asymmetric because superior competitors pre-empt resources, then competition is most likely to be asymmetric when it occurs for resources that are most liable to be pre-empted. Specifically, competition amongst plants for light, in which a superior competitor can overtop and shade an inferior, might be expected to lend itself far more readily to pre-emptive resource capture than competition for soil nutrients or water, where the roots of even a very inferior competitor will have more immediate access to at least some of the available resources than the roots of its superiors. This expectation is borne out by the results of an experiment in which morning glory vines (*Ipomoea tricolor*) were grown as single plants in pots ('no competition'), as several plants rooted in their own pots but with their stems intertwined on a single stake ('shoots competing'), as several plants rooted in the same pot, but with their stems growing up their own stakes ('roots competing') and as several plants rooted in the same pot with their stems intertwined on one stake ('shoots and roots competing') (Figure 5.29). Root competition was more intense than shoot competition, in the sense that it led to a far greater decrease in the mean weight of individual plants. However, it was shoot competition for light that led to a much greater increase in size inequality and determined which individuals were dominant – a general conclusion drawn by Kiær *et al.* (2013) in their meta-analysis of root and shoot competition.

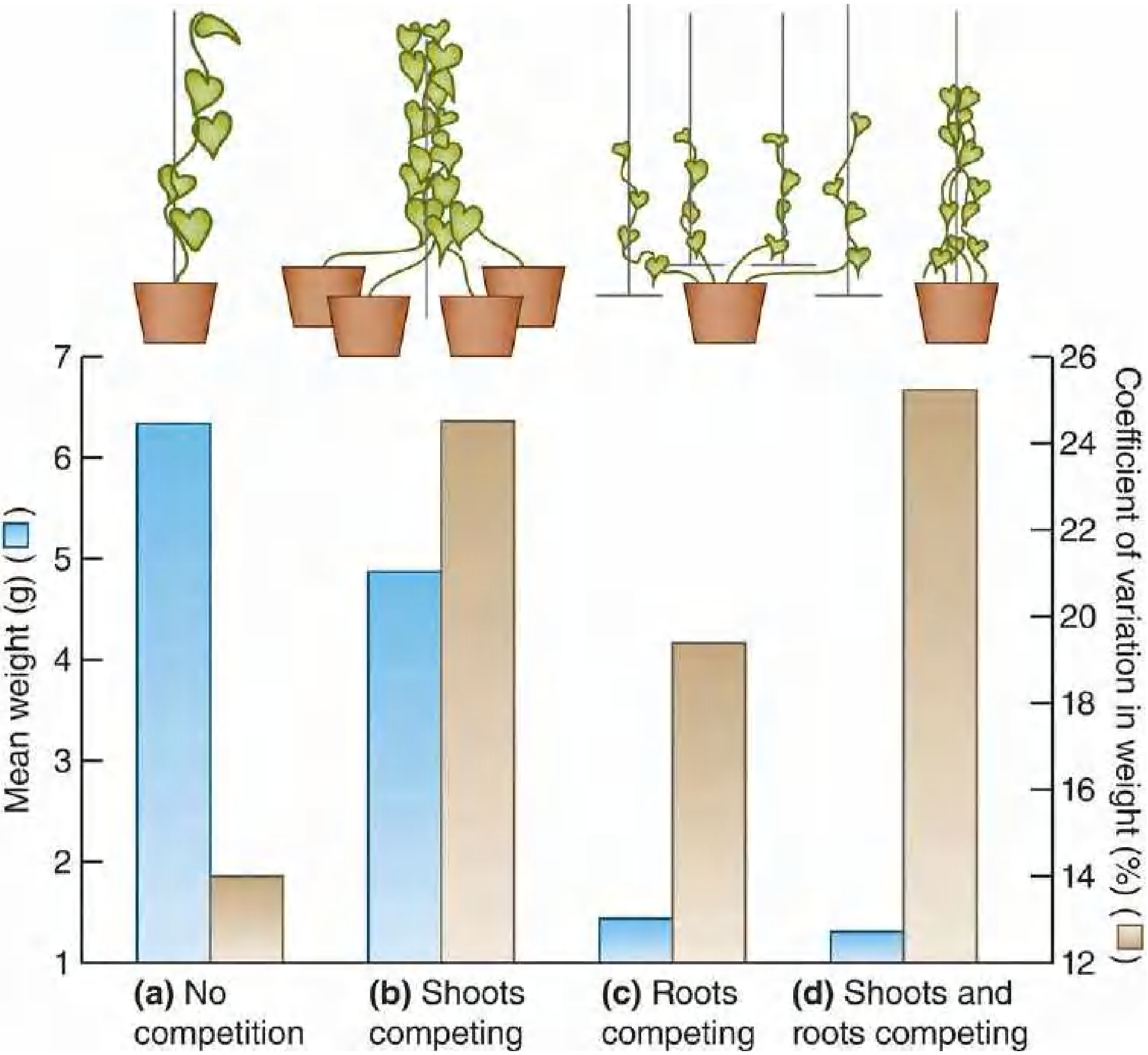


Figure 5.29 Root and shoot competition can have contrasting effects on mean size and size inequalities. When morning glory vines competed, root competition was most effective in reducing mean plant weight (treatments significantly different, $P < 0.01$, for all comparisons except (c) with (d)), but shoot competition was most effective in increasing the degree of size inequality, as measured by the coefficient of variation in weight (significant differences between treatments (a) and (b), $P < 0.05$, and (a) and (d), $P < 0.01$).

Source: After Weiner (1986).

5.8.3 Asymmetry enhances regulation

One final, important general point is that these asymmetries tend to reinforce the regulatory powers of intraspecific competition, in much the same way as exactly-compensating or 'contest' competition (Section 5.3) gives rise to the most tightly regulated populations. This can be seen especially in a much longer term study that will almost certainly never be repeated, where a population of the herbaceous perennial, *Anemone hepatica*, in Sweden was visited and revisited every year from 1943 to 1956, with each plant being tracked individually (Figure 5.30; Tamm, 1956). Crops of seedlings entered the population each year between 1943 and 1952, but nonetheless, the most important factor determining which individuals survived to 1956 was whether or not they were established in 1943. Of the 30 individuals that were already of a large or intermediate size in 1943, 28 survived until 1956, and some of these had branched. By contrast, of the 112 plants that were either small in 1943 or appeared as seedlings subsequently, only 26 survived to 1956, and not one of these was sufficiently well established to have flowered. Tamm's established plants were successful competitors – winners in a contest – year after year, but his small plants and seedlings were repeatedly unsuccessful. This guaranteed a near constancy in the number of established, winning plants between 1943 and 1956, accompanied by a variable number of 'losers' that not only failed to grow, but usually, in due course, died. Similar patterns can be observed in tree populations. The survival rates, the birth rates and thus the fitnesses of the few established adults are high; those of the many seedlings and saplings are low.

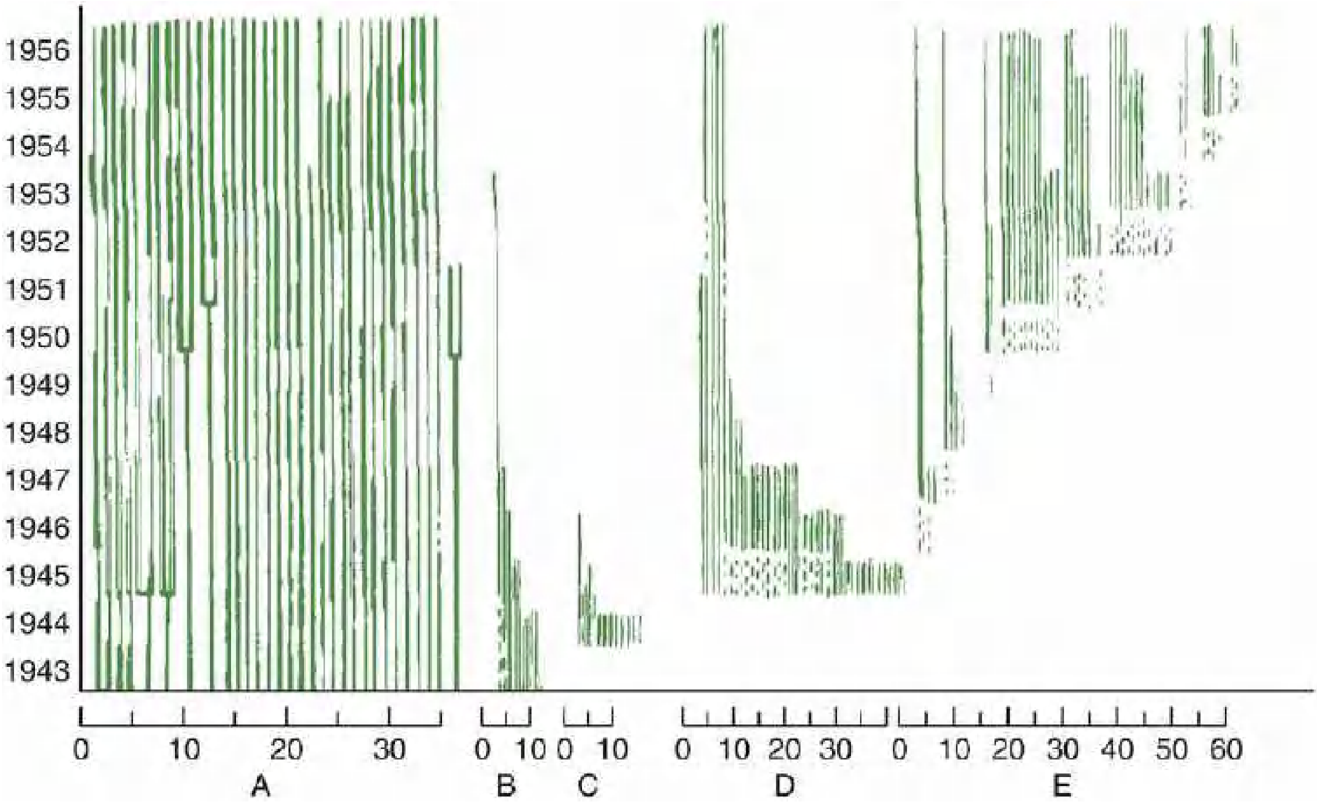


Figure 5.30 Asymmetric competition enhances population size regulation in an anemone. Space pre-emption in a perennial, *Anemone hepatica*, in a Swedish forest. Each line represents one individual: straight for unramified ones, branched where the plant has ramified (i.e. produced new plants by vegetative growth), bold where the plant flowered and broken where the plant was not seen that year. Group A were alive and large in 1943, group B alive and small in 1943, group C appeared first in 1944, group D in 1945 and group E thereafter, presumably from seedlings.

Source: After Tamm (1956).

5.8.4 Territoriality

Territoriality refers to active interference between individuals, such that a more or less exclusive area, the territory, is defended against intruders by a recognizable pattern of behaviour. It is one widespread and important example of asymmetric intraspecific competition.

territoriality is a contest and regulates population size

Individuals of a territorial species that fail to obtain a territory, commonly referred to as 'floaters', often make no contribution whatsoever to future generations. Territoriality, therefore, is a 'contest'. There are winners (those that come to hold a territory) and losers (those that do not), and at any one time there can be only a limited number of winners. The exact number of territories (winners) in a population may vary, depending on resources or environmental conditions. We see this, for example, for the number of tawny owl (*Strix aluco*) territories in a Finnish population, and especially the number of productive territories, which varies with the abundance of their vole prey (Figure 5.31). Nonetheless, the contest nature of territoriality ensures, like asymmetric competition generally, a comparative constancy in the number of surviving, reproducing individuals. In the past, some felt that these regulatory consequences of territoriality must themselves be the root cause underlying the evolution of territorial behaviour – territoriality being favoured because the population as a whole benefitted from the rationing effects, which guaranteed that the population did not overexploit its resources (e.g. Wynne-Edwards, 1962). However, there are powerful and fundamental reasons for rejecting this 'group selectionist' explanation (essentially, it stretches evolutionary theory beyond reasonable limits): the ultimate cause of territoriality must be sought within the realms of natural selection, in some advantage accruing to the individuals concerned.

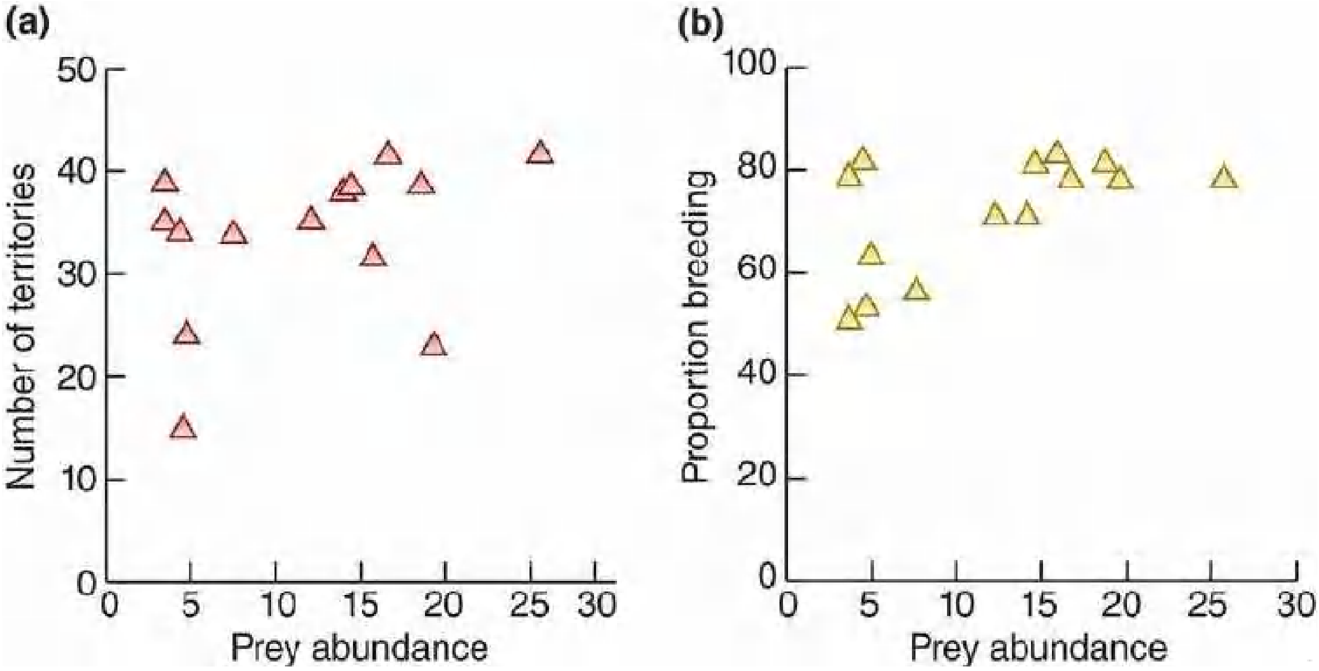


Figure 5.31 The number of (successful) territories may increase at higher resource levels. (a) The number of territories occupied by tawny owls, *Strix aluco*, in a study in Finland tends to increase as the abundance of their vole prey increases, though not significantly, but (b) the proportion of these in which territory holders are successful in breeding increases significantly with prey abundance ($P = 0.007$).

Source: After Karell *et al*. (2009).

benefits and costs of territoriality

Any benefit that an individual does gain from territoriality, of course, must be set against the costs of defending the territory. In some animals this defence involves fierce combat between competitors, while in others there is a more subtle mutual recognition by competitors of one another's keep-out signals (e.g. song or scent). Even when the chances of physical injury are minimal, territorial animals typically expend energy in marking, patrolling and advertising their territories. Hence, for territoriality to be favoured by natural selection, we expect the benefits to exceed these costs. Indeed, we expect territories to be an optimum size – 'as large as necessary but as small as possible' (Kittle *et al*., 2015) – in that as territory size increases, defence costs will also increase but resource access is increasingly likely to exceed the animals' requirements. This in turn suggests that as habitat quality (and resource 'concentration') increases, territory sizes should decrease (they do not need to be as large), which is precisely what we see for the lion territories in Figure 5.32.

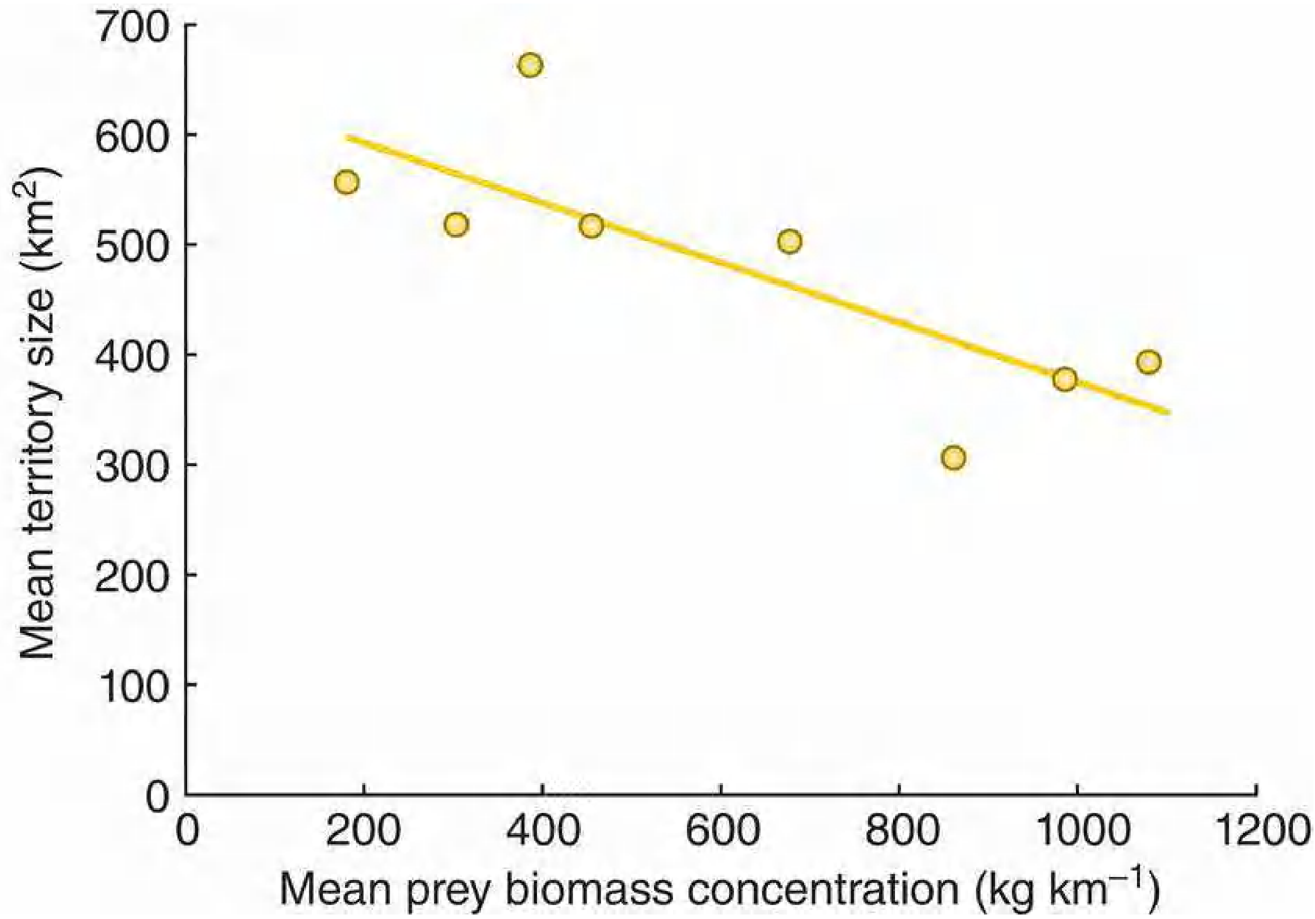


Figure 5.32 Territory sizes occupied by male lions (*Panthera leo*) in Zimbabwe decreased as habitat quality increased. Habitat quality was equated with the estimated concentration of the biomass of a wide range of prey species.

Source: After Loveridge *et al.* (2015).

floater? or territory holder?

On the other hand, explaining territoriality only in terms of a net benefit to the territory owner is rather like history always being written by the victors. A trickier question, which has received less attention, asks to what extent floaters, without a territory, are simply losers or are themselves net beneficiaries of a strategy they have adopted. Three main hypotheses, not mutually exclusive and all receiving some support, have been proposed for what determines whether an individual, as a result of a contest, becomes a floater or a territory holder: (i) territory holders are in some sense 'stronger', (ii) they have skills that they have acquired with age, and (iii) they are winners because they have acquired familiarity, as holders of the territory already (Sergio *et al.*, 2009). Variants of these hypotheses also include there being a 'convention', respected by the participants, through which the larger animal or the territory holder *always* wins. There is no doubt that some floaters are simply losers: too small, too young or too lacking in local knowledge, and therefore *en route* to dying without making any contribution to future generations. However, we see in Figure 5.33, for example, that for black kites, *Milvus migrans*, in the Doñana National Park in Spain, where floaters are neither smaller, nor in worse condition, and there is no 'resident always wins' rule, floaters are predominantly the youngest individuals, and territory holders are all members of the older age classes. Hence, for many of these floaters, it is not so much a case of reproduction-denied as reproduction-deferred. It therefore makes evolutionary sense for the young floaters to invest little in territorial contests in their earliest years (and in that sense ‘accept’ their floater status, temporarily) so as to maximise their chances of surviving long enough to win a territory later.

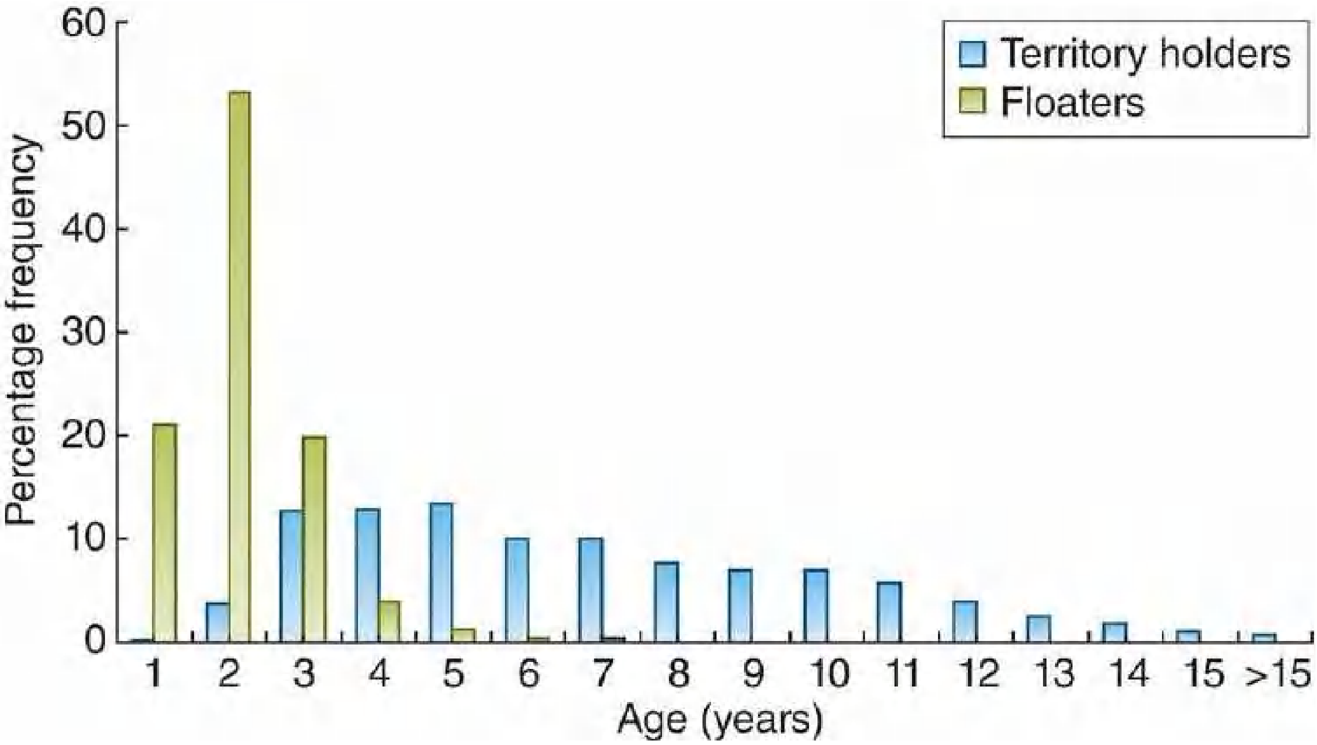


Figure 5.33 Older individuals hold the territories in a black kite population. The age profile of territory holders and non-territory holding floaters in a black kite, *Milvus migrans*, population in the Doñana National Park in Spain.

Source: After Sergio *et al.* (2009).

dear enemies and nasty neighbours

The idea that territorial strategies will be favoured that minimise costs to the territory holders also implies that the territory holders should, where possible, tailor their level of effort to the level of threat being posed. This has led to two contrasting hypotheses. The ‘dear enemy’ hypothesis proposes that more effort should be exerted against strangers unfamiliar to the territory holder, lacking territories themselves, than against territory-holding neighbours, since, once a territorial boundary has been established, it pays both neighbours to minimise their investment in maintaining it (Fisher, 1954). But on the other hand, the ‘nasty neighbour’ hypothesis proposes that more aggression should be displayed against neighbours than strangers, and proposes it especially for group-living species, where out-competing your neighbour allows your group to swell in size (Temeles, 1994). There is evidence for both (Figure 5.34), and while the dear enemy effect seen for the rodents in Figure 5.34a appears to be the more common, the nasty neighbour effect does indeed appear to be most often found in group-living species, such as the ants in Figure 5.34b.

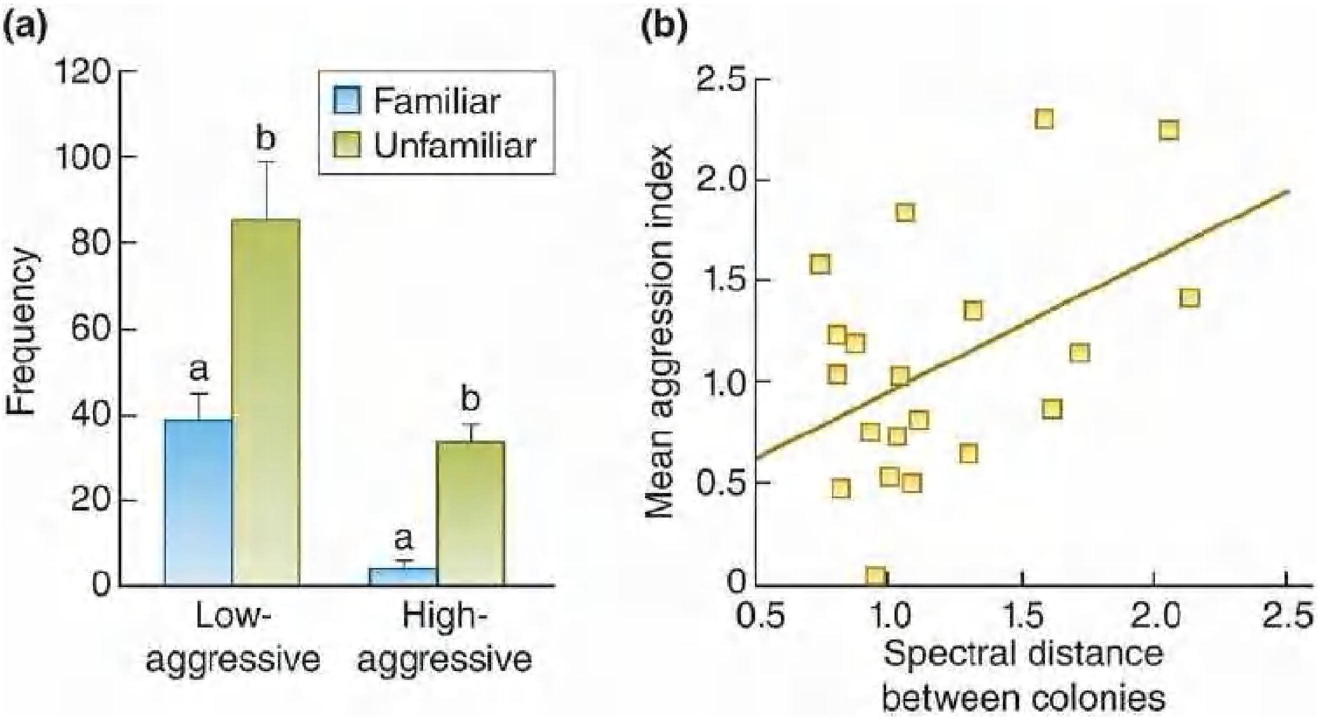
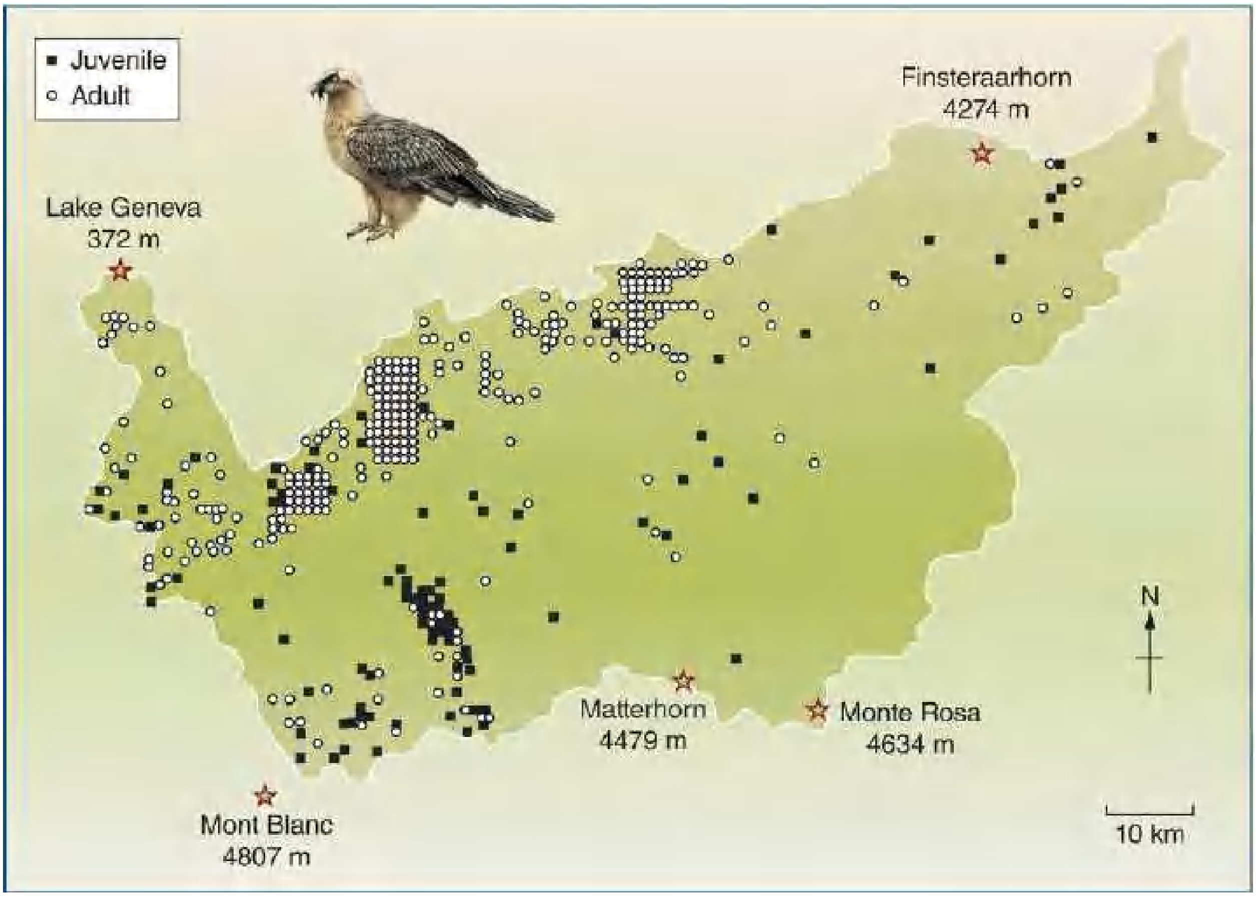


Figure 5.34 'Dear enemy' and 'nasty neighbour' effects. (a) Male subterranean rodents, *Ctenomys talarum* (tuco-tucos) in Argentina display more aggressive (especially high-aggressive) behaviour towards unfamiliar than towards familiar opponents in staged contest. Familiarity was gained by previously exposing contestants to the odour of their opponents. In the 'unfamiliar' treatment, exposure was to an odour of an animal other than the opponent. Bars are SEs; different letters denote significant differences ($P < 0.05$). (b) Weaver ant colonies, *Oecophylla smaragdina*, in Queensland, Australia, behaved more aggressively towards other colonies the less related they were to them in terms of the chemicals in their cuticle (less related = greater 'spectral distance').

Source: (a) After Zenuto (2010). (b) After Newey *et al.* (2010).

APPLICATION 5.3 Reintroduction of territorial vultures

Having seen that many species compete for territories related to the availability of resources rather than for the resources themselves, it is perhaps not surprising that when we come to manage such species, ensuring the availability of territories is a top priority. A good example comes from a study of bearded vultures, *Gypaetus barbatus*, which became extinct in the European Alps more than a century ago, and have been the focus of a reintroduction programme since 1986 (Figure 5.35). Captive-reared individuals were released from four widely dispersed sites from which they spread to new areas, and this spread was monitored in the Valais region of Switzerland (not one of the release sites). During an initial phase, from 1987 to 1994, the sightings were of subadults, and the most important factor explaining the distribution of these sightings was the biomass of ibex, *Capra ibex*, whose carcasses are an important resource for the vultures. However, during the subsequent phase, from 1995 to 2001, when adults were finally settling in the region, the presence of the vultures was most closely correlated with the distribution of craggy limestone crags, which are the ideal base for their territories, providing nest sites, thermal conditions for soaring, and limestone screes for bone breaking and food storage. Food availability was of only secondary significance. It seems clear, therefore, that future reintroductions in the area should focus precisely on the availability of these viable territories.



[Figure 5.35](#) **The importance of good territories for the conservation of bearded vultures.** Map of the Valais region of Switzerland, where bearded vultures, *Gypaetus barbatus*, have spread following their reintroduction. Black squares are 1 km squares where juvenile vultures were sighted during an initial, 'prospecting' stage (1987–94). White circles are 1 km squares where adult vultures were sighted during a subsequent, 'settling' stage (1995–2001).

Source: After Hirzel *et al*. (2004).

5.9 Self-thinning

We have seen throughout this chapter that intraspecific competition can influence the number of deaths, the number of births and the amount of growth within a population. We have illustrated this largely by looking at the end results of competition. But in practice the effects are often progressive. As a cohort ages, the individuals grow in size, their requirements increase, and they therefore compete at a greater and greater intensity. This in turn tends to increase their risk of dying. But if some individuals die, then the density is decreased as is the intensity of competition – which affects growth, which affects competition, which affects survival, which affects density, and so on.

In trying to understand these interconnected processes it is important to distinguish three types of study: (i) those in which the 'final' performance of competitors is monitored over a range of densities and hence over a range of intensities of competition; (ii) those in which density and performance are monitored together over time as groups of competitors grow and undergo density-dependent mortality; and (iii) those which seek relationships between density and performance in sets of populations, each observed just once (Weiner & Freckleton, 2010). Each type of study involves density and the performance of either individual competitors or the whole population, but the three tend to be aimed at addressing rather different questions. We examined the first in Section 5.2.2 when we discussed constant final yield. We turn here to the second and third.

5.9.1 Dynamic thinning lines

Starting with the second, the patterns that emerge in growing, crowded cohorts of individuals were originally the focus of particular attention in plant populations. For example, perennial rye grass (*Lolium perenne*) was sown at a range of densities, and samples from each density were harvested after 14, 35, 76, 104 and 146 days (Figure 5.36a). Figure 5.36a has the same logarithmic axes – density and mean plant weight – as Figure 5.7: what we referred to previously as a type (i) study. In Figure 5.7, each line represented a separate yield–density relationship at different ages of a cohort, and the points along a line represented different initial sowing densities. In Figure 5.36, on the other hand, each line itself represents a different initial density and successive points along a line represent populations at different ages. Each line is therefore a trajectory that follows a cohort through time, as indicated by the arrows in Figure 5.36, pointing from many small, young individuals (bottom right) to fewer, larger, older individuals (top left).

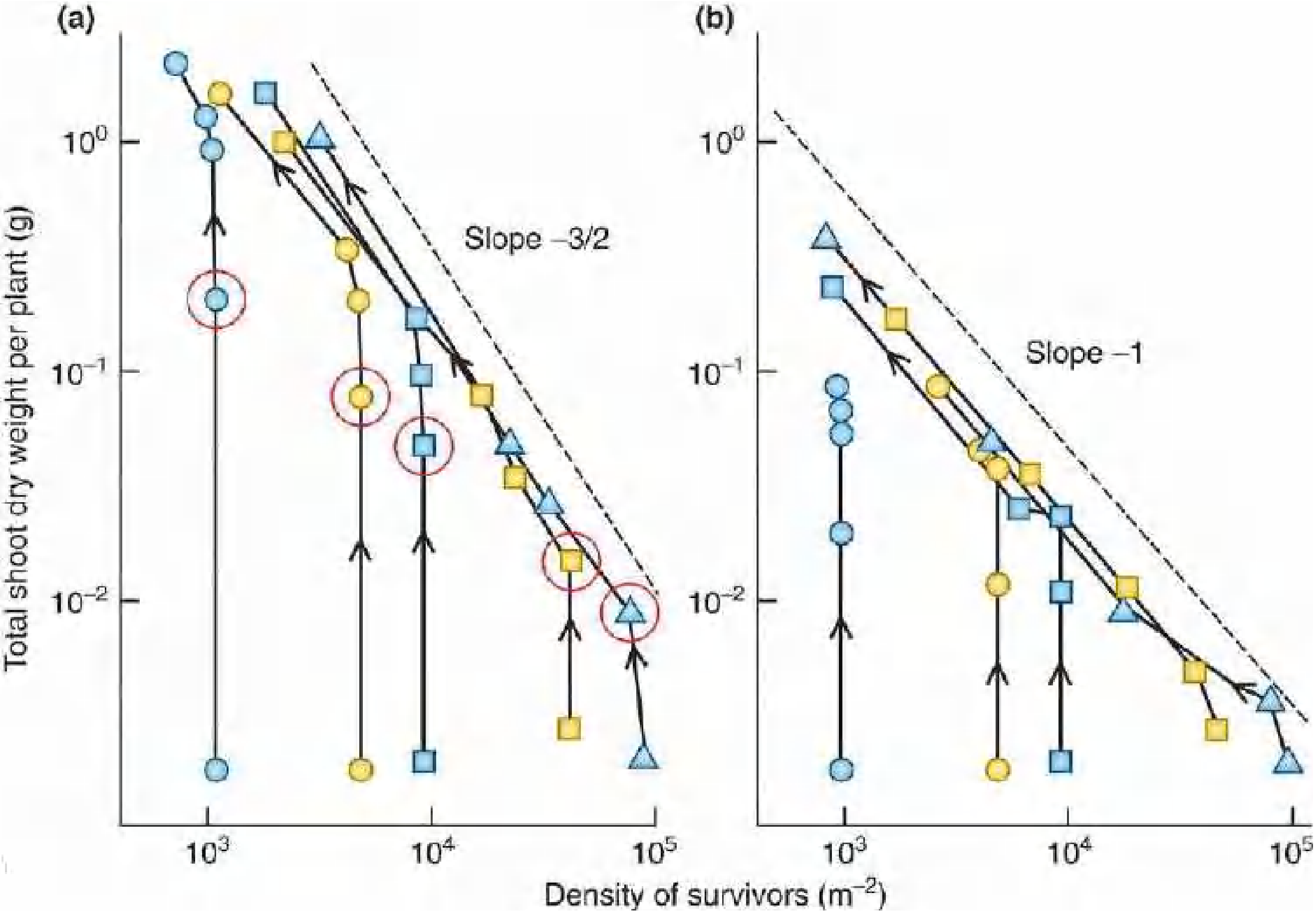


Figure 5.36 Crowded plant populations typically approach and then track self-thinning lines. Self-thinning in *Lolium perenne* sown at five densities: 1000 (🔵), 5000 (🟡), 10 000 (🟨), 50 000 (🟦) and 100 000 (▲) 'seeds' m^{-2}, in (a) 0% shade, where the observations after 35 days are circled in red, and (b) 83% shade. The lines join populations of the five sowing densities harvested on five successive occasions. They therefore indicate the trajectories, over time, that these populations would have followed. The arrows indicate the directions of the trajectories, i.e. the direction of self-thinning. For further discussion, see text.

Source: After Lonsdale & Watkinson (1983).

We can see that mean plant weight at a given age was always greatest in the lowest density populations (illustrated, for example, after 35 days (circled points) in Figure 5.36a). It is also clear that the highest density populations were the first to suffer substantial mortality. What is most noticeable, though, is that eventually, in all cohorts, density declined and mean plant weight increased in unison: populations progressed along roughly the *same* straight line. The populations are said to have experienced *self-thinning* (i.e. a progressive decline in density in a population of growing individuals), and the line that they approached and then followed is known as a *dynamic thinning line* (Weller, 1990).

The lower the sowing density, the later was the onset of self-thinning. In all cases, though, the populations initially followed a trajectory that was almost vertical, reflecting the fact that there was little mortality. Then, as they neared the thinning line, the populations suffered increasing amounts of mortality, so that the slopes of all the self-thinning trajectories gradually approached the dynamic thinning line and then progressed along it. Note also that Figure 5.36 has been drawn, following convention, with log density on the *x*-axis and log mean weight on the *y*-axis. This is not meant to imply that density is the independent variable on which mean weight depends. Indeed, it can be argued that the truth is the reverse of this: that mean weight increases naturally during plant growth, and this determines the decrease in density. The most satisfactory view is that density and mean weight are wholly interdependent.

’the –3/2 power law’

Plant populations (if sown at sufficiently high densities) have repeatedly been found to approach and then follow a dynamic thinning line. For many years, all such lines were widely perceived as having a slope of roughly –3/2, and the relationship was often referred to as the ‘–3/2 power law’ (Yoda *et al*., 1963 ; Hutchings, 1983), since density (*N*) was seen as related to mean weight $(\overline{w})$ by the equation:

$$\log \overline{w} = \log c - 3/2 \log N \qquad (5.22)$$

or:

$$\overline{w} = cN^{-3/2} \qquad (5.23)$$

where *c* is constant. (In fact, as we shall see, this is even further from being a universal law than the ‘law’ of constant final yield, discussed previously.)

Note, however, that there are statistical problems in using Equations 5.22 and 5.23 to estimate the slope of the relationship in that $\overline{w}$ is usually estimated as *B*/*N*, where *B* is the total biomass per unit area, and so $\overline{w}$ and *N* are inevitably correlated, and any relationship between them is, to a degree, spurious (Weller, 1987). It is therefore preferable to use the equivalent relationships relating overall biomass to density, lacking autocorrelation:

$$\log B = \log c - 1/2 \log N \qquad (5.24)$$

or:

$$B = cN^{-1/2}. \qquad (5.25)$$

5.9.2 Species and population boundary lines

If we now turn to the third type of study, listed in Section 5.9, we can note that although, again, combinations of density and mean weight for a particular species are being used to generate a relationship between them, it is not a single cohort that has been followed over time, but a series of crowded populations at different densities (and possibly different ages). In such cases, it is more correct to speak not of a thinning line but of a species boundary line – a line beyond which combinations of density and mean weight appear not to be possible for that species (Weller, 1990). Indeed, since what is possible for a species will vary with the environment in which it is living, the *species* boundary line will itself subsume a whole series of *population* boundary lines, each of which defines the limits of a particular population of that species in a particular environment (Sackville Hamilton *et al*., 1995).

dynamic thinning and boundary lines need not be the same

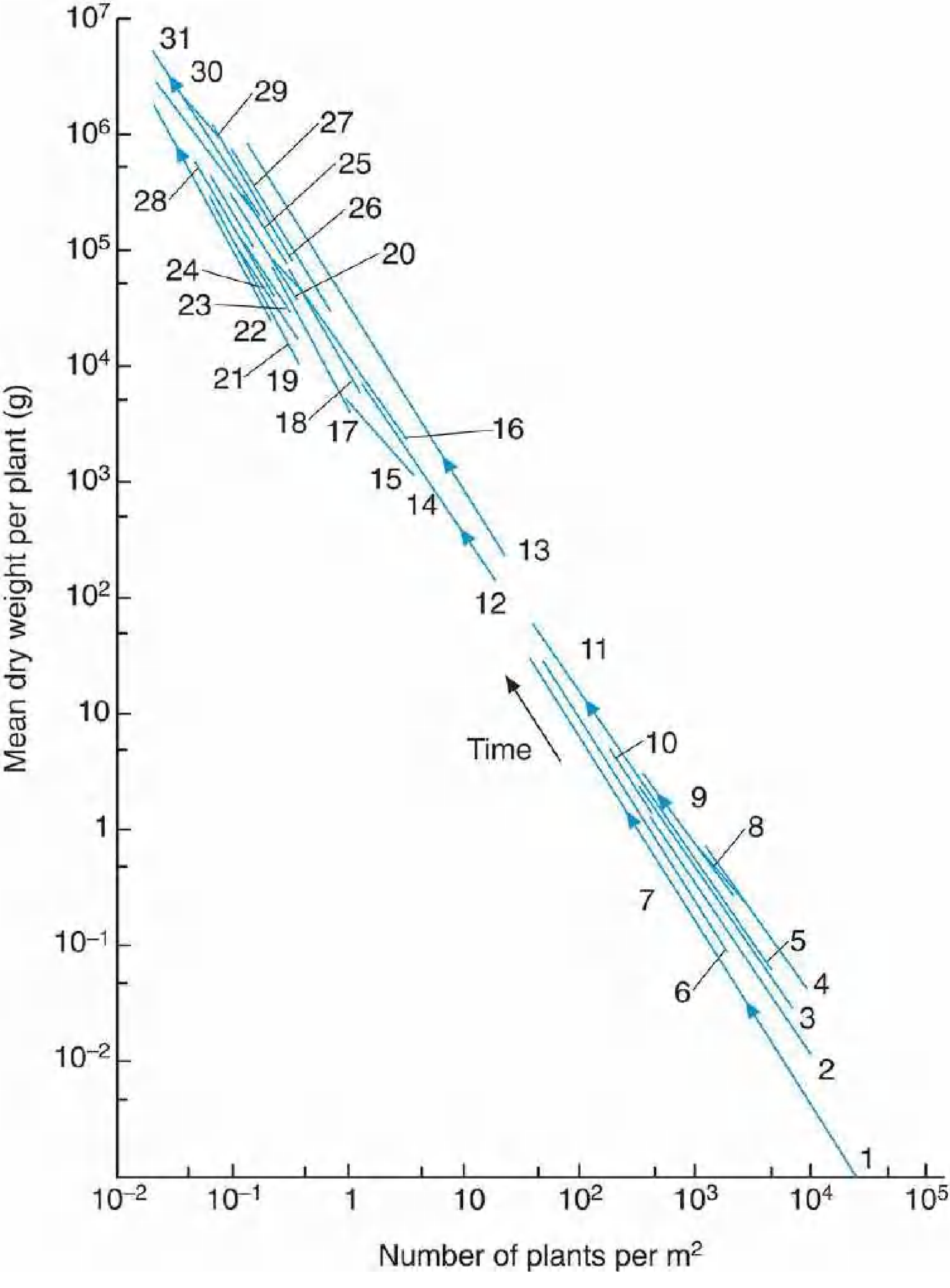
Thus, a self-thinning population should approach and then track its population boundary line, which, as a trajectory, we would call its dynamic thinning line – but this need not also be its *species* boundary line. The light regime, soil fertility, spatial arrangement of seedlings, and no doubt other factors may all alter the boundary line (and hence the dynamic thinning line) for a particular population (Weller, 1990 ; Sackville Hamilton *et al*., 1995). Soil fertility, for example, has been found in different studies to alter the slope of the thinning line, the intercept, neither, or both (Morris, 2002).

thinning slopes of –1

The influence of light, in particular, is worth considering in more detail, since it highlights a key feature of thinning and boundary lines. A slope of roughly –3/2 means that mean plant weight is increasing faster than density is decreasing, and hence that total biomass is increasing. But eventually this must stop: total biomass cannot increase indefinitely. Instead, the thinning line might be expected to change to a slope of –1, which would mean that loss through mortality is being exactly balanced by the growth of survivors, such that the total biomass remains constant. This can be seen when the populations of *Lolium perenne* discussed previously were grown at low light intensities (Figure 5.36b). A thinning line with a slope of –1 was apparent at much lower densities than it would otherwise be. This emphasises that boundary lines with negative slopes steeper than –1 (whether or not they are exactly –3/2) imply limits to the allowable combinations of plant densities and mean weights that set in before the maximum biomass from an area of land has been reached. Possible reasons are discussed in the next section.

5.9.3 A single boundary line for all species?

Intriguingly, when the thinning and boundary lines of all sorts of plants are plotted on the same figure, they all appear to have approximately the same slope and also to have intercepts (i.e. values of c in Equation 5.24) falling within a narrow range (Figure 5.37). To the lower right of the figure are high-density populations of small plants (annual herbs and perennials with short-lived shoots), while to the upper left are sparse populations of very large plants, including coastal redwoods (*Sequoia sempervirens*), the tallest known trees. Fashions change in science as in everything else. At one time, ecologists looked at Figure 5.37 and saw uniformity – all plants marching in –3/2 time, with variations from the norm seen either as ‘noise’ or of only minor interest (e.g. White, 1980). Subsequently, however, serious doubt was cast on the conformity of individual slopes to –3/2, and on the whole idea of a single, ideal thinning line (Weller, 1987 , 1990 ; Zeide, 1987; Lonsdale, 1990). This is reminiscent of discussions we described in Section 3.9 about whether there is a general rule (and a common slope) at the heart of a metabolic theory of ecology, and again, there really is no contradiction. On the one hand, the lines in Figure 5.37 occupy a very much smaller portion of the graph than one would expect by chance alone. There is apparently some fundamental phenomenon linking this whole spectrum of plant types: not an invariable ‘rule’ but an underlying trend. On the other hand, the variations between the lines are real and important and in as much need of explanation as the general trend.



Mean dry weight per plant (g)
10^7
10^6
10^5
10^4
10^3
10^2
10
1
10^{-1}
10^{-2}
Number of plants per m^2
10^{-2}
10^{-1}
1
10
10^2
10^3
10^4
10^5
Time
31
30
29
27
25
28
26
20
24
23
22
21
19
18
17
16
15
14
13
12
11
10
9
8
7
5
4
6
3
2
1

Figure 5.37 Self-thinning in a wide variety of herbs and trees. Each line is a different species, and the line itself indicates the range over which observations were made. The blue arrows, drawn on representative lines only, indicate the direction of self-thinning over time. The figure is based on Figure 2.9 of White (1980), which also gives the original sources and the species names for the 31 datasets.

5.9.4 An areal basis for self-thinning

We proceed, therefore, by examining possible bases for the general trend, and then asking why different species or populations might display their own variations on this common theme. Two broad types of explanation for the trend have been proposed. The first (and for many years the only one) is areal and based on a resource falling on the organisms from above (like light); the second is based on resource allocation in organisms of different sizes. Again, the similarities to the arguments at the heart of a metabolic theory of ecology, discussed in Section 3.9, are clear.

Limiting ourselves for now to plants, the areal argument runs as follows. In a growing cohort, as the mass of the population increases, the leaf area index (L, the leaf area per unit area of land) does not keep on increasing because beyond a certain point the canopy is full and so L remains constant irrespective of plant density (N). It is, in fact, precisely beyond this point that the population follows the dynamic thinning line. We can express this by saying that beyond this point:

$$L = \lambda N = \text{constant}, \tag{5.26}$$

where λ is the mean leaf area per surviving plant. However, the leaf area of individual plants increases as they grow, and so too therefore does their mean, λ. We expect λ, because it is an area, to be related to linear measurements of a plant, such as stem diameter, D, by a formula of the following type:

$$\lambda = aD^2 \tag{5.27}$$

where a is a constant. Similarly, we expect mean plant weight, P, to be related to D by:

$$\overline{w} = bD^3 \tag{5.28}$$

where b is also a constant. Putting Equations 5.26–5.28 together, we obtain:

$$\overline{w} = b(L/a)^{3/2} \cdot N^{-3/2}. \tag{5.29}$$

This is structurally equivalent to the −3/2 power relationship in Equation 5.23, with the intercept constant, c, given by $b(L/a)^{3/2}$.

It is apparent, therefore, why thinning lines might generally be expected to have slopes of approximately −3/2. Moreover, if the relationships in Equations 5.27 and 5.28 were roughly the same for all plant species, and if all plants supported roughly the same leaf area per unit area of ground (L), then the constant c would be approximately the same for all species. On the other hand, suppose that L is not quite the same for all species, or that the powers in Equations 5.27 and 5.28 are not exactly 2 or 3, or that the constants in these equations (a and b) either vary between species or are not actually constants at all. Thinning lines will then have slopes that depart from −3/2, and slopes and intercepts that vary from species to species. It is easy to see why, according to the areal argument, there is a broad similarity in the behaviour of different species, but also why, on closer examination, there are likely to be variations between species and no such thing as a single, 'ideal' thinning line.

complications of the areal argument

Furthermore, contrary to the simple areal argument, the yield–density relationship in a growing cohort need not depend only on the numbers that die and the way the survivors grow. We have seen (see Section 5.8) that competition is frequently highly asymmetric. If those that die in a cohort are predominantly the very smallest individuals, then density (*individuals* per unit area) will decline more rapidly as the cohort grows than it would otherwise do. It seems possible, too, to use departures from the assumptions built into Equations 5.26–5.29 to explain at least some of the variations from a 'general' −3/2 rule. We see this in a study by Osawa and Allen (1993), who estimated a number of the parameters in these equations from data on the growth of individual plants from two species: mountain beech (*Nothofagus solandri*) and red pine (*Pinus densiflora*). They estimated, for instance, that the exponents in Equations 5.27 and 5.28 were not 2 and 3, but 2.08 and 2.19 for mountain beech, and 1.63 and 2.41 for red pine. These suggest thinning slopes of −1.05 in the first case and −1.48 in the second, which compare quite remarkably well with the slopes that they observed: −1.06 and −1.48 (Figure 5.38). The similarities between the estimates and observations for the intercept constants were equally impressive. These results show, therefore, that thinning lines with slopes other than −3/2 can occur, but can be explicable in terms of the detailed biology of the species concerned. They also show that even when slopes of −3/2 do occur, they may do so, as with red pine, for the 'wrong' reason (−2.41/1.63 rather than −3/2).

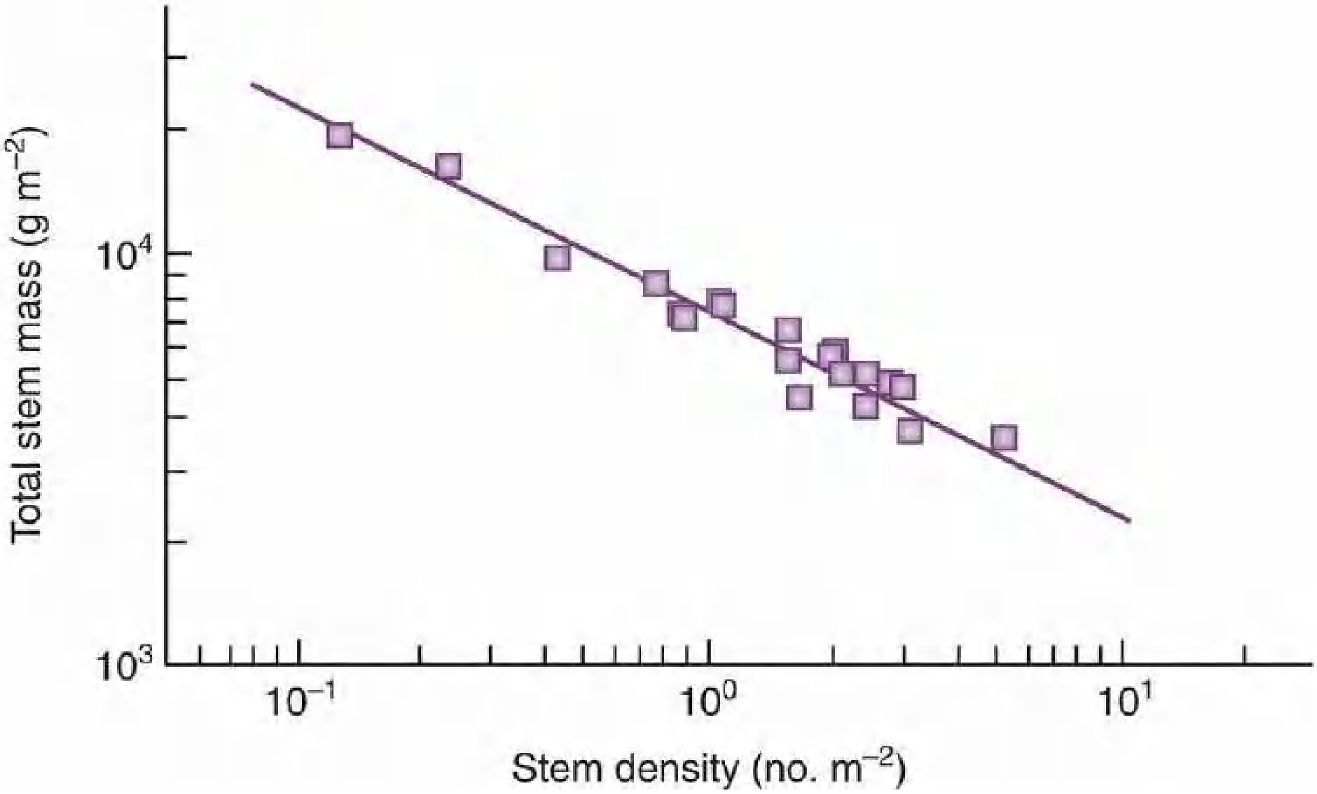


Figure 5.38 The species boundary line for populations of red pine, *Pinus densiflora*, from northern Japan (slope = −1.48).

Source: After Osawa & Allen (1993).

self-thinning in sessile animals

Animals must also 'self-thin', insofar as growing individuals within a cohort increasingly compete with one another and reduce their own density. And in the case of some sessile, aquatic animals, we can think of them, like plants, as being reliant on a resource falling from above (typically food particles in the water) and therefore needing to pack 'volumes' beneath an approximately constant area. It is striking, therefore, that in studies on rocky-shore invertebrates, mussels have been found to follow a thinning line with a slope of −1.4, barnacles a line with a slope of −1.6 (Hughes & Griffiths, 1988), and gregarious tunicates a slope of −1.5 (Guiñez & Castilla, 2001). There is, however, nothing linking all animals quite like the shared need for light interception that links all plants.

5.9.5 A resource-allocation basis for thinning boundaries

This need to include all types of organisms in considerations of self-thinning is reflected in studies seeking alternative explanations for the underlying trend itself. Most notably, Enquist *et al.* (1998) made use of the much more general model of West *et al.* (1997) that we discussed in Chapter 3, which considered the most effective architectural designs of organisms. We saw there that the rate of resource use per individual, u, or more simply their metabolic rate, should be related to mean organism body mass, M, according to the equation:

$$u = aM^{3/4}, \tag{5.30}$$

where a is a constant and the value ¾ is the ‘allometric exponent’.

−4/3 or −3/2?

They then argued that we can expect organisms to have evolved to make full use of the resources available, and so if S is the rate of resource supply per unit area and N_{max} the maximum density of organisms possible at this supply rate, then:

$$S = N_{\max}u \tag{5.31}$$

or, from Equation 5.30:

$$S = aN_{\max}\overline{w}^{3/4}. \tag{5.32}$$

But if the organisms have arrived at an equilibrium with the rate of resource supply, then S should itself be constant. Hence:

$$\overline{w} = cN_{max}^{-4/3}, \tag{5.33}$$

where c is another constant. In short, the expected slope of a population boundary on this argument is −4/3 rather than −3/2. Similarly, the figure in biomass–density relationships (equations 5.24 and 5.25) would be −1/3 rather than −1/2.

Enquist and colleagues themselves considered the available data to be more supportive of their prediction of a slope of −4/3 than the more conventional −3/2, though this had not been the conclusion drawn from previous data surveys. Indeed, as we saw in Chapter 3, the idea of ¾ being a consistent or universal allometric exponent, and a slope of −4/3 therefore being expected, has itself been called increasingly into question (e.g. Glazier, 2005). Nonetheless, Begon *et al.* (1986) had found a mean value of −1.29 (close to a value of −4/3) for self-thinning in experimental cohorts of grasshoppers, and Elliott (1993) a value of −1.35 for a population of sea trout, *Salmo trutta*, in the English Lake District. On the other hand, studies on house crickets, *Acheta domesticus*, have shown that the allometric exponent in in their case is not ¾ but 0.9 (Jonsson, 2017), and the estimated slope of their self-thinning line −1.11 (the exact reciprocal of the allometric exponent; Figure 5.39a).

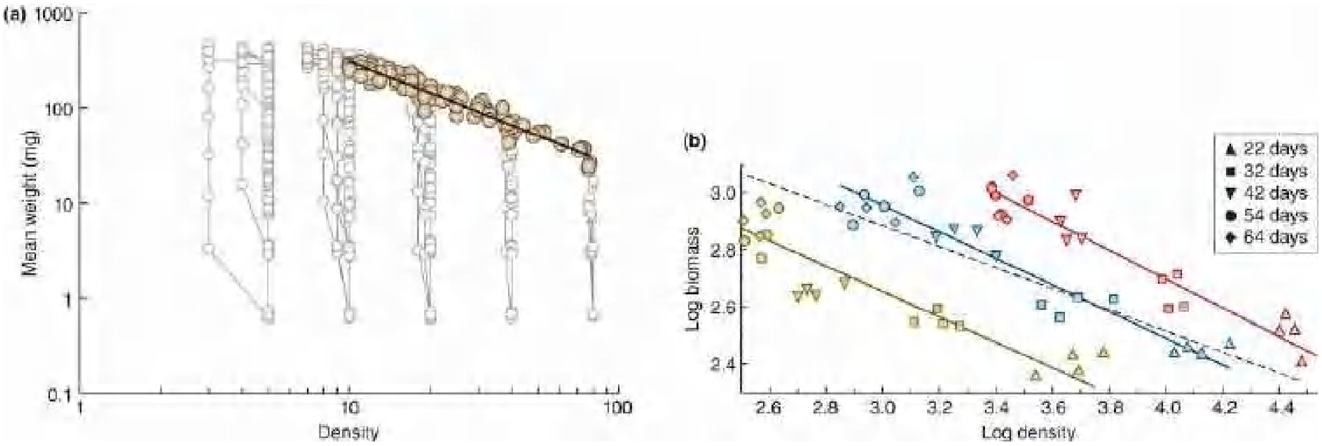


Figure 5.39 Self-thinning lines vary in their support for the metabolic theory. (a) Self-thinning in house crickets, *Acheta domesticus*, plotting mean weight against density on log scales. Replicate populations were established with between five and 80 newly hatched nymphs and followed until all survivors had hatched into adults. Lines join points from the same replicate, with the exception of the regression line fitted to self-thinning populations from the three highest densities only (slope ± 95% CI, −1.11 ± 0.05). (b) Self-thinning in common buckwheat, *Fagopyrum esculentum*, plotting log biomass against log density. Three initial densities, 8000 (green), 24 000 (blue) and 48 000 (red) individuals m^{-2} were harvested after 22, 32, 42, 54 and 64 days. The dashed line is fitted to all data combined (slope, 95% CI: −0.38 (−0.30 to −0.47)), and the solid lines are fitted to the individual initial densities (slopes, 95% CIs: −0.45 (−0.36 to −0.55), −0.47 (−0.40 to −0.55) and −0.50 (−0.43 to −0.59) for 8000, 24 000 and 48 000, respectively).

Source: (a) After Jonsson (2017). (b) After Li *et al.* (2013).

Moreover, when experimental populations of common buckwheat, *Fagopyrum esculentum*, were grown at a range of densities, the best estimate for the slope of the biomass–density relationship overall was −0.38 (Figure 5.39b), very similar to the value of −0.33 predicted by the metabolic theory (and significantly different from −0.5, predicted by the areal argument). But if separate lines were fitted for each of the three initial densities, the slopes were −0.45, −0.47 and 0.50, all consistent with the areal argument and significantly different from −0.33 (Figure 5.39b). The different intercepts of the three lines (plants sown at higher initial densities had greater biomass) seemed to reflect an effect of initial density on growth form and perhaps on the degree of asymmetry in the competitive process (Li *et al.*, 2013). This suggests, in turn, that light interception may drive patterns in individual populations while metabolic constraints set limits in a species overall.

What seems clear is that we have moved further from, not closer to, anything that could be called a self-thinning ‘law’. But this represents progress in the important sense of acknowledging the range of forces acting on growing, competing cohorts of individuals, and recognising, too, that the details of a species’ morphology or physiology may influence the way in which those forces act and the slopes of the resulting relationships. The patterns we observe are likely to be the combined effect of a range of forces, even if in some cases one of those forces may dominate – metabolic constraints in mobile animals, light interception in many plants; light interception in individual populations, metabolic constraints in a species overall. Universal rules have their attractions but Nature is not so easily seduced.

APPLICATION 5.4 Density management diagrams

The precise nature of self-thinning and species boundary lines, and of the forces shaping them, are important and interesting issues, but from the point of view of managing growing, crowded, single-species cohorts, those details are arguably less important than a simple recognition of the fundamental patterns that underlie all variants of these lines – that as cohorts grow and compete, there are boundaries in density-biomass or density-mean size space beyond which they cannot go, and trajectories that they tend to follow. This underpins, for example, one particular approach to the exploitation of commercially important growing cohorts: the construction and use of density management diagrams (DMDs) (Jack & Long, 1996). We conclude here, therefore, by looking at the use of a DMD for Norway spruce, *Picea abies*, in central-southern European montane regions (Figure 5.40), focusing on general principles rather than going into detailed calculations. Data had been compiled from France, Germany, Italy, the Czech Republic, Romania and Bulgaria on the density of trees and their average diameter in a total of 1609 plots, selected as having at least 80% Norway spruce and being markedly ‘even-aged’ (a strongly unimodal and non-skewed diameter distribution). These were then used to estimate a species-boundary – in this case, a linear combination of density and individual size on a log–log plot (Figure 5.40) that encompasses all but the 2% of plots with the greatest biomass. (Using this rather than the absolute maximum biomass prevents the line being overly influenced by a few rogue populations.) Other analyses of the data were able to estimate the speed and direction with which populations are likely to proceed through this size-density space.

We can see that the DMD encapsulates what the data can tell us about self-thinning in Norway spruce. The use of the DMD in managing any particular Norway spruce population (or type of population) then proceeds as follows: (1) the starting position of the population on the DMD is identified; (2) the target position is also identified, along with the likely trajectory to it in an unmanaged population; (3) the trajectory and time-scale of an alternative route to the target is estimated, usually based on a managed reduction in density, designed to prevent or delay the onset of competition-related mortality (Vacchiano *et al.*, 2013). In this case, examples include managing future timber production (illustrated in Figure 5.40), ensuring mechanical stability against wind damage (where susceptibility is greatest in stands with slender trees that arise when the intensity of competition is high), enhancing the protective powers of stands against avalanches and rockfalls (similar to windfirmness, but with the additional need to avoid gaps between trees that may ‘release’ an avalanche), and minimising vulnerability to spruce bark beetle attack (achieved by delaying canopy closure).

For the timber production in Figure 5.40, the target tree diameter, for commercial purposes, is 40 cm. From an initial population of trees 10 cm in diameter, growing at a density of 2600 trees per hectare, the DMD suggests that unmanaged self-thinning would take the population to that target diameter in around 90 years with around 650 trees per hectare. An alternative suggested in the figure, however, would be a period of managed thinning, taking the density to around 400 trees per hectare, following which they would grow unaffected by competition-induced mortality and reach the target diameter in a total of only 70 years. Although the overall harvest would be smaller, getting a commercial return 20 years earlier will often make perfect economic sense. Understanding both the principles and the details of Norway spruce self-thinning helps the forest managers achieve this.

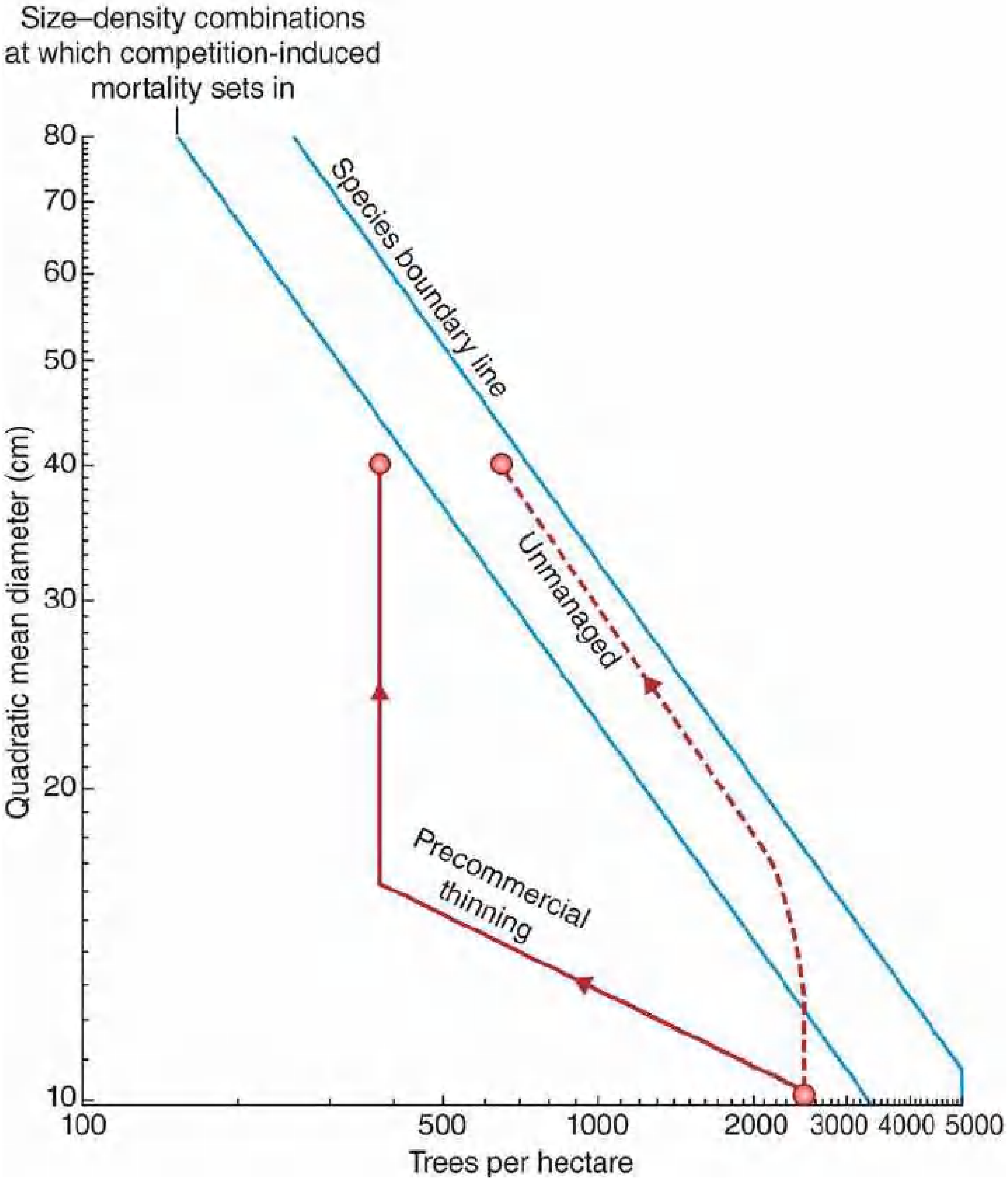


Figure 5.40 A density management diagram (DMD) for Norway spruce in central-southern European montane regions, based on observed combinations of stand density and average tree size (calculated as the quadratic mean diameter – the square root of the mean of the squared diameters), both plotted on log scales. Alternative trajectories over time are plotted from an initial population of 2600 10 cm diameter trees per hectare (bottom right, red circle). Unmanaged self-thinning would give rise to a population of *c*. 650 40 cm diameter trees per hectare in 90 years. Managed precommercial thinning would give rise to a population of *c*. 400 40 cm diameter trees per hectare in 70 years.

Source: After Vacchiano *et al*. (2013).

Chapter 6
Movement and Metapopulations

6.1 Introduction

All organisms in nature are where we find them because they have moved there. This is true for even the most apparently sedentary of organisms, such as oysters and redwood trees. Their movements range from the passive transport that affects many plant seeds to the apparently purposeful actions of mobile animals. Movement is itself linked directly to the structure of an organism's environment. Patterns of movement will only be favoured by natural selection insofar as they tend to take the individual from a less favourable to a more favourable location.

However, we should not be fooled into thinking that whenever an organism moves away from (escapes) its immediate environment, it has necessarily sensed that the environment is unfavourable and responded to that sensation; or that whenever an organism ceases its movement and arrives in a new location, it has necessarily sensed the attractiveness of its new home. This can often be the case, especially for animals. But there are also many examples where the behaviour has evolved in response to predictable patterns in the environment that the organism need not sense for itself in order for the response to be elicited. Seeds typically escape from their parent plant because the environment around the parent is predictably crowded. The behaviour has evolved over many generations: the seeds do not need to sense this for themselves. Likewise, many birds migrate polewards for the summer to regions with longer day lengths (and hence feeding periods) – but evolution has taken care of this: the birds typically respond to a proximate cue, probably the changing day length, rather than to diminishing resource availability per se.

A useful distinction in this context is between geographic space and environmental space. Studies of geographic space are focused on the home ranges of individuals or the geographic distributions of species. Studies of environmental space are focused on patterns in the selection and use of resources. Van Moorter *et al.* (2016) characterise individual movements as the glue, binding the two together. Individual movements reflect an immediate or evolutionary response to patterns in the distribution of resources; we observe the consequences in home ranges and species' distributions.

Whatever the precise details in particular cases, it is useful to divide individual movements into three phases: starting (leaving), moving itself, and stopping (South *et al.*, 2002), or, put another way, emigration, transfer and immigration (Ims & Yoccoz, 1997). The questions we ask about the three phases differ both from a behavioural point of view (for example, what triggers the initiation and cessation of movement?) and from a demographic point of view (the distinction between loss and gain of individuals). A broader perspective still proposes a four-part framework onto which all kinds of movement studies can fit (Nathan *et al.*, 2008). The first is the organism's internal state (why move?); the second is the nature of the movement itself; third are the environmental drivers of different patterns of movement (when and where to move?); and last are the ecological and evolutionary consequences of movement. We touch on all of these in what follows.

‘dispersal’ and ‘migration’

Dispersal is most often taken to mean the movement of individuals away from others, and is therefore an appropriate description for several kinds of movements: (i) of plant seeds or starfish larvae away from each other and their parents; (ii) of voles from one area of grassland to another, usually leaving residents behind and being counterbalanced by the dispersal of other voles in the other direction; and (iii) of land birds amongst an archipelago of islands (or aphids amongst a mixed stand of plants) in the search for a suitable habitat. A common and somewhat narrower view of dispersal, and one that distinguishes it more clearly from simply ‘movement’, defines two categories of dispersal: *natal* dispersal, the movement of a prereproductive individual from its site of birth to its site of reproduction, and *postbreeding* dispersal, the movement of individuals between successive sites of reproduction (Matthysen, 2012). Natal dispersal is more common, if for no other reason than it is the only dispersal available to organisms, like plants, with a static reproductive stage, and it has been the main preoccupation of ecologists working on dispersal.

Migration, then, is most often taken to mean the movement, frequently directional, either of individuals or of collections of individuals from a starting to a terminal location, which is often predetermined. The term therefore applies to classic migrations (the movements of locust swarms, the intercontinental journeys of birds) but could also be applied, for example, to the to and fro movements of shore animals following the tidal cycle. We deal first here with migration and then move on to dispersal more generally.

technological advances

Before we do so, however, we should acknowledge the enormous increase there has been recently in our ability to collect and analyse movement data. As Börger (2016) describes, we have been tracking movements for more that 200 years by marking an individual in one location (for example, a leg ring on a bird) so that we can recognise that same individual when we recapture or resight it somewhere else, some time later. But as it has become increasingly easy to monitor resightings without ever capturing an animal – for example, using camera traps linked to photoidentification or non-invasive genetic typing – this approach has become possible in an ever-widening range of cases. Even more profound have been advances in the technology of electronic tags (and decreases in their size and weight) with, for example, signals indicating an individual’s location transmittable not simply to a nearby investigator but to satellite systems that can also store the location data being collected. Note, too, that information on location, alone, is of limited value. It is much better to have this allied to a characterisation of the habitat at that location, and there have been equally significant advances in satellite-based remote-sensing of the environment and in geographical information systems (GIS) able to manage those data and link them, where appropriate, to movement data. Needless to say, these advances in the development of tools to collect data have been paralleled by advances in computing power and software sophistication. As in other sciences, the challenge will be to use these new tools to answer, perhaps for the first time, the most important ecological questions rather than being seduced by them into collecting datasets simply ‘because we can’.

6.2 Patterns of migration

migrating to track resources

Individuals of many species move *en masse* from one habitat to another and back again repeatedly during their life. The timescale involved may be hours, days, months or years. As we shall see throughout this book, Nature usually defies our attempts to place species' behaviours into neat, water-tight categories, but there is something to be said for dividing migratory movements into those either tracking the same environment or moving between quite distinct environments, or perhaps into being driven either by tracking the environment, finding a breeding site or finding a refuge (Shaw, 2016). The movement of crabs on a shoreline is an example of a tracking migration: they move with the advance and retreat of the tide. Monarch butterflies, *Danaus plexippus*, in North America migrate south to Mexico to overwinter (taking refuge from the harsh northern winters) but then take four generations to migrate northwards again over the summer, tracking the availability of their host plant, the milkweed, *Asclepias* spp.

APPLICATION 6.1 Tracking the tracking migrations of locusts

Many locust species also have tracking migrations, following the availability of food, itself driven by patterns of rainfall. However, these migrations, even in those species where they occur, are irregular and complex. The desert locust, *Schistocerca gregaria*, which lives across broad swathes of north Africa and east to India, and which can be devastating to crops throughout this region, is a good example. It passes through a series of wingless nymphal stages before moulting into an adult, and crucially can exist as either of two 'phases': the solitary phase, which they display when densities are low, and the gregarious phase, which they transition into and remain in when densities are high. The phases differ morphologically, but most important, their behaviour also differs as the phase names suggest: those in the solitary phase acting as individuals but the gregarious locusts behaving in a concerted way as coherent units. All stages have voracious (collective) appetites and can destroy crops. Gregarious nymphs form 'bands' (having transitioned in their solitary phase through 'groups') that then march across the landscape, tracking available vegetation. Gregarious adults form 'swarms', also having transitioned through groups, which then take to the skies, sometimes in their hundreds of millions. Crucially, by flying downwind for up to 200 km in a day, these swarms are migrating to where rain (and fresh vegetation) is most likely rather than tracking its current availability. Regardless of whether the migrations are tracking or predictive, though, nymphal or adult, they themselves are tracked by the Food and Agricultural Organization of the United Nations (the FAO), providing farmers and governments throughout the region with early warning alerts through their Desert Locust Watch scheme (Figure 6.1), combining data from satellites, GIS and even drones.

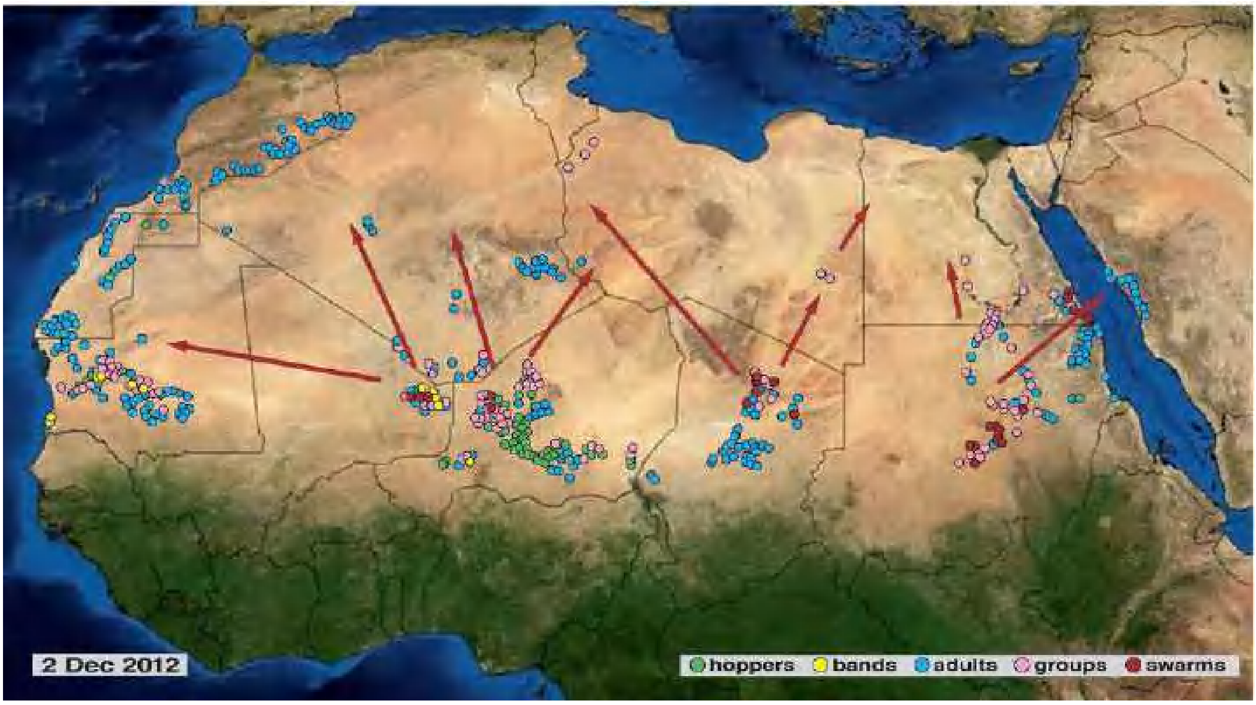


Figure 6.1 **The movements of locusts and their impending threat.** An example from 2 December 2012 of an alert issued by the Food and Agricultural Organization of the United Nations of the impending threat from desert locusts in north Africa. Hoppers (wingless nymphs) and adults are in the solitary phase; groups are transitioning into the gregarious phase; and bands and swarms are the large (often massive) aggregation of nymphs and adults, respectively. The red arrows show the current (2 December 2012) migration paths.

Source: After FAO (2012).

Not all migrations are easily categorised. American elk (*Cervus elaphus*) and mule deer (*Odocoileus hemionus*) move up into high mountain areas in the summer and down to the valleys in the winter. We can think of this as taking refuge from high altitude winters, but there is also a strong element of their tracking the most favourable supply of food. On a much larger scale, the Arctic tern (*Sterna paradisaea*) travels from its Arctic breeding ground to the Antarctic pack ice and back each year – about 10 000 miles (16 100 km) each way (although unlike many other migrants it can feed on its journey). So this is a breeding migration, and in that sense the terns are deriving something quite different from their Arctic and Antarctic environments. But they are also migrating to maintain a year-round supply of essentially the same kind of food, and taking refuge at both ends of their journey from polar winters that they could not survive. Similarly, many terrestrial birds in the northern hemisphere move north in the spring to breed when food supplies become abundant during the warm summer period, and move south to savannas in the autumn when food becomes abundant only after the rainy season; while baleen whales in the southern hemisphere move south in summer to feed in the food-rich waters of the Antarctic, and move north in winter to breed (but scarcely to feed) in tropical and subtropical waters.

migrating between distinct resources

In other cases, migration may involve a more straightforward movement between two distinct environments. For example, many planktonic algae descend to the depths at night to accumulate phosphorus and perhaps other nutrients but move to the surface during the day to photosynthesise. Likewise, but on a very different time scale, many amphibians (frogs, toads and newts) migrate between an aquatic breeding habitat in spring and a terrestrial environment for the remainder of the year, the young developing (as tadpoles) in water with a different food resource from what they later eat on land.

one-way tickets

Many long-distance migrants, however, make only one return journey during their lifetime. They are born in one habitat, make their major growth in another habitat, but then return to breed and die in the home of their infancy. Eels and migratory salmon provide classic examples. The European eel (*Anguilla anguilla*) travels from European rivers, ponds and lakes across the Atlantic to the Sargasso Sea, where it is thought to reproduce and die (although spawning adults and eggs have never actually been caught there). The American eel (*A. rostrata*) makes an analogous journey from areas ranging between the Guianas in the south, to south-west Greenland in the north. Salmon make a comparable transition, but from a freshwater egg and juvenile phase to mature as a marine adult. The fish then returns to freshwater sites to lay eggs. After spawning, all Pacific salmon (*Oncorhynchus nerka*) die without ever returning to the sea. Many Atlantic salmon (*Salmo salar*) also die after spawning, but some survive to return to the sea and then migrate back upstream to spawn again.

Most migrations occur seasonally. They are typically triggered by some external seasonal phenomenon (e.g. changing day length), and perhaps also by an internal physiological clock. They are often preceded by quite profound physiological changes such as the accumulation of body fat and represent strategies evolved in environments where seasonal cycles of habitat favourability are reliably repeated from year to year. By contrast, the economically disastrous migration plagues of locusts in arid and semiarid regions, for example, are tactical, forced by events such as overcrowding, and typically have no cyclicity or regularity. These are most common in environments where rainfall is not seasonally reliable.

APPLICATION 6.2 The conservation of migratory species

Species that spend part of their time in one habitat and part in another can be badly affected by human activities that influence the ability to move between them. The declining populations of river herrings (*Alosa pseudoharengus*, also known as the alewife, and *A. aestivalis*) in the north-eastern USA are a case in point. Adults ascend coastal rivers to spawn in lakes between March and July, and the young fish then remain in fresh water for three to seven months before migrating to the ocean. The species are commercially important as both food and bait, but overfishing, pollution and the construction of dams on their migration paths have prompted the US National Marine Fisheries Service to classify them as a 'species of concern'. Yako *et al.* (2002) sampled river herrings three times per week from June to December in the Santuit River downstream of Santuit Pond, Massachusetts, USA which contains the only spawning habitat in the catchment. The pond is dammed and discharge of water from it managed – but when are the best times to discharge the water? To answer this, Yako *et al*. classified periods of downstream juvenile migration as either 'peak' (>1000 fish per week) or 'all' (>30 fish per week, obviously including the peak), and by simultaneously measuring a range of physicochemical and biotic variables, they aimed to identify factors that could predict the timing of this migration (Figure 6.2). Peak migration was most likely to occur during the new moon (dark nights perhaps reducing the risk from piscivorous fish and birds) and when the density of important zooplankton prey (*Bosmina* spp.) was low, and migration at any level tended to occur when water clarity was low (perhaps because their own ability to forage successfully was impaired and/or because risk from piscivores was lower) and during decreased periods of rainfall. With this information, predictive models could be built to help managers identify periods when river discharge needs to be maintained to coincide with migration. On the other side of the pond, in 2013 a new 'fish ladder' was opened (small ascending steps in the river's profile) to help spawning fish reach the pond from the sea.

keeping routes open for commercial fisheries

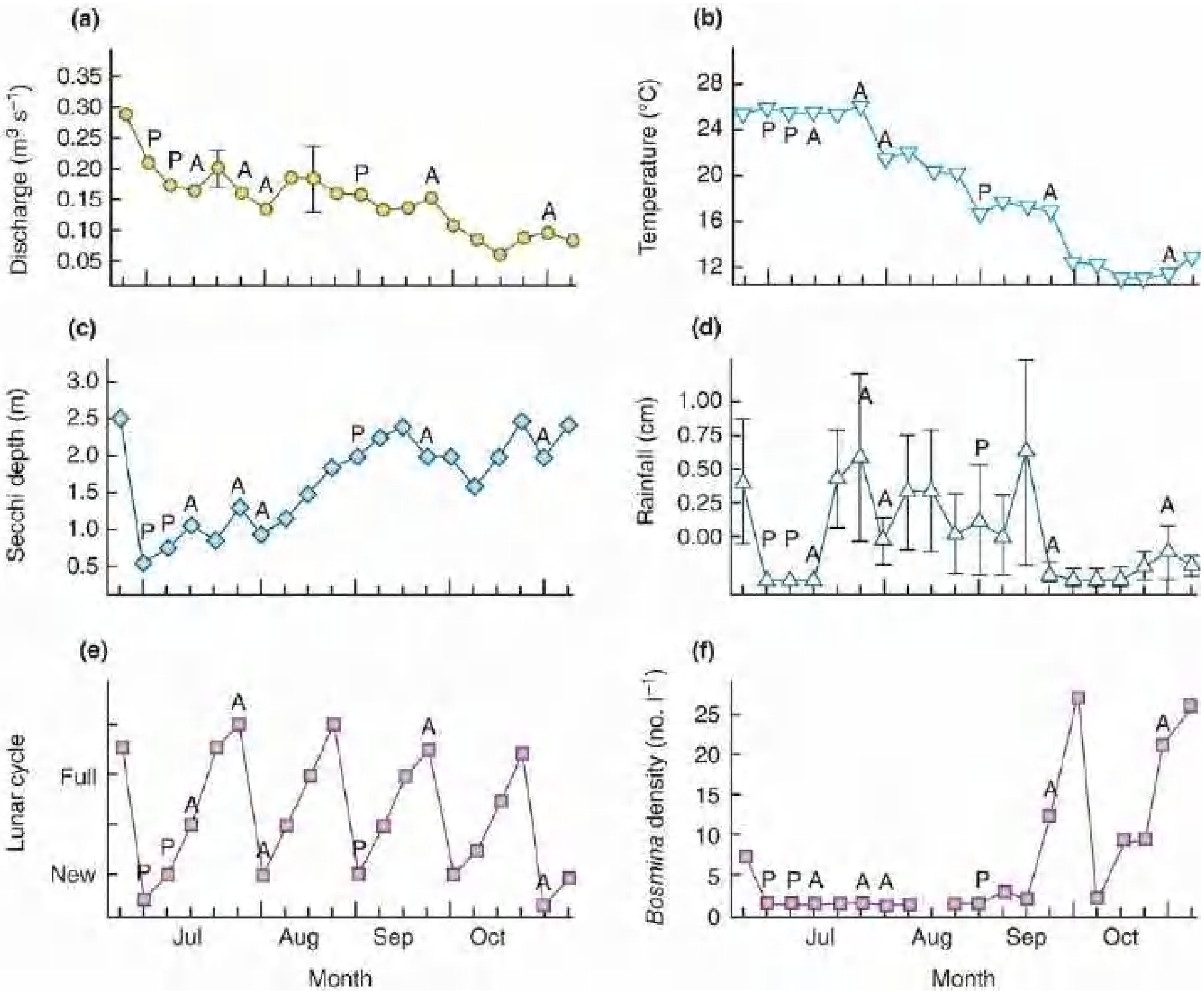


Figure 6.2 Variation in physical and biotic variables in the Santuit River, USA during the migratory period of river herrings. (a) Discharge, (b) temperature, (c) water clarity, measured as Secchi disc depth (low values indicate high turbidity), (d) rainfall, (e) lunar cycle and (f) *Bosmina* (prey) density. P denotes 'peak' periods of migration (>1000 fish per week). P and A (>30 fish per week) together denote all periods of migration. Bars are SEs.

Source: After Yako *et al.* (2002).

More broadly, an understanding of the migratory behaviour of species at risk can assist managers in devising conservation strategies. One example is a scheme to divert the southwards migration route of lesser white-fronted geese (*Anser erythropus*) from arriving in south-eastern Europe, where they tend to get shot, to spending their winters in the Netherlands (Sutherland, 1998). To do this, a captive population of barnacle geese (*Branta leucopsis*) that breeds in Stockholm Zoo but overwinters in the Netherlands was used. Individual barnacle geese were taken from Stockholm to Lapland where they nested and were given lesser white-fronted goose eggs to rear. Then, when the adoptive barnacle goose parents flew to the Netherlands for the winter, the young lesser white-fronted geese flew with them. But crucially, next spring the young lesser white-fronted geese returned to Lapland and bred with conspecifics there, subsequently returning again to the Netherlands along what was now their accustomed route.

designing nature reserves for pandas

Migratory movements may also play a role in the design of nature reserves. The Qinling Province in China is home to approximately 220 giant pandas (*Ailuropoda melanoleuca*), representing about 20% of the wild representatives of one of the world's most imperilled mammals. The pandas in this region are elevational migrants. They need both low and high elevation habitat to survive. But current nature reserves did not cater for this. Pandas are extreme dietary specialists, primarily consuming a few species of bamboo. In Qinling Province, from June to September, pandas eat *Fargesia spathacea*, which grows from 1900 to 3000 m. But as colder weather sets in, they travel to lower elevations, and from October to May they feed primarily on *Bashania fargesii*, which grows from 1000 to 2100 m.

manipulating the seasonal migration of geese

Loucks *et al.* (2003) sought to identify regions of the landscape that would meet the long-term needs of the species. They first excluded areas lacking giant pandas, forest block areas that were smaller than 30 km^2 (the minimum area needed to support a pair of giant pandas over the short term) and forest with roads, settlements or plantations. Based on this, Figure 6.3b maps summer habitat (1900–3000 m; *F. spathacea* present), autumn/winter/spring habitat (1400–2100 m; *B. fargesii* present) and a small amount of year-round habitat (1900–2100 m, both bamboo species present). Four areas of core panda habitat (A–D) were identified that provide for the migrational needs of the pandas. Figure 6.3b also shows the location of the current nature reserves. They cover only 45% of the core habitat. Loucks *et al.* (2003) recommend that the four core habitat areas identified should be incorporated into a reserve network. They also note the importance of promoting linkage between these core areas, because extinction in any one area (and ultimately in all combined) is more likely if the populations are isolated from each other (see Section 6.7). Thus, they further identify two important linkage regions for protection, between areas A and B where steep topography means few roads exist, and between B and D across high elevation forests.

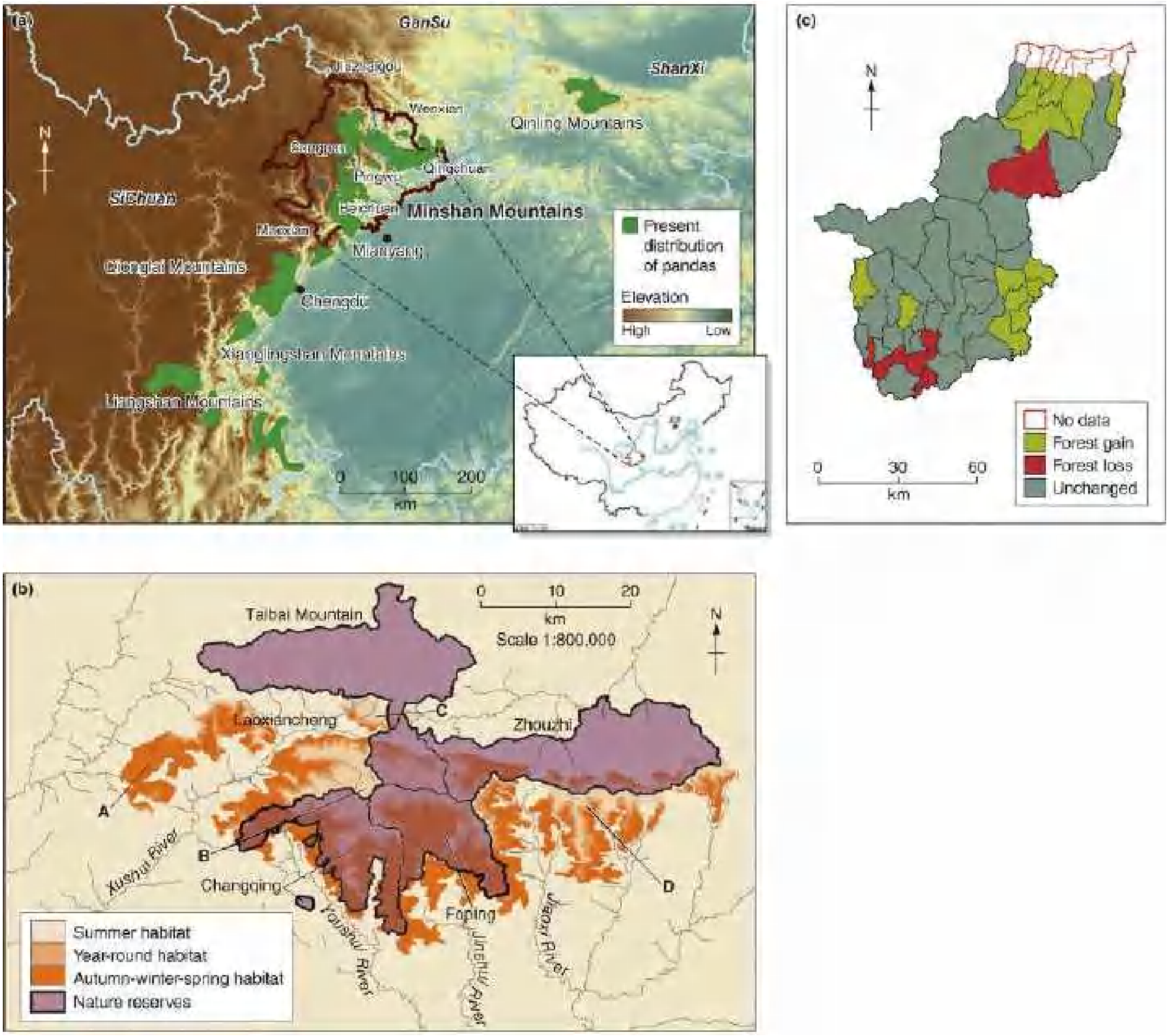


Figure 6.3 The optimal location of nature reserves for giant pandas in China must take account of their elevational migrations. (a) The present distribution of the giant panda, *Ailuropoda melanoleuca*, with the location within this of the Qinling Mountains (examined further in this figure) and the Minshan Mountains (see Figure 6.34). (b) Core panda habitats (A–D) in the Qinling Mountains, each of which caters for the year-round needs of the elevational migration of giant pandas. Superimposed are nature reserves as of 2003 (purple) and their names. (c) Subsequent assessment of forest cover change between 2001 and 2008 in the southern section of the Qinling Mountains region – part of the Changqing reserve in (b).

Source: (a) After Shen *et al*. (2015). (b) After Loucks *et al*. (2003). (c) After Li *et al*. (2013).

Of course, making recommendations is easier that implementing them, especially when there are conflicting demands being made on the landscape. It is therefore encouraging that a subsequent analysis of forest cover change in the region between 2001 and 2008 showed an overall increase (Figure 6.3c), and that this tended to occur where the Chinese Government had implemented its Grain-To-Green Programme (converting agricultural land to forest), though there was less of an increase (and in some cases decreases), because of the need to grow food, at lower altitudes and where a higher proportion of people were agricultural workers.

6.3 Modes of dispersal

6.3.1 Passive dispersal

seed rains and seed shadows

Much seed dispersal is passive. Most seeds fall close to the parent and their density declines with distance from that parent. This is the case for wind-dispersed seeds and also for those that are ejected actively by maternal tissue (e.g. many legumes). There is a 'seed rain' generated by the mother plant that produces a 'seed shadow' on the ground. The eventual destination of the dispersed offspring is determined by the original location of the parent and by the relationship between disperser density and distance from parent, but the detailed microhabitat of that destination is left largely to chance. Dispersal is non-exploratory. Some animals have essentially this same type of dispersal. For example, the dispersal of most pond-dwelling organisms without a free-flying stage depends on resistant wind-blown structures (e.g. gemmules of sponges, cysts of brine shrimps).

rare but important long-distance dispersal

The density of seeds is often low immediately under the parent, rises to a peak close by and then falls off steeply with distance (Figure 6.4). However, there are immense practical problems in studying seed dispersal (i.e. in following the seeds), and these become increasingly irresolvable further from the source. Greene and Calogeropoulos (2001) liken any assertion that 'most seeds travel short distances' to a claim that most lost keys and contact lenses fall close to streetlights. Indeed, for seeds, but also more generally, the relatively rare long-distance dispersers may be especially important for the invasion of new habitats or habitat patches (see Section 6.4.1). And when studies have sought evidence for long-distance dispersers they have often found it. For example, the data compiled to estimate the dispersal profiles of the trees in Figure 6.4 were obtained from sites within which the distributions of the four study species had been mapped (and their sizes determined in order to estimate their fecundity) and seeds collected from a network of seed traps within the sites. Spatial models were then fitted to these data assuming either that all seeds were derived from the known, local trees, or that there was an additional background input of long-distance dispersers that could not be attributed to parent trees within 50 m. In all cases the models with long-distance dispersers were a much better fit to the data. Hence, while there was a fall off within around 5–10 m in the seed rain from individual trees, the tail of that distribution (the outer reaches of the seed shadow) extended, at significant levels, 10 or more times further, as judged by the input from equivalent trees outside the study site (Figure 6.4)

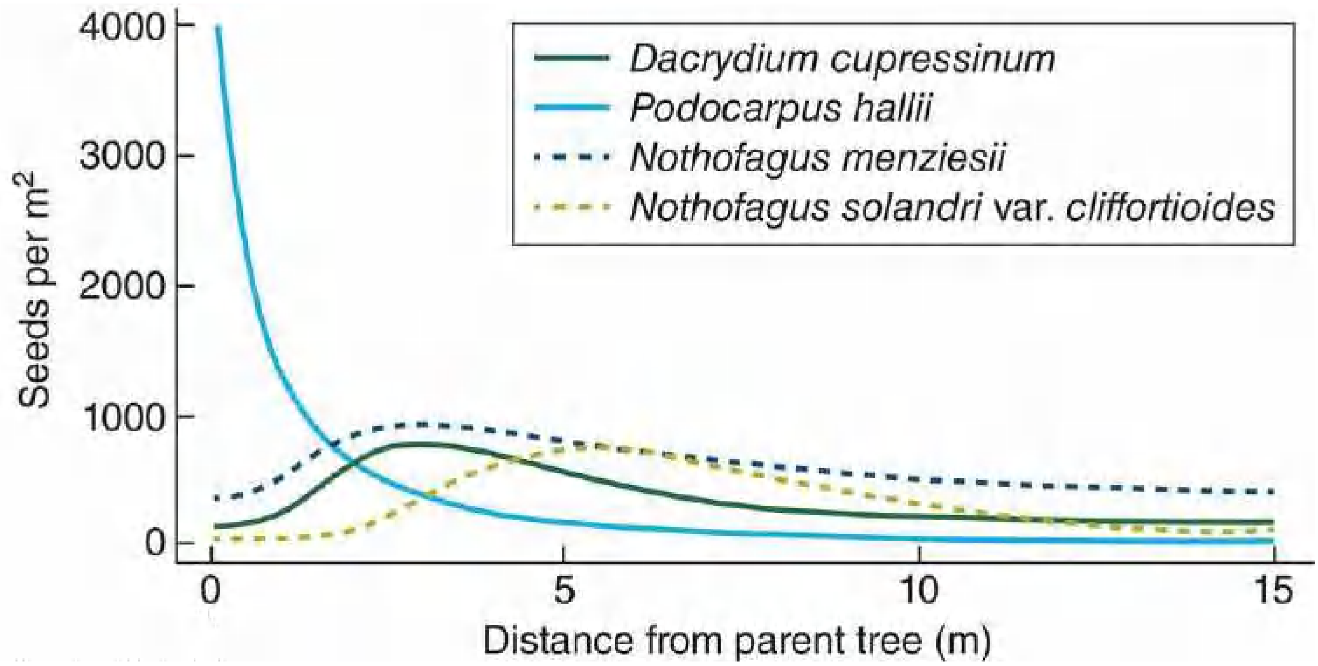


Figure 6.4 The 'seed rains' of four tree species from a temperate rainforest in New Zealand. These were standardised for tree size (60 cm diameter at breast height), and combined the estimated total seed production (TSP) and input from outside the immediate area for the study year when TSP was highest.

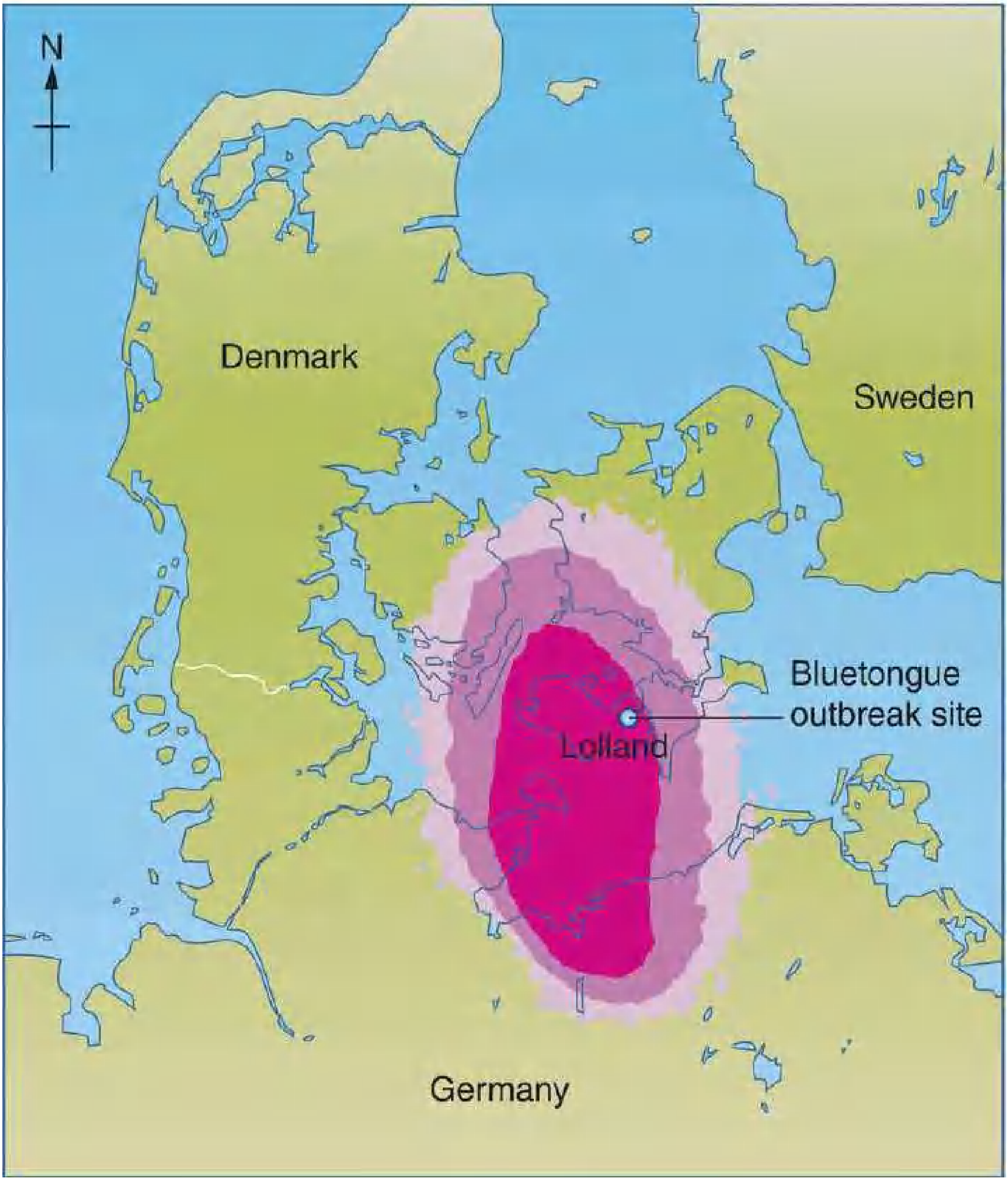
Source: After Canham *et al.* (2014).

6.3.2 An active–passive continuum

The reality, however, is that if we wish to classify the dispersal movements of organisms away from one another, we need a scale from passive to active, and there is often no clear distinction between the two. Young spiders, for example, climb to high places and then release gossamer threads that carry them on the wind where they are at the mercy of air currents, but their 'starting' is active even if moving itself is effectively passive. Even the wings of insects are often simply aids to what is effectively passive movement.

APPLICATION 6.3 Winds predict the arrival of midges carrying bluetongue virus

We see this, for example, in the success with which wind patterns are able to predict the invasion of bluetongue virus into new areas of Europe. Bluetongue is an economically important disease of ruminants (especially sheep and cattle), found in many parts of the world, that since around 1998 has been spreading from north Africa and the Middle East northwards through Europe, carried by its vector: biting midges from the genus *Culicoides*. The midges can, of course, fly, but their powers of flight alone tell us very little about their spreading of the disease. However, a model that treats midges just as it would any other wind-borne particle does an excellent job of predicting risk areas for new outbreaks, and indeed, linking new outbreaks to their source. This model is the United Kingdom Meteorological Office's Numerical Atmospheric-dispersion Modelling Environment programme (NAME), modified to incorporate key aspects of midge biology like their daily and seasonal activity cycles. The midges' wings give them the capacity to be carried by the wind, but the wind plays the major role in determining where they will go. An example is shown in Figure 6.5.



[**Figure 6.5**](#) **Predicting an outbreak of bluetongue virus.** Results from the modelling programme NAME (see main text) showing the estimated relative likelihood of *Culicoides* midges from northern Germany, to the south, potentially carrying bluetongue virus, reaching the region around the island of Lolland in southern Denmark overnight 22–23 September 2007. The darker the pink, the greater the likelihood. The blue dot shows the actual bluetongue outbreak site.

Source: After Burgin *et al*. (2012).

passive dispersal by an active agent

In the case of seed dispersal, too, the role of the seed itself may be passive, but uncertainty as to the destination will be reduced if there is also an active agent of dispersal. The seeds of many herbs of the woodland floor have spines or prickles, increasing their chances of being carried passively on the coats of animals and tending to concentrate their destinations in nests or burrows when the animal grooms itself. The fruits of many shrubs and lower canopy trees are fleshy and attractive to birds, and the seed coats resist digestion in the gut. The destination of the seed then depends on the defaecating behaviour of the bird. With fruits, therefore, the associations are ‘mutualistic’ (beneficial to both parties – see Chapter 13): the seed is dispersed and the disperser consumes the fleshy ‘reward’. There are also important examples in which animals are dispersed by active, mutualistic agents. For instance, many species of mite are taken very effectively and directly from dung pat to dung pat, or from one piece of carrion to another, by attaching themselves to dung or carrion beetles. The mites gain a dispersive agent, and many of them attack and eat the eggs of flies that would otherwise compete with the beetles.

We can see the combination of active and passive processes in the patterns of dispersal of seeds by frugivores (fruit-eaters) in Figure 6.6. Seeds of the St Lucie cherry, *Prunus mahaleb*, at a site in Spain containing 196 cherry trees, were detected and then assigned to a parent tree by DNA-based genotyping and to a frugivore type by noting the faeces in which the seed was found. The frugivore types were mammals (for example foxes and badgers), small birds (including warblers, redstarts and robins) and two medium-sized birds, the mistle thrush (*Turdus viscivorous*) and the carrion crow (*Corvus corone*). The seed shadows overall were reminiscent of the ‘passive’ patterns for wind-dispersed seeds in Figure 6.4: a rapid decline in numbers close to the source, combined with a long tail (though the distances in this case were much greater). However, there were also important differences in pattern generated by the characteristic activities of the different frugivores. The small birds deposited most seeds only a short distance from the parent tree and did so into covered microhabitats (under cherry trees and various shrubs). These birds, therefore, provided the seedlings with their highest chances of establishing. The mammals and larger birds, on the other hand, were the most important groups in terms of dispersal between cherry populations and in starting new populations (assuming the seeds germinated and established), because they dispersed seeds longer distances and did so selectively into more open habitats (the mammals and crows) or into pine forests (the thrushes).

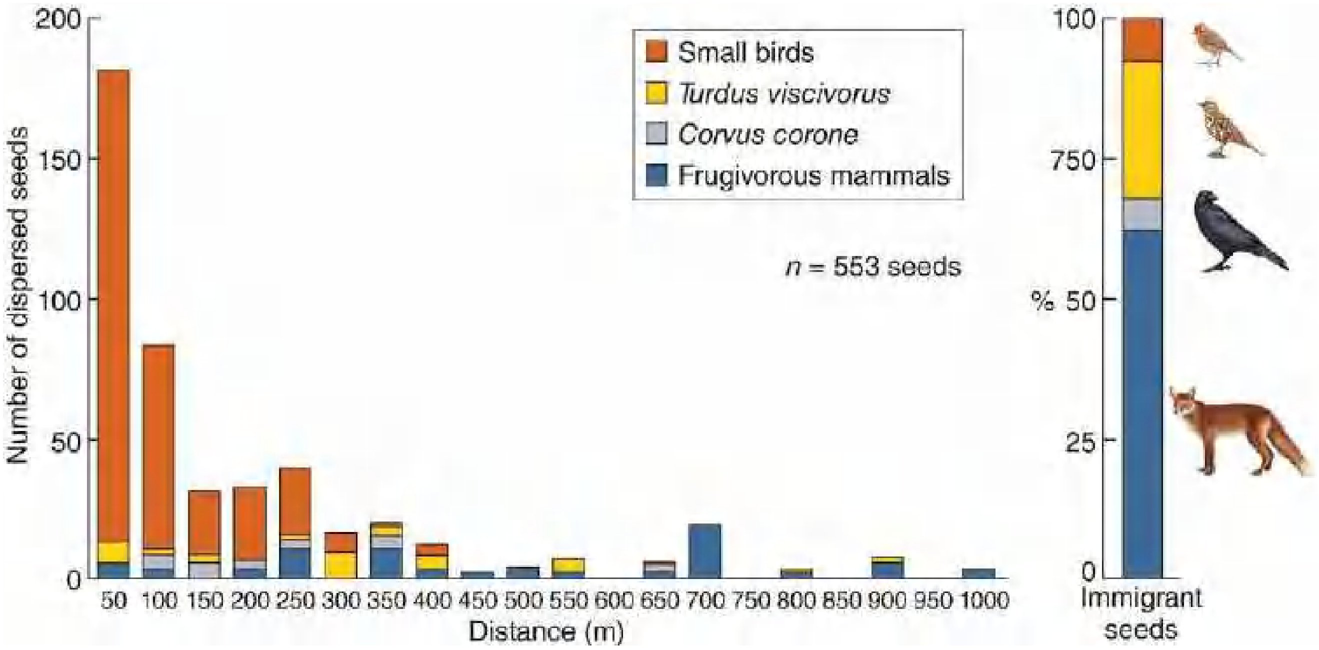


Figure 6.6 Seed dispersal by frugivores can show a variety of patterns. The left panel shows the profile of distances dispersed by seeds of the St Lucie cherry, *Prunus mahaleb*, in Spain carried by four types of frugivore, as indicated in the right panel, which shows the percentage contributions of the four groups to long distance dispersal (>1500 m).

Source: After Jordano *et al.* (2007).

active and passive processes interacting in migration?

We can gain further insight into the interplay between active and passive processes in the movements of organisms by turning again to migration and to a study that contrasted the nocturnal movements over Sweden of a species of moth and of small birds, tracked by radar (Figure 6.7). For both groups, and for all flying animals, there is a balance to be struck between, literally, going with the flow – allowing the wind to carry them, expending minimal energy, but having little control over direction or ultimate landing site – and on the other hand, flying actively, and expensively, in a 'heading direction' (the way the animal is facing) that combines with the wind direction to carry the animal along a 'track' (its movement in relation to the ground) that is its preferred path. Figure 6.7 shows that both the moths and the birds adopted strategies that were somewhere between these two extremes. For the moths this came closer to going with the flow. They were selective in when they set off, limiting themselves to nights with stronger winds and winds blowing in downwind directions that would take them along their preferred track: northwards in the spring and southwards in the autumn. As a result, they tended to move faster across the ground than the birds, despite the birds' greater intrinsic capacity for fast flight, but they were less accurate than the birds: the moths' track directions were more affected by the winds. The birds, by contrast, much more frequently compensated for the wind. In the autumn especially, when the winds over Sweden predominantly blow to the east, the birds mostly headed to the south-west – but doing so took their tracks to the south. The moths flew less often and less accurately than the birds – but more cheaply.

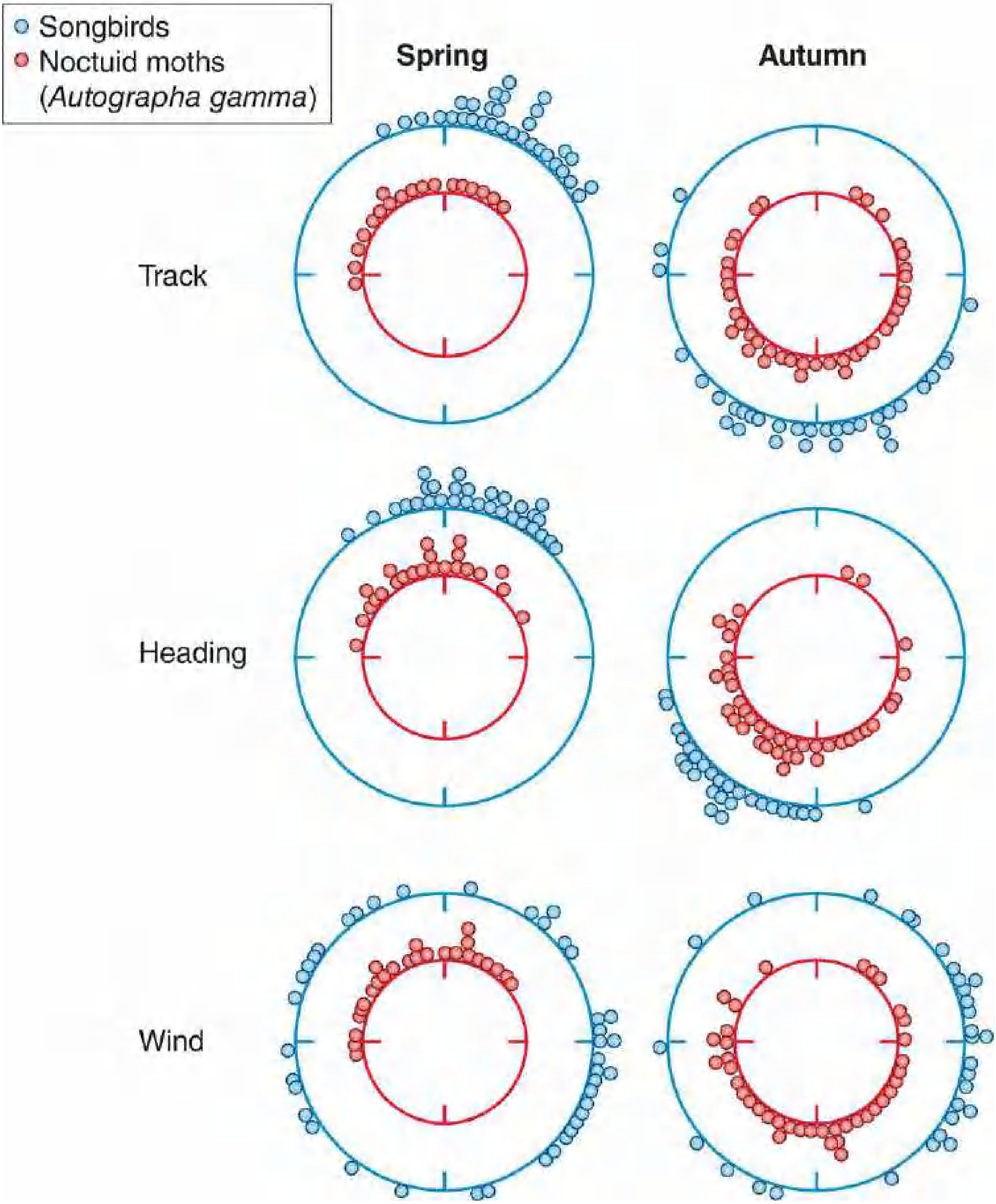


Figure 6.7 Birds and moths reflect contrasting combinations of active and passive processes in their migrations. Overnight movements for songbirds (blue, outer circle) and noctuid moths, *Autographa gamma* (red, inner circle) in Sweden, in spring and autumn. The compass directions of observed movements (dots) are shown for 'Track' movements of the animals (movements in relation to the ground), 'Heading' movements (the direction the animal is facing), and also of 'Wind' flow while the animals are moving.

Source: After Chapman *et al*. (2015).

Many other animals, of course, are land-based in their dispersal and control when they leave, where they travel, and when they stop.

6.3.3 Clonal dispersal

In almost all modular organisms (see Section 4.2.1), an individual genet branches and spreads its parts around it as it grows. There is a sense, therefore, in which a developing tree or coral actively disperses its modules into, and explores, the surrounding environment. The interconnections of such a clone often decay, so that it becomes represented by a number of dispersed parts. This may result ultimately in the product of one zygote being represented by a clone of great age that is spread over great distances. Some clones of the rhizomatous bracken fern (*Pteridium aquilinum*) were estimated to be more than 1400 years old and one extended over an area of nearly 14 ha (Oinonen, 1967).

guerrillas and phalanx-formers

We can recognise a continuum of strategies in clonal dispersal (Lovett Doust & Lovett Doust, 1982). At one extreme, the connections between modules are long and the modules themselves are widely spaced. These have been called 'guerrilla' forms, because they give the plant, hydroid or coral a character like that of a guerrilla army. Fugitive and opportunist, they are constantly on the move, disappearing from some territories and penetrating into others. At the other extreme are 'phalanx' forms, named by analogy with the phalanxes of a Roman army, tightly packed with their shields held around them. Here, the connections between the modules are short, and the organisms expand their clones slowly, retain their original site occupancy for long periods, and neither penetrate readily amongst neighbouring plants nor are easily penetrated by them.

Even amongst trees, it is easy to see that the way in which the buds are placed gives them a guerrilla or a phalanx growth form. The dense packing of shoot modules in species like cypresses (*Cupressus*) produces a relatively undispersed and impenetrable phalanx canopy, whilst many loose-structured, broad-leaved trees (*Acacia*, *Betula*) can be seen as guerrilla canopies, bearing buds that are widely dispersed and shoots that interweave with the buds and branches of neighbours. The twining or clambering lianas in a forest are guerrilla growth forms *par excellence*, dispersing their foliage and buds over immense distances, both vertically and laterally.

The way in which modular organisms disperse and display their modules affects the ways in which they interact with their neighbours. Those with a guerrilla form will continually meet and compete with other species and with other genets of their own kind. With a phalanx structure, however, most meetings will be between modules of a single genet. We see this, for example, for 12 grassland species in Figure 6.8.

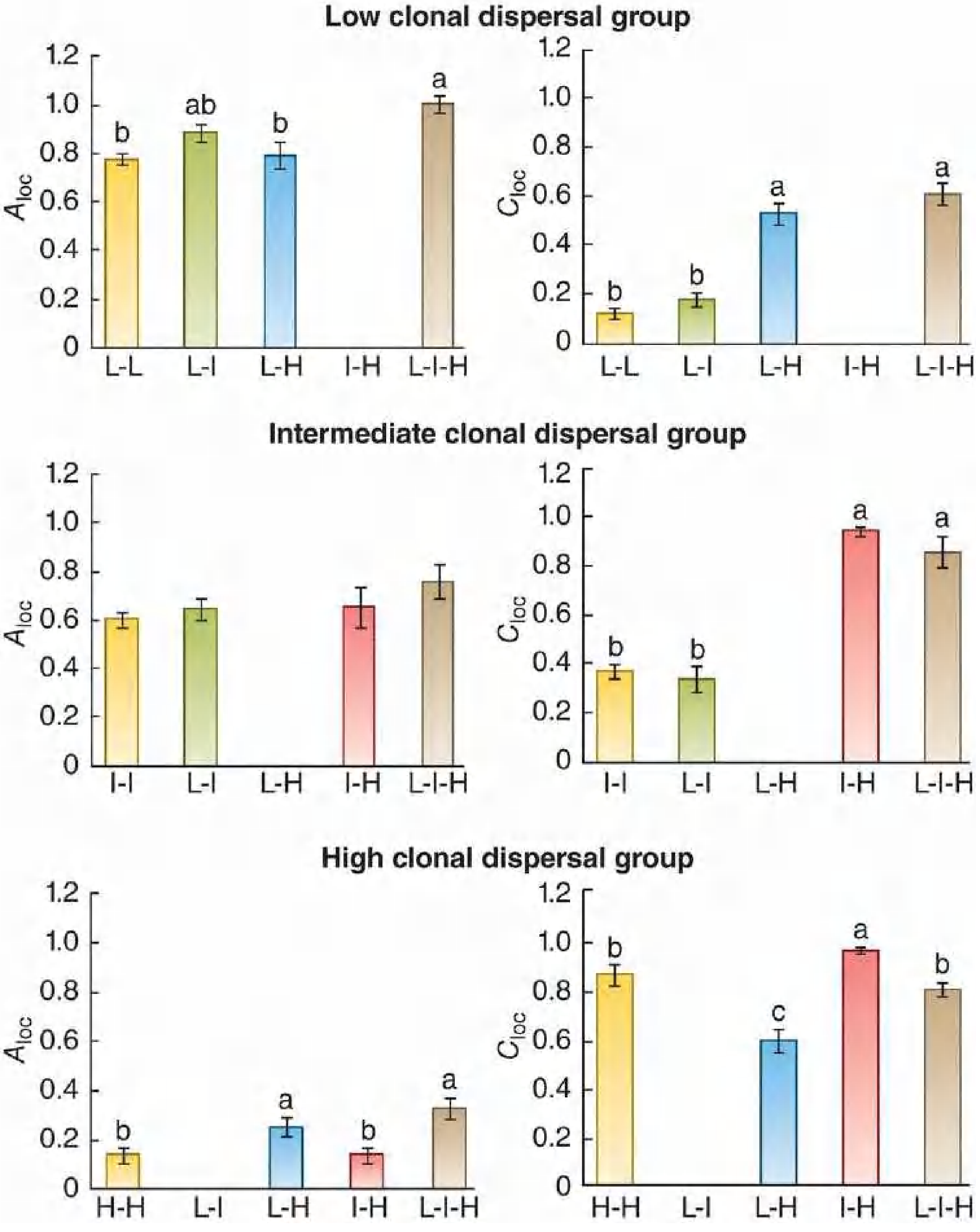

Low clonal dispersal group
A_loc
1.2
1.0
0.8
0.6
0.4
0.2
0
b
ab
b
a
L-L
L-I
L-H
I-H
L-I-H
C_loc
1.2
1.0
0.8
0.6
0.4
0.2
0
b
b
a
a
L-L
L-I
L-H
I-H
L-I-H
Intermediate clonal dispersal group
A_loc
1.2
1.0
0.8
0.6
0.4
0.2
0
I-I
L-I
L-H
I-H
L-I-H
C_loc
1.2
1.0
0.8
0.6
0.4
0.2
0
b
b
a
a
I-I
L-I
L-H
I-H
L-I-H
High clonal dispersal group
A_loc
1.2
1.0
0.8
0.6
0.4
0.2
0
b
a
b
a
H-H
L-I
L-H
I-H
L-I-H
C_loc
1.2
1.0
0.8
0.6
0.4
0.2
0
b
c
a
b
H-H
L-I
L-H
I-H
L-I-H

Figure 6.8 Phalanx-type plants aggregate locally and co-occur little with other species: guerilla-type plants show high levels of clonal dispersal and co-occur often with other species. An index of local aggregation (A_{loc}, left panels) and an index of local co-occurrence (C_{loc}, right panels) for the interactions of 12 grassland plant species. These are grouped into species with low clonal dispersal, L (four species), intermediate, I (four species), and high, H (four species). They interacted with members of their own or other groups as indicated along the horizontal axis. High levels of local aggregation indicate a tendency to disperse (clonally) into volumes of soil, 'cells', nearby. These were observed especially in plants with low clonal dispersal (more phalanx types). High levels of local co-occurrence indicate a tendency to co-occur in cells with other species. These were observed especially in plants with high clonal dispersal (more guerilla types). Standard errors are shown. Bars with different letters above them are significantly different.

Source: After Benot *et al.* (2013).

APPLICATION 6.4 Invasive fragmenting aquatic weeds

Clonal growth is most effective as a means of dispersal in aquatic environments. Many aquatic plants fragment easily, and the parts of a single clone become independently dispersed because they are not dependent on the presence of roots to maintain their water relations. The major aquatic weed problems of the world are caused by plants that multiply as clones and fragment and fall to pieces as they grow: duckweeds (*Lemna* spp.), the water hyacinth (*Eichhornia crassipes*), Canadian pond weed (*Elodea canadensis*) and the water fern *Salvinia*. The biological, chemical and mechanical control methods available to counter this threat are reviewed by Hussner *et al.* (2017).

6.4 Patterns of dispersion

The movements of organisms, whether we classify them as migration or dispersal, have causes and consequences. The proximate (immediate) causes are the triggers that prompt the movements, but as evolutionary biologists, we take it that the ultimate causes are the forces of natural selection that have favoured the consequences we observe. At the individual level, those consequences are the individuals' locations in particular habitats and in relation to other individuals. At the population level, the consequences are the spatial patterns of their distribution – their *dispersion*. We can recognise three main patterns of dispersion, although they form part of a continuum (Figure 6.9).

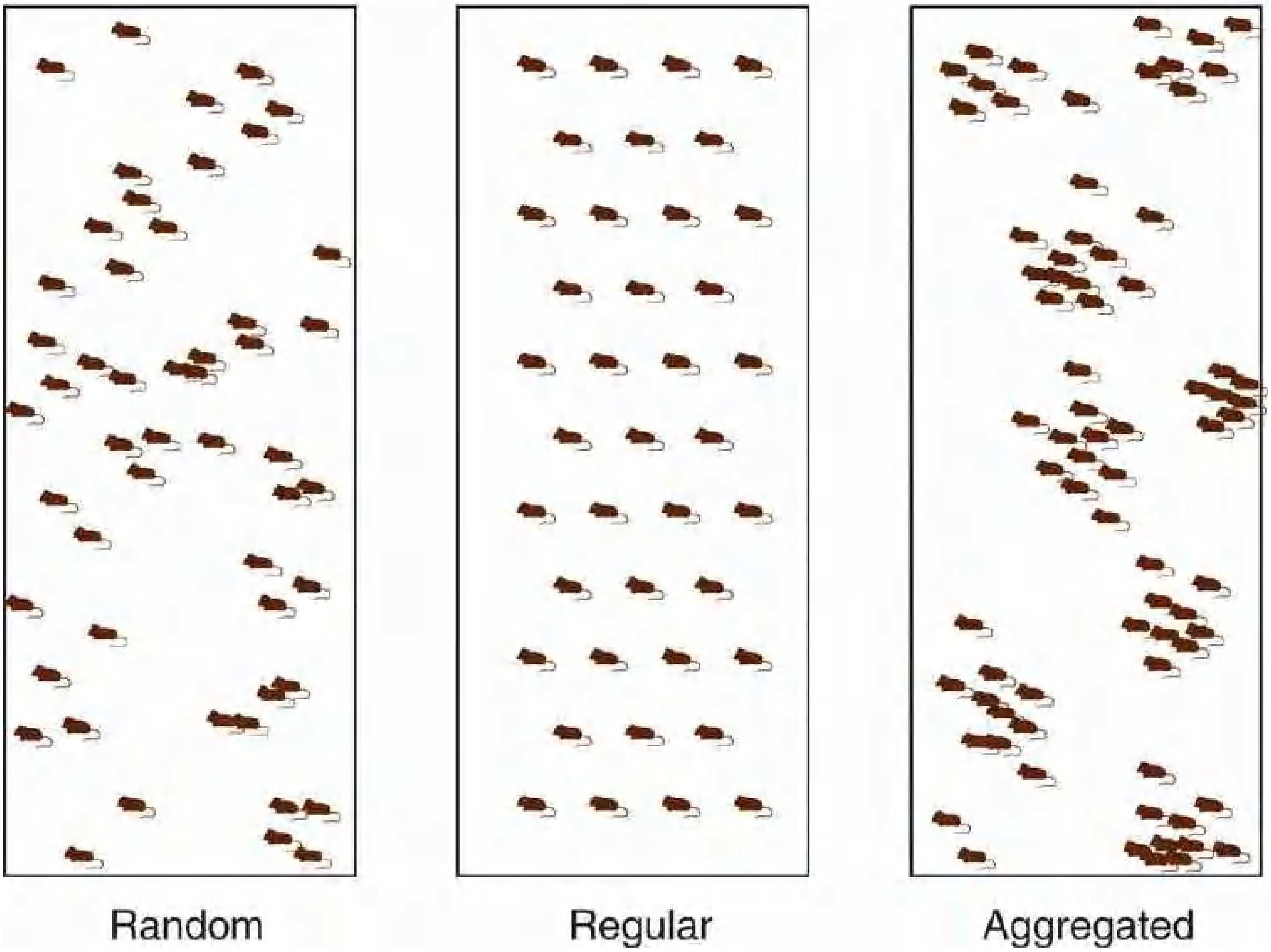


Figure 6.9 Three generalised spatial patterns that may be exhibited by organisms across their habitats.

random, regular and aggregated distributions

Random dispersion occurs when there is an equal probability of an organism occupying any point in space (irrespective of the position of any others). The result is that individuals are unevenly distributed because of chance events.

Regular dispersion (also called a *uniform* or *even* distribution or *overdispersion*) occurs either when an individual has a tendency to avoid other individuals, or when individuals that are especially close to others die. The result is that individuals are more evenly spaced than expected by chance.

Aggregated dispersion (also called a *contagious* or *clumped* distribution or *underdispersion*) occurs either when individuals tend to be attracted to (or are more likely to survive in) particular parts of the environment, or when the presence of one individual attracts, or gives rise to, another individual close to it. The result is that individuals are closer to others than expected by chance.

How these patterns appear to an observer, however, and their relevance to the life of other organisms, depends on the spatial scale at which they are viewed. Consider the distribution of an aphid living on a particular species of tree in a woodland. At a large scale, the aphids will appear to be aggregated in woodlands as opposed to other types of habitat. If samples are smaller and taken only in woodlands, the aphids will still appear to be aggregated, but now on their host tree species rather than on trees in general. However, if samples are smaller still (25 cm^2, about the size of a leaf) and are taken within the canopy of a single tree, the aphids might appear to be randomly distributed over the tree as a whole. And at an even smaller scale (*c*.1 cm^2), we might detect a regular distribution because individual aphids on a leaf avoid one another.

6.4.1 Patchiness

fine- and coarse-grained environments

At some scale, at least, all habitats are patchy. Patchiness may be a feature of the physical environment: islands surrounded by water, rocky outcrops in a moorland, and so on. Equally important, patchiness may be created by the activities of organisms themselves; by their grazing, the deposition of dung, or the local depletion of water and mineral resources. It is crucial that we describe dispersion within and across patches at scales that are relevant to the lifestyle of the organisms concerned. MacArthur and Levins (1964) introduced the concept of environmental *grain* to make this point. For example, from the point of view of a bird like the scarlet tanager (*Piranga olivacea*) that forages indiscriminately in both oaks and hickories, the canopy of an oak–hickory forest is *fine grained*. It is patchy, but the birds experience the habitat as an oak–hickory mixture. The habitat is *coarse grained*, however, for defoliating insects that attack *either* oaks *or* hickories preferentially: they experience the habitat one patch at a time, moving from one preferred patch to another (Figure 6.10).

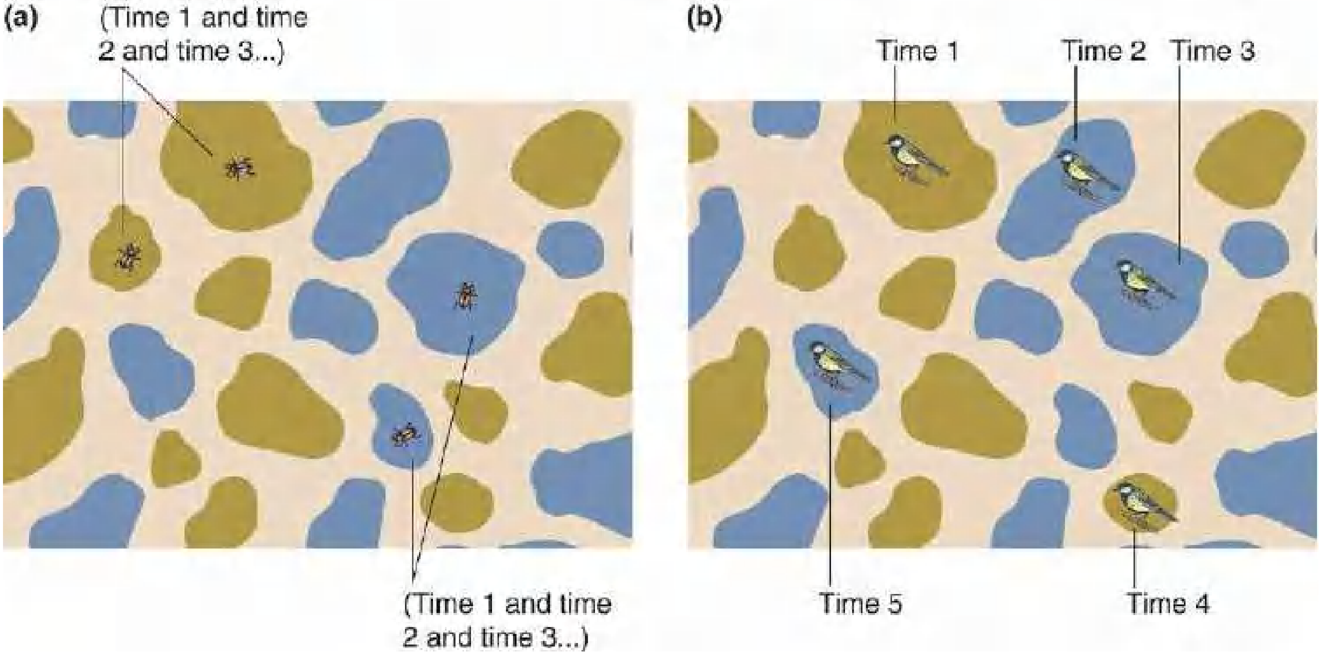


Figure 6.10 The ‘grain’ of the environment must be seen from the perspective of the organism concerned. (a) An organism that is small or moves little is likely to see the environment as coarse grained: it experiences only one habitat type within the environment for long periods or perhaps all of its life. (b) An organism that is larger or moves more may see the same environment as fine grained: it moves frequently between habitat types and hence samples them in the proportion in which they occur in the environment as a whole.

turnover and invasion dispersal

Patchiness, dispersal and scale are tied intimately together. Thus, it is useful to distinguish between *turnover dispersal* and *invasion dispersal* (Bullock *et al.*, 2002) and to do so at local and landscape scales (though what is ‘local’ to a worm is very different from what is local to the bird that eats it). Turnover dispersal at the local scale refers to the movement of an individual into a gap from occupied habitat immediately surrounding the gap. But that gap may also be ‘invaded’ by individuals moving in from elsewhere in the surrounding community. Similarly, at the landscape scale, dispersal may be part of an on-going turnover of extinction and recolonisation of occupiable patches within an otherwise unsuitable habitat matrix (see Section 6.7 on metapopulation dynamics), or there may be invasion of an area by dispersal of a ‘new’ species expanding its range.

6.4.2 Forces favouring aggregation

The simplest evolutionary explanation for the patchiness of populations is that organisms aggregate where they find the habitat patches that are most favourable for reproduction and survival. It pays them (and has paid them in evolutionary time) to disperse to these patches. There are, however, other specific ways in which organisms may gain from being close to neighbours.

aggregation and the selfish herd

A classic theory identifying a selective advantage to individuals that aggregate with others was proposed by Hamilton (1971) in his paper ‘Geometry for the selfish herd’. He argued that the risk to an individual from a predator may be lessened if it places another potential prey individual between itself and the predator. The consequence of many individuals doing this is that they form an aggregation in space. The principle of the selfish herd is also pertinent to the aggregated (synchronous) appearance of organisms in time. We return to this question of so-called ‘mast’ years in Section 10.2.4.

philopatry

Individuals may also tend to aggregate, or at least not disperse away from one another, if they exhibit philopatry or ‘home-loving’ behaviour (Lambin *et al.*, 2001). This can come about because there are advantages of inhabiting a familiar environment, or individuals may cooperate with, or at least be prepared to tolerate, related individuals in the natal habitat that share a high proportion of their genes, or dispersers may face intolerance or aggression from groups of unrelated individuals (Hestbeck, 1982). Lambin and Krebs (1993), for example, found in Townsend’s voles, *Microtus townsendii*, in Canada, that the nests of females that were first degree relatives (mother–daughters, littermate sisters) were closer than those that were second degree relatives (non-littermate sisters, aunt–nieces), which were closer than those that were more distantly related, which in turn were closer than those not related at all. *M. townsendii* also provides an example where the fitness advantages of philopatry have been confirmed. Lambin and Yoccoz (1998) manipulated the relatedness of groups of breeding females, mimicking either a situation where the population had experienced philopatric recruitment followed by high survival (‘high kinship’), or where the population had experienced either low philopatric recruitment or high mortality of recruits (‘low kinship’). Survival of pups, especially early in their life, was significantly higher in the high kinship than in the low kinship treatment.

sociality and cooperation

Individuals may also gain from living in groups (Krause & Ruxton, 2002) if this helps to locate food, or gives warning of predators, or if it pays for individuals to join forces in fighting off a predator (Figure 6.11); and there are also many species – estimated to be up to 9% of all bird species, for example – that exhibit cooperative breeding, often arising out of philopatry amongst offspring who remain on their natal territory and help their parents raise subsequent broods (Hatchwell, 2009).

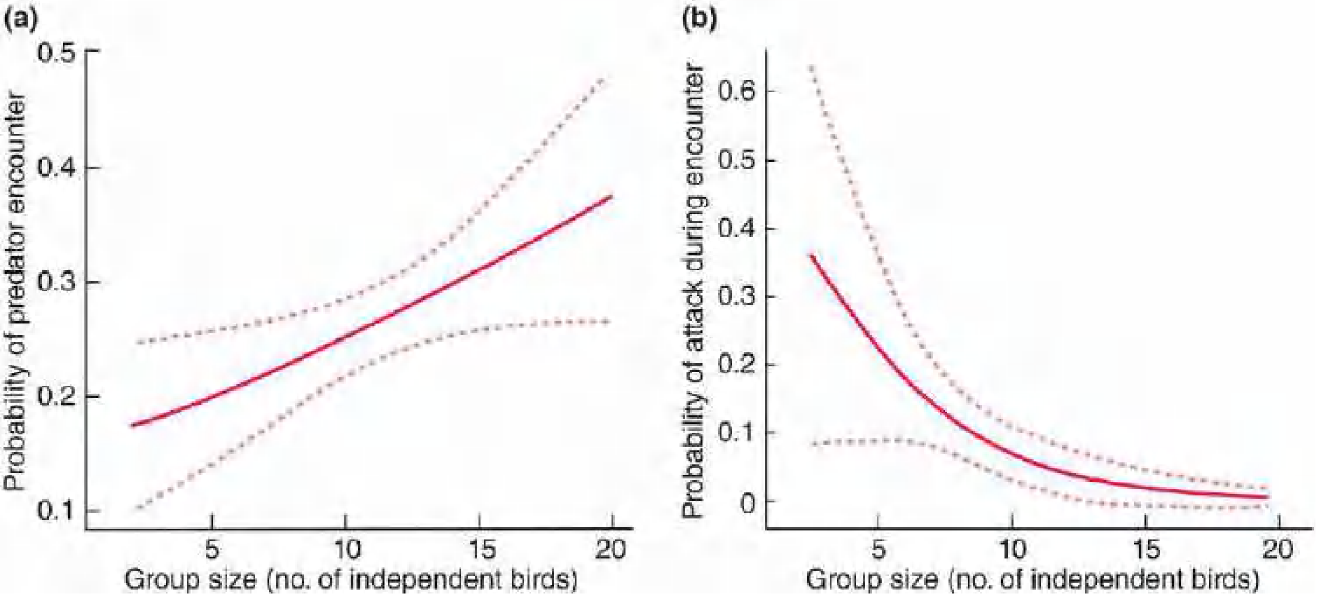


Figure 6.11 Group living protects against predation in the chestnut-crowned babbler, *Pomatostomus ruficeps*, in New South Wales, Australia. (a) Predators (various falcons and hawks) were more likely to encounter (i.e. be attracted to) larger groups of babblers. (b) But despite this, predators were less likely to attack babblers in larger groups. In each case, lines are from models fitted to the data after other explanatory variables have been taken into account, with standard errors shown by dotted lines.

Source: After Sorato *et al.* (2012).

6.4.3 Forces diluting aggregations: density-dependent dispersal

On the other hand, there are strong selective pressures that can act against aggregation. The foremost of these is undoubtedly the more intense competition suffered by crowded individuals (see Chapter 5) and the direct interference between such individuals even in the absence of a shortage of resources. We see this, for example, in the even spacing that results from territoriality (Section 5.8.4).

inbreeding and outbreeding

Another important driver diluting aggregations is not so much ecological as evolutionary. When closely related individuals breed, their offspring are likely to suffer an ‘inbreeding depression’ in fitness (Charlesworth & Charlesworth, 1987), especially as a result of the expression in the phenotype of recessive deleterious alleles. With limited dispersal, inbreeding becomes more likely, and inbreeding avoidance is thus a force favouring dispersal, particularly natal dispersal. On the other hand, many species show local adaptation to their immediate environment (see Section 1.2). Longer distance dispersal may therefore bring together genotypes adapted to different local environments, which on mating give rise to low-fitness offspring adapted to neither habitat. This is called ‘outbreeding depression’, resulting from the break-up of co-adapted combinations of genes – a force acting against dispersal. The situation is complicated by the fact that inbreeding depression is most likely amongst populations that normally outbreed, since inbreeding itself will purge populations of their deleterious recessives. Nonetheless, natural selection can be expected to favour a pattern of dispersal that is in some sense intermediate – maximising fitness by avoiding both inbreeding and outbreeding depression, though these will clearly not be the only selective forces acting on dispersal. Certainly, there are several examples in plants of inbreeding and outbreeding depression when pollen is transferred from either close or distant donors, and in some cases both effects can be demonstrated in a single experiment. For example, when larkspur (*Delphinium nelsonii*) offspring were generated by hand pollinating with pollen brought from 1, 3, 10 and 30 m to the receptor flowers (Figure 6.12), both inbreeding and outbreeding depression in fitness were apparent.

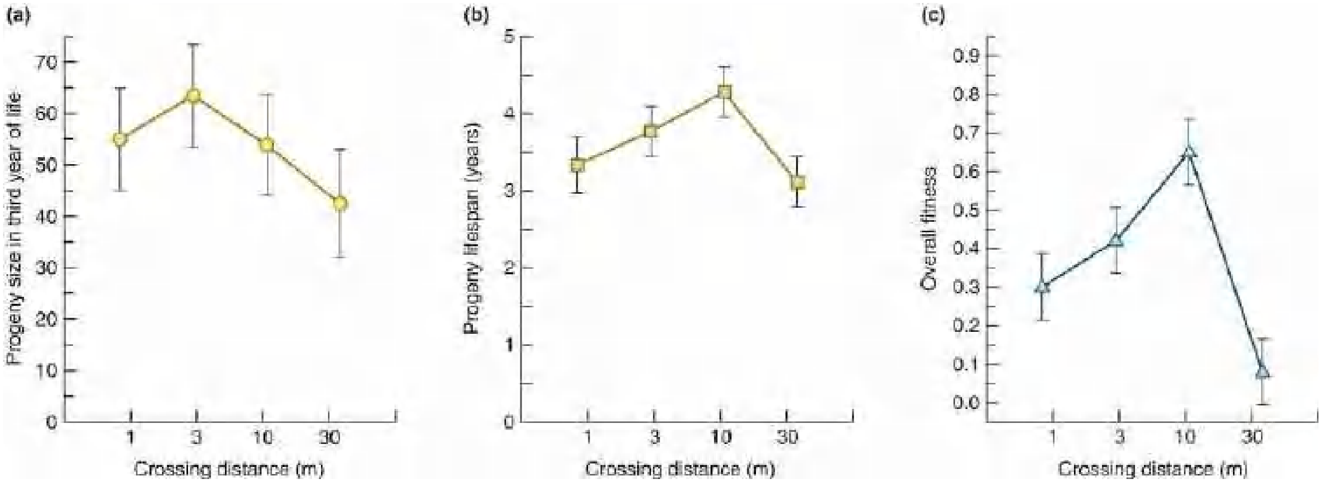


Figure 6.12 **Inbreeding and outbreeding depression in *Delphinium nelsonii*.** (a) Progeny size in the third year of life, (b) progeny lifespan and (c) the overall fitness of progeny cohorts were all lower when progeny were the result of crosses with pollen taken close to (1 m) or far from (30 m) the receptor plant. Bars are SEs.

Source: After Waser & Price (1994).

avoiding kin competition

Dispersal of offspring away from their close relatives may also be favoured because it decreases the likelihood of competitive effects being directed at close kin. This was explained in a classic modelling paper by Hamilton and May (1977) (see also Gandon & Michalakis, 2001). In a population of stay-at-home non-dispersers, a rare mutant disperser that keeps some offspring at home but commits others to dispersal will suffer no competition in its own patch from non-dispersers, but it will compete against those non-dispersers in their home patches. Dispersers will thus direct much of their competitive effects at non-dispersing non-kin, while non-dispersers direct all of their competition at their own kin. The propensity to disperse will therefore increase in frequency in the population. On the other hand, if the majority of the population are dispersers, a rare mutant non-disperser will again do worse, since it can never displace any of the dispersers from their patches but has itself to contend with dispersers from outside. Dispersal is therefore said to be an evolutionarily stable strategy (ESS) (Maynard Smith, 1972 ; Parker, 1984) – it increases when rare but cannot be displaced when common. Hence, the avoidance of both inbreeding and kin competition seem likely to give rise to higher emigration rates at higher densities, when these forces are most intense.

There is indeed evidence for kin competition playing a role in driving offspring away from their natal habitat, but much of it is indirect (Lambin *et al.*, 2001). In some cases, though, it has been demonstrated experimentally, for example in the spider mite, *Tetranychus urticae*, where individuals not only dispersed further, on average, when surrounded by more related individuals, but also had a dispersal distribution in which more individuals dispersed the greatest distances (Figure 6.13).

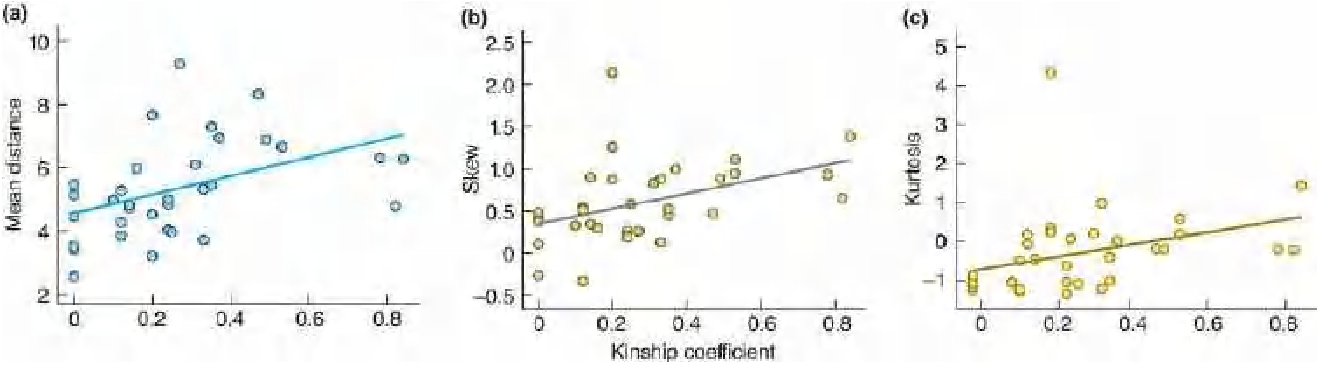


Figure 6.13 Kin competition may drive offspring away from their natal habitat. The effect of the relatedness of animals around them (kinship coefficient) on dispersal in the mite, *Tetranychus urticae*. As relatedness increased (a) mean distance dispersed increased ($F = 8.51$, $P < 0.01$), (b) the distribution was more skewed towards longer distances and fewer near the mean ($F = 7.94$, $P < 0.01$), and (c) kurtosis was greater: there were more individuals towards the tail of the distribution and less near the mean ($F = 4.74$, $P < 0.05$).

Source: After Bitume *et al*. (2013).

density-dependent dispersal

Putting these considerations together, we can see that the types of distribution over available patches found in nature are bound to be evolved compromises between forces attracting individuals towards one another and those provoking them to disperse apart. We return later to the 'ideal free distribution' and other spatial distributions that may result from this (Section 10.4.3). We can also see that if dispersal occurs to avoid competition, then it should occur at greater rates at higher densities. That is, we should see density-dependent dispersal. If, however, dispersal occurs to avoid kin competition in particular, then we should especially expect to see density-dependent *natal* dispersal, and we should see it in response to overall kin densities, irrespective of sex. But if density-dependent natal dispersal occurs to avoid inbreeding, then we would expect to see males responding in particular to the densities of females, and females to the densities of males. There are certainly examples of density-dependent emigration, both as regards a tendency to disperse and the distance dispersed (Figure 6.14). But we need to acknowledge the alternative drivers behind such patterns, and to see how these might combine with opposing forces like philopatry, that hold individuals together. Thus, we must recognise the importance of distinguishing between the dispersal patterns of males and females, and of the old and the young, which we do in the next section.

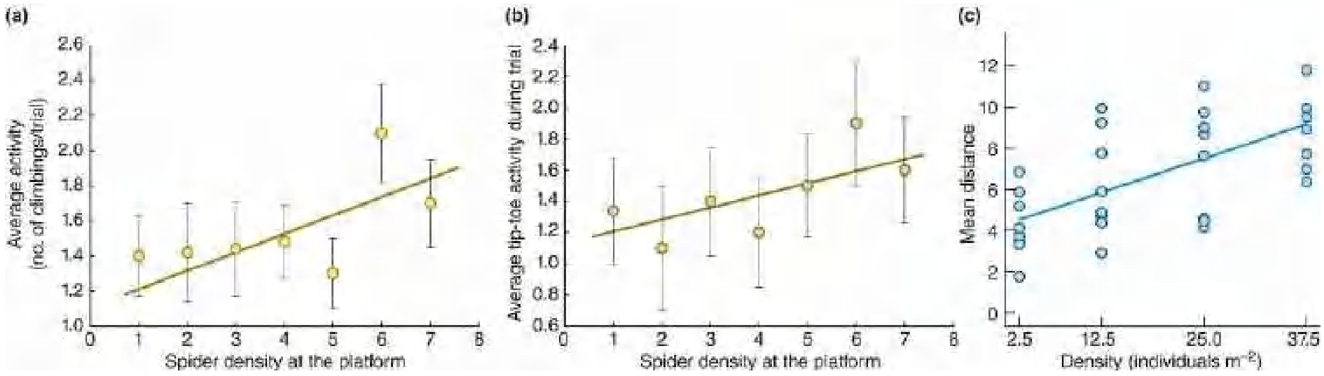


Figure 6.14 **Density-dependent emigration in spiders and mites.** In laboratory experiments, spiders, *Erigone atra*, emigrated from platforms in a wind tunnel, tethered by gossamer threads, at a greater rate at higher densities, as evidenced by (a) their tendency to climb vertical sticks from where the wind carried them away ($F = 11.27$, $P < 0.001$), and (b) exhibiting 'tip-toeing' behaviour which made take-off more likely ($F = 4.77$, $P < 0.05$). In both cases, standard errors are shown. (c) The mean distance dispersed by the mite, *Tetranychus urticae*, increased with initial density in the experiment also shown in Figure 6.13 ($F = 9.61$, $P < 0.001$).

Source: (a, b) After De Meester & Bonte (2010). (c) After Bitume *et al*. (2013).

6.5 Variation in dispersal within populations

6.5.1 Dispersal polymorphism

One source of variability in dispersal within populations is polymorphism ('many types') amongst the progeny of a single parent. This is typically associated with habitats that are variable or unpredictable. A number of plants, for example, produce two different sorts of seed or fruits – *heterocarpy* – one of them dispersive and the other stay-at-home; and worldwide, around 50 of these species exhibit *amphicarpy*, where the same plant may either produce subterranean flowers at its home location that self-pollinate and then germinate underground, usually early in the season, or aerial flowers that cross-pollinate and produce fruits for long-range dispersal. One example is the devil's thorn, *Emex spinosa*, an annual originally from semiarid regions in the Old World that has now spread globally. The dry, aerial one-seeded fruits ('achenes') are spiny, buoyant and light and hence adapted for dispersal by animals, water and wind. The subterranean achenes, by contrast, lack spines and are never shed by the mother but are more competitive than their aerial counterparts. Greenhouse experiments show that the ratio of aerial to subterranean achene mass is greater at lower densities but at higher nutrient concentrations (Figure 6.15). This supports a view of subterranean production as part of a 'bad-year' strategy: when the plants are crowded and short of resources, they stay at home, investing more of their (scarce) resources safely where the past suggests they may thrive. High investment in aerial achenes, then, constitutes a dispersive, 'good-year' strategy of relatively risky investment that may, however, yield the benefit of the establishment of a new population 'elsewhere'.

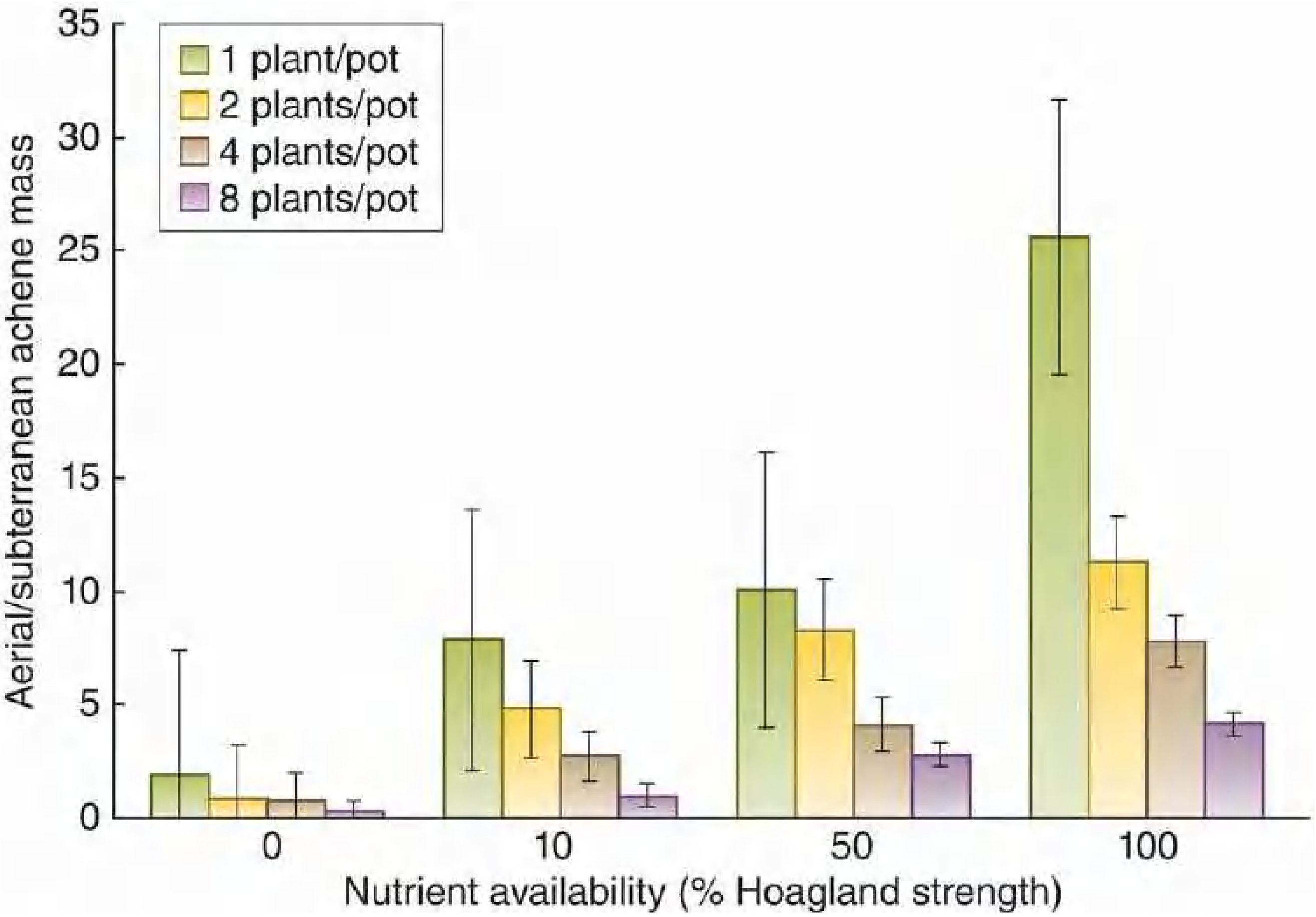


Figure 6.15 **Plants staying at home in bad years and dispersing in good years.** Investment by the annual plant, *Emex spinosa*, in dispersive aerial fruits (achenes) as opposed to stay-at-home subterranean ones was greater at lower densities ($F = 8.59$, $P < 0.001$) and higher nutrient concentrations ($F = 6.52$, $P < 0.001$). Means and SEs are shown.

Source: After Sadeh *et al*. (2009).

A dimorphism of dispersers and non-dispersers (winged and wingless progeny), or of more- and less-dispersive offspring, is also a common phenomenon amongst insects. The pea aphid, *Acyrthosiphon pisum*, for example, produces more winged morphs in the presence of predators (Figure 6.16a), presumably as an escape response from an adverse environment; while the parasitoid wasp, *Melittobia digitata*, produces more of its more dispersive long-winged morph from larger, more crowded clutches, again presumably as an escape response (Figure 6.16b).

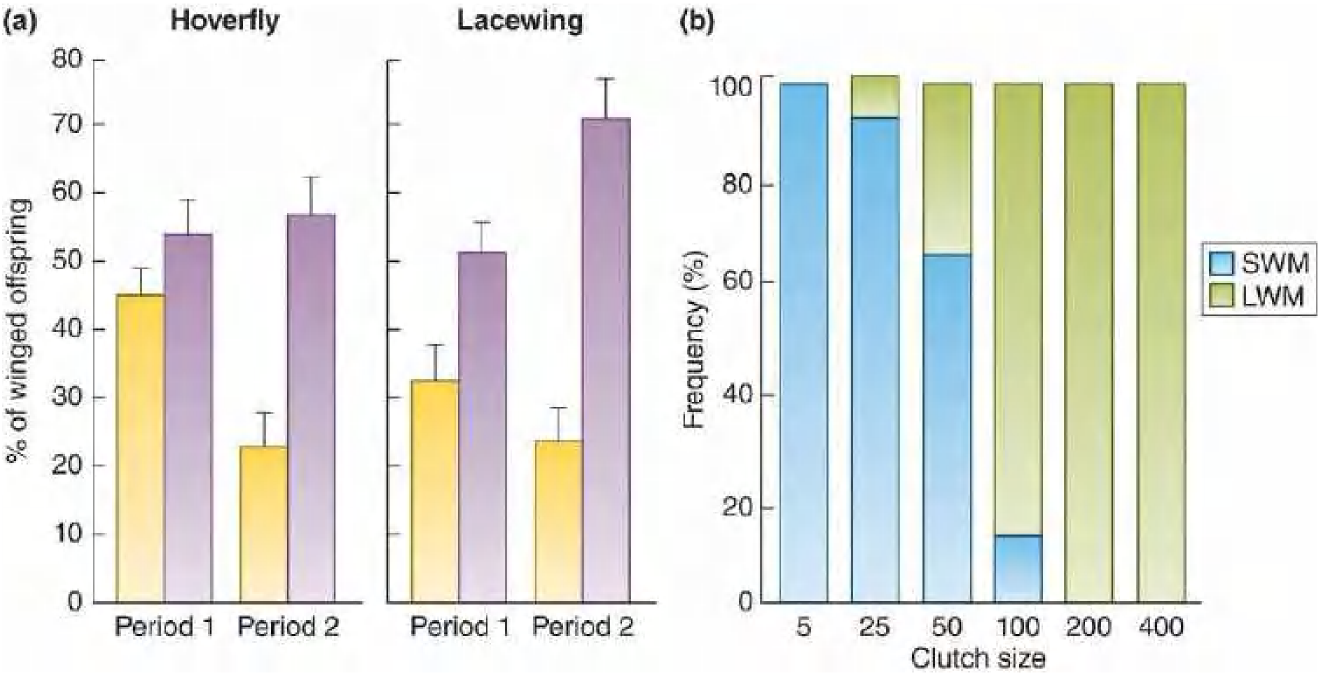


Figure 6.16 **Dispersal polymorphisms.** (a) The mean proportion (± SE) of winged morphs of the pea aphid, *Acyrthosiphon pisum*, produced after two separate periods of exposure to each of two predators: left: hoverfly larvae, and right: lacewing larvae. Purple bars, predator treatment; yellow bars, control. (b) The percentage, in the parasitoid wasp, *Melitobia digitata*, of the relatively dispersive long-winged morph (LWM) at the expense of the short-winged morph (SWM) increased in larger (more crowded) clutches of eggs ($X^2 = 446.2$, $P < 0.001$).

Source: (a) After Kunert & Weisser (2003). (b) After Consoli & Vinson (2002).

6.5.2 Sex- and age-related differences

We have noted already that much dispersal is natal, involving the movement of an organism from its place of birth to its place of first breeding. This age differential is linked intimately to sex differences in dispersal because the forces shaping dispersal – philopatry, avoidance of kin competition or inbreeding, and so on – often themselves act differently on the two sexes. In a seminal paper that set the scene for much of the subsequent work in this area, Greenwood (1980) contrasted the sex-biased natal dispersal of birds and mammals. He noted that many birds are socially monogamous (pairs breed only with one another, at least within each breeding season) and that the males establish and defend territories (ideally in areas with which they are familiar) to which females are then attracted. He argued that it was for this reason that females were the dispersive sex far more often than the philopatric males. Many mammals, on the other hand, are polygynandrous (both males and females have multiple mates over the course of a breeding season), with males defending access to groups of females. Hence, here it is the males that disperse more in order to give themselves the opportunity of obtaining this most crucial resource. Of course, the real distinction is not between birds and mammals but between the contrasting lifestyles and selection pressures that they may be said to typify. An entirely equivalent argument can be applied to monogamous and polygynandrous mammals (Dobson, 2013). A review of sex-biased dispersal covering a much wider (animal) taxonomic range (Trochet *et al.*, 2016) confirmed the broad distinction between birds and mammals, but showed that overall, male and female bias are more or less equally common (Figure 6.17). Analysis of the traits associated with sex bias provided some support for Greenwood's hypothesis (an association of the direction of sex bias with the mating system and territoriality) but in fact the link was more strongly related to the type of parental care and sexual dimorphism.

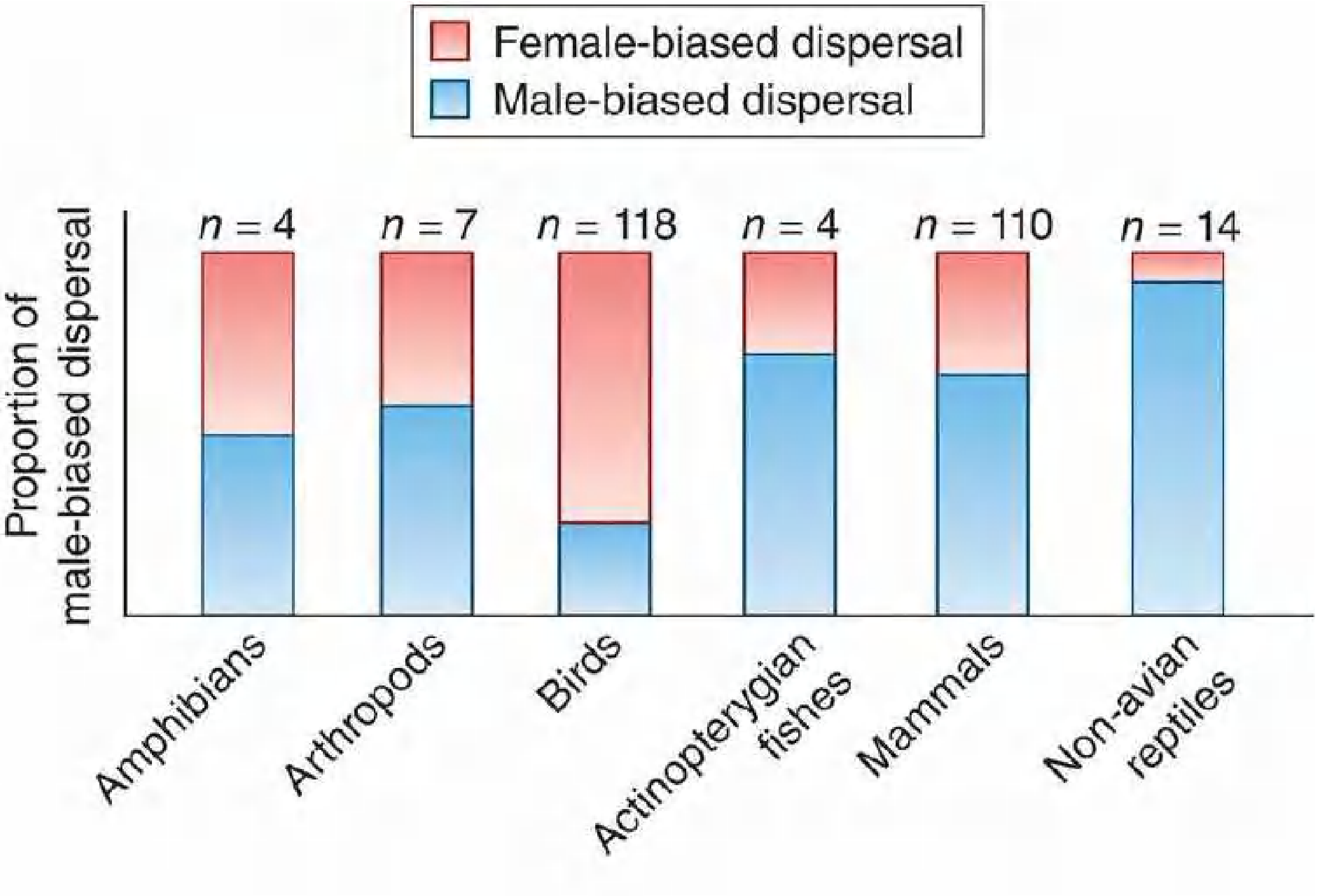


Figure 6.17 The proportion of male-biased as opposed to female-biased dispersal in studies carried out on different taxonomic groups.

Source: After Trochet *et al.* (2016).

The more important, general point, therefore, is that dispersal differentials will reflect differences, however they arise, in the costs and benefits of philopatry, the nature of the key resource being competed for, and so on. These are set against a background of dispersal differences between the sexes being favoured in their own right as a means of reducing inbreeding (Lambin *et al.*, 2001 ; Dobson, 2013), though of course, if individuals can recognise their own kin, they can avoid breeding with them without the need for differential dispersal. Hence, alongside the common themes there are many apparent 'special cases'. The superb starling, *Lamprotorbis superbis*, for example, studied in Kenya, is a cooperative breeder in which philopatric males remain in the natal territory to help rear their parents' subsequent broods while their female equivalents disperse, thereby reducing the possibilities of inbreeding but also of cooperation between female kin. In fact, however, 26% of dispersed females occurred in a group with a full sister, and these individuals were significantly more likely to breed, suggesting that the sisters formed some kind of alliance (Pollack & Rubenstein, 2015). In another example, in Tanzania, spotted hyena males, *Crocuta crocuta*, could be either philopatric, remaining with their natal clan, or dispersers, despite the fact that a well-developed female mate-choice in the species would be expected to drive male-biased dispersal in an archetypal mammalian pattern. In practice, though, dispersal decisions appeared much more pragmatic and opportunistic. Males choosing to be philopatric entered the social hierarchy near the top, whereas dispersers had to work their way up over several years. As a consequence, philopatric males mated more often than dispersers did with the highest-ranking females, which compensated for their lower mating rate with other females, leading to similar fitnesses of the two types overall (Davidian *et al.*, 2016). There is undoubtedly a shared list of drivers of age- and sex-related patterns of dispersal, but an enormous variety of ways in which these drivers can be combined.

6.6 The demographic significance of dispersal

Dispersal can have a potentially profound effect on the dynamics of populations. In practice, however, many studies have paid little attention to dispersal. The reason often given is that emigration and immigration are approximately equal, and they therefore cancel one another out. One suspects, though, that the real reason is that dispersal is usually extremely difficult to quantify.

6.6.1 Dispersal and the demography of single populations

The studies that have looked carefully at dispersal have tended to bear out its importance. In a long-term and intensive investigation of a population of great tits, *Parus major*, near Oxford, UK, it was observed that 57% of breeding birds were immigrants rather than born in the population (Greenwood *et al*., 1978). And in many cases the rapid spread of species into new areas, including economically important pests, is a compelling testament to the power of dispersal in determining the abundance we observe (see, for example, Figure 6.18).

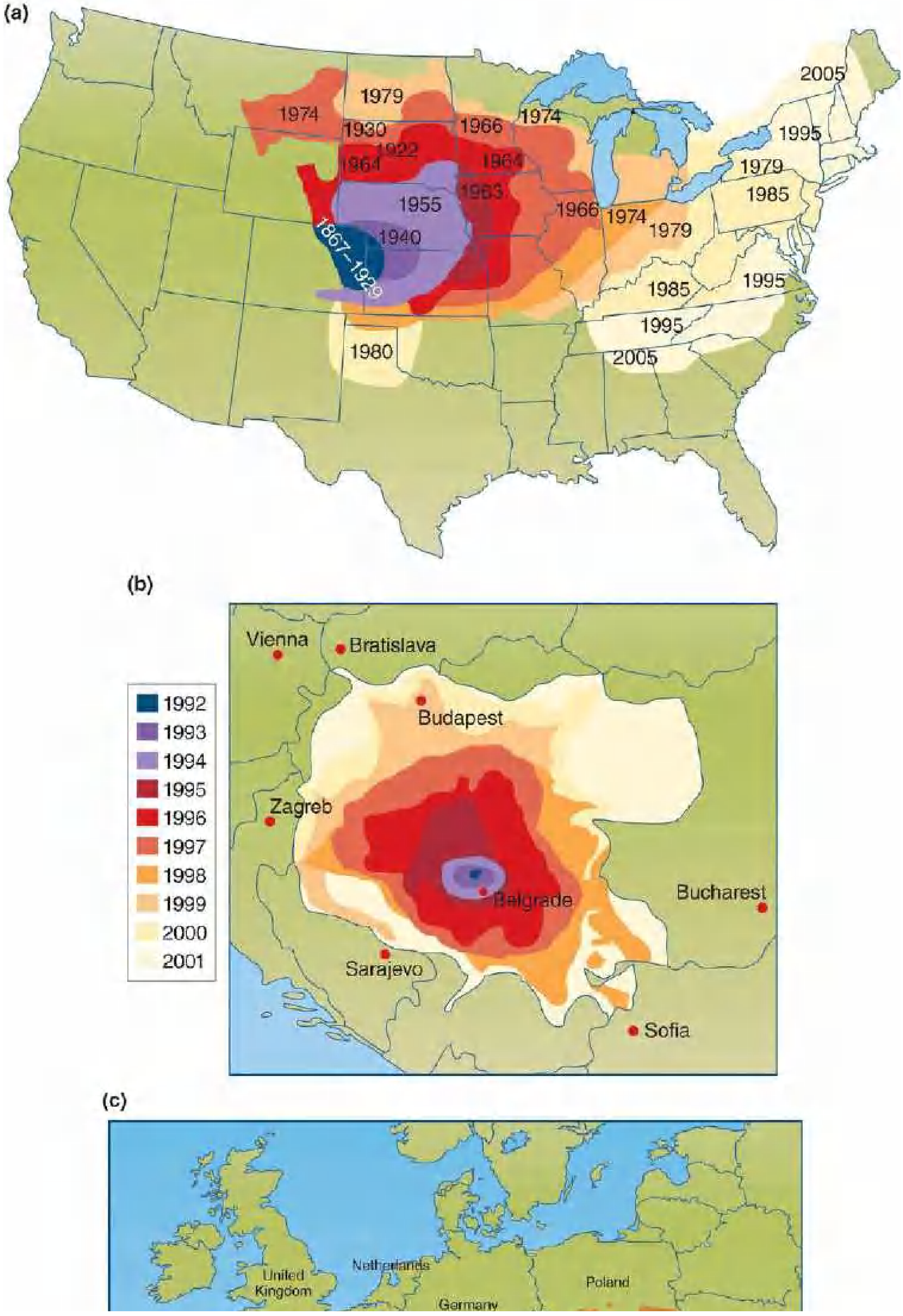

(a)
2005
1979
1974
1930
1966
1974
1995
1922
1964
1964
1979
1955
1963
1985
1867–1929
1940
1966
1974
1979
1985
1995
1995
1980
2005
(b)
Vienna
Bratislava
Budapest
Zagreb
Belgrade
Bucharest
Sarajevo
Sofia
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
(c)
United Kingdom
Netherlands
Germany
Poland

Figure 6.18 The rapid spread of the western corn rootworm, *Diabrotica virgifera virgifera*, one of the most important pests of maize worldwide, (a) in the USA following its break-out from a stronghold around eastern Nebraska in the 1940s; (b) in south-eastern Europe, from 1992 to 2001, following its probable introduction from material carried by plane to Belgrade airport; and (c) to elsewhere in Europe by 2007.

Source: (a, c) After Gray *et al.* (2009). (b) After European Environment Agency (2002).

A profound effect of dispersal on the dynamics of a population was seen in a study of *Cakile edentula*, a summer annual plant growing on the sand dunes of Martinique Bay, Nova Scotia. The population was concentrated in the middle of the dunes, and declined towards both the sea and the land. Only in the area towards the sea, however, was seed production high enough and mortality sufficiently low for the population to maintain itself year after year. At the middle and landward sites, mortality exceeded seed production. Hence, one might have expected the population to become extinct (Figure 6.19). But the distribution of *Cakile* did not change over time. Instead, large numbers of seeds from the seaward zone dispersed to the middle and landward zones. Indeed, more seeds were dispersed into and germinated in these two zones than were produced by the residents. The distribution and abundance of *Cakile* were directly due to the dispersal of seeds in the wind and the waves.

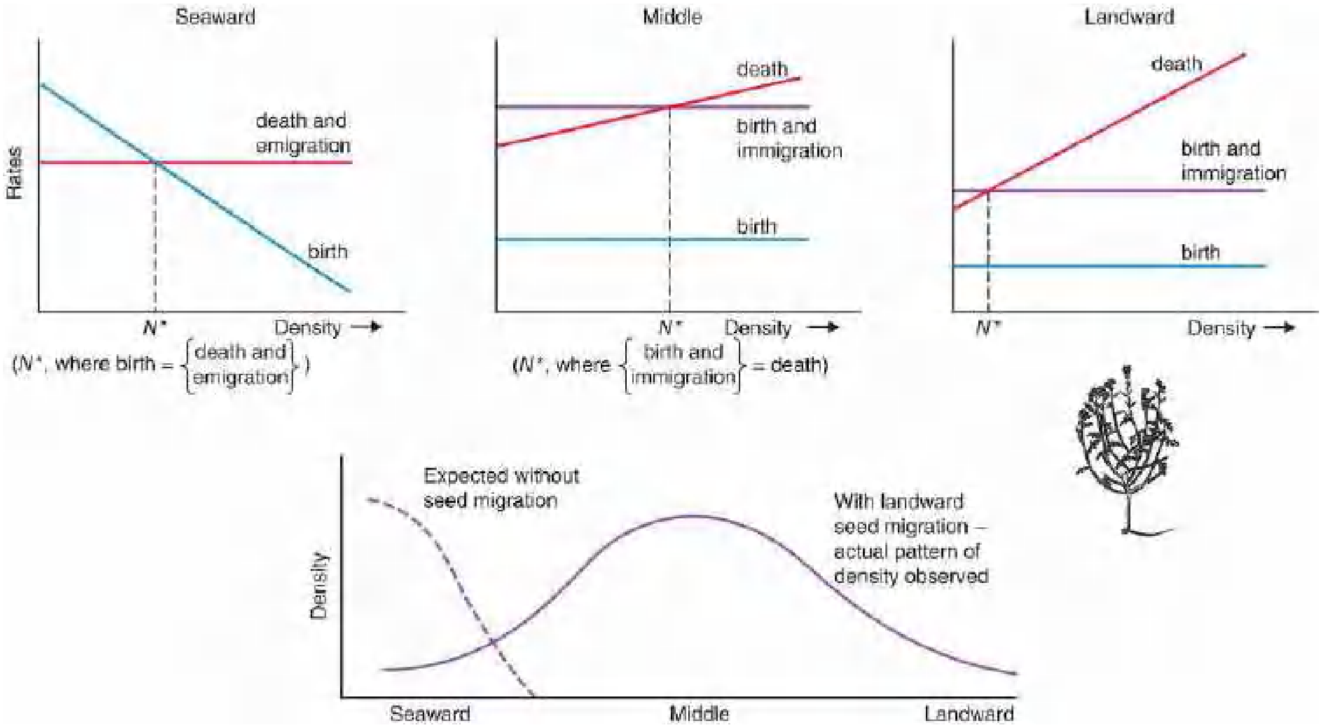


Figure 6.19 **Dispersal drives the local distribution of a sand-dune plant.** Diagrammatic representation of variations in mortality and seed production of *Cakile edentula* in three areas along an environmental gradient from open sand beach (seaward) to densely vegetated dunes (landward). In contrast to other areas, seed production was prolific at the seaward site. Births, however, declined with plant density, and where births and deaths were equal, an equilibrium population density can be envisaged, *N**. In the middle and landward sites, deaths always exceeded births resulting from local seeds, but populations persisted there because of the landward drift of the majority of seed produced by plants on the beach (seaward site). Thus, the sum of local births plus immigrating seeds can balance mortality in the middle and landward sites, resulting in equilibria at appropriate densities.

APPLICATION 6.5 Habitat restoration for a declining squirrel population

We can see the importance of dispersal, too, in the development of plans to conserve populations of flying squirrels (*Pteromys volans*), which have declined dramatically since the 1950s in Finland, mainly because of habitat loss, habitat fragmentation and reduced habitat connectivity associated with intensive forestry practices. Areas of natural forest are now separated by clear-cut and regenerating areas. The core breeding habitat of the flying squirrels only occupies a few hectares, but individuals, particularly males, move to and from this core for temporary stays in a much larger ‘dispersal’ area (1–3 km^2), and juveniles permanently disperse within this range. Reunanen *et al.* (2000) compared the landscape structure around known flying squirrel home ranges (63 sites) with randomly chosen areas (96 sites) to determine the forest patterns that favour the squirrels. They first established that landscape patch types could be divided into optimal breeding habitat (mixed spruce–deciduous forests), dispersal habitat (pine and young forests) and unsuitable habitat (young sapling stands, open habitats, water). Figure 6.20 shows the amount and spatial arrangement of the breeding habitat and dispersal habitat for examples of a typical flying squirrel site and a random forest site. Overall, flying squirrel landscapes contained three times more suitable breeding habitat within a 1 km radius than random landscapes. Crucially, however, squirrel landscapes also contained about 23% more dispersal habitat than random landscape, and significantly, squirrel dispersal habitat was much better connected (fewer fragments per unit area) than random landscapes. Reunanen *et al.* (2000) recommend that forest managers should restore and maintain a deciduous mixture, particularly in spruce-dominated forests, for optimal breeding habitat. But they also need to ensure good physical connectivity between the optimal squirrel breeding and dispersal habitats.

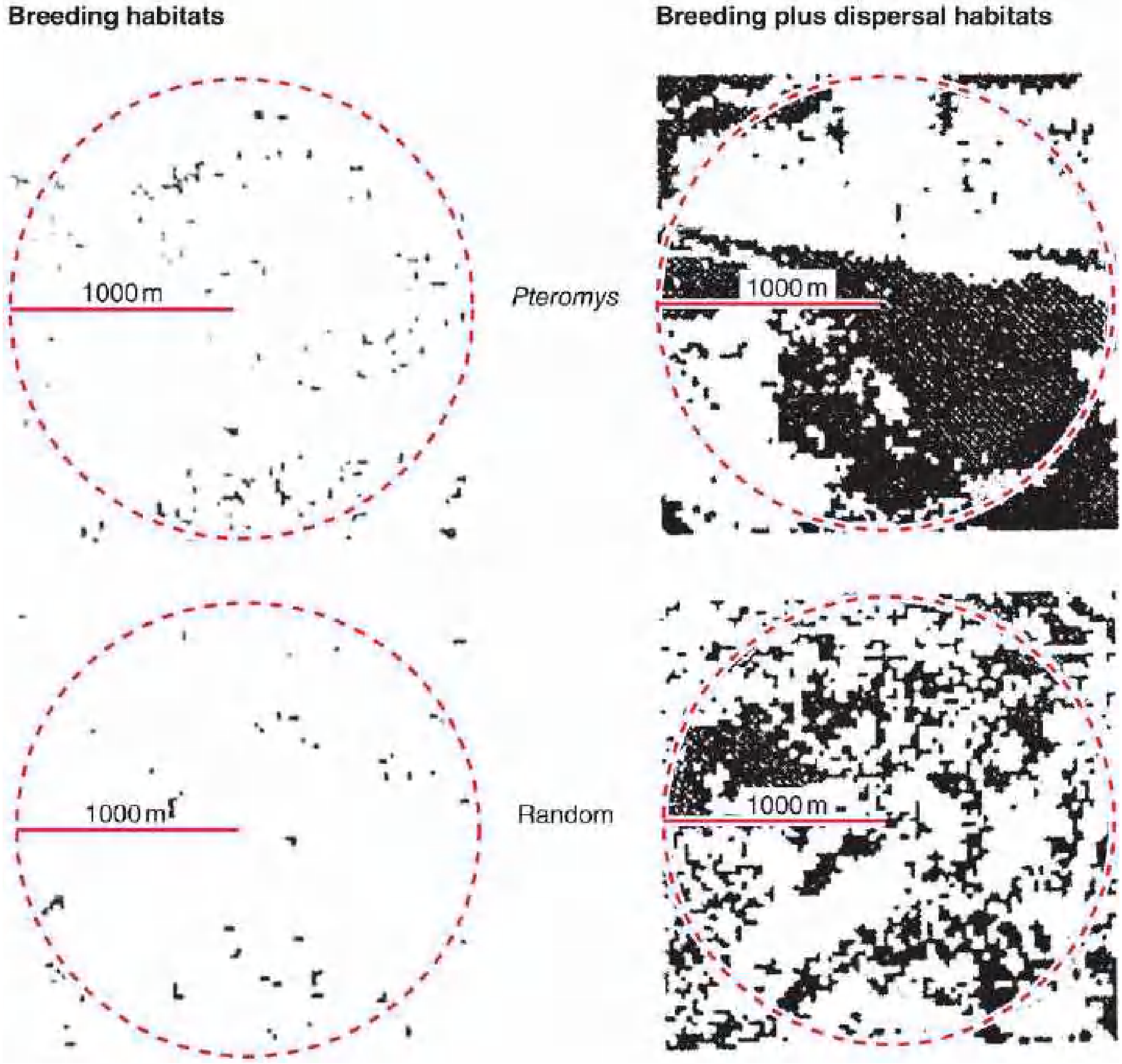


Figure 6.20 Flying squirrels in Finland preferentially occupy habitat favouring effective dispersal. The spatial arrangement of patches (dark) of breeding habitat (left hand panels) and breeding plus dispersal habitat (right hand panels) in a typical landscape containing flying squirrels (*Pteromys*) (top panels) and a random forest location (bottom panels). This flying squirrel landscape contains 4% breeding habitat and 52.4% breeding plus dispersal habitat, compared with 1.5% and 41.5% for the random landscape. Dispersal habitat in the squirrel landscape is much more highly connected (fewer fragments per unit area) than in the random landscape.

Source: After Reunanen *et al*. (2000).

Probably the most fundamental consequence of dispersal for the dynamics of single populations, though, is the effect of density-dependent emigration in reducing resource over-exploitation (see Section 6.3.3). Locally, all that was said in Chapter 5 regarding density-dependent mortality applies equally to density-dependent emigration. Globally, of course, the consequences of the two may be quite different. Those that die are lost forever and from everywhere. With emigration, one population's loss may be another's gain.

6.6.2 Invasion dynamics

the importance of eccentric dispersers

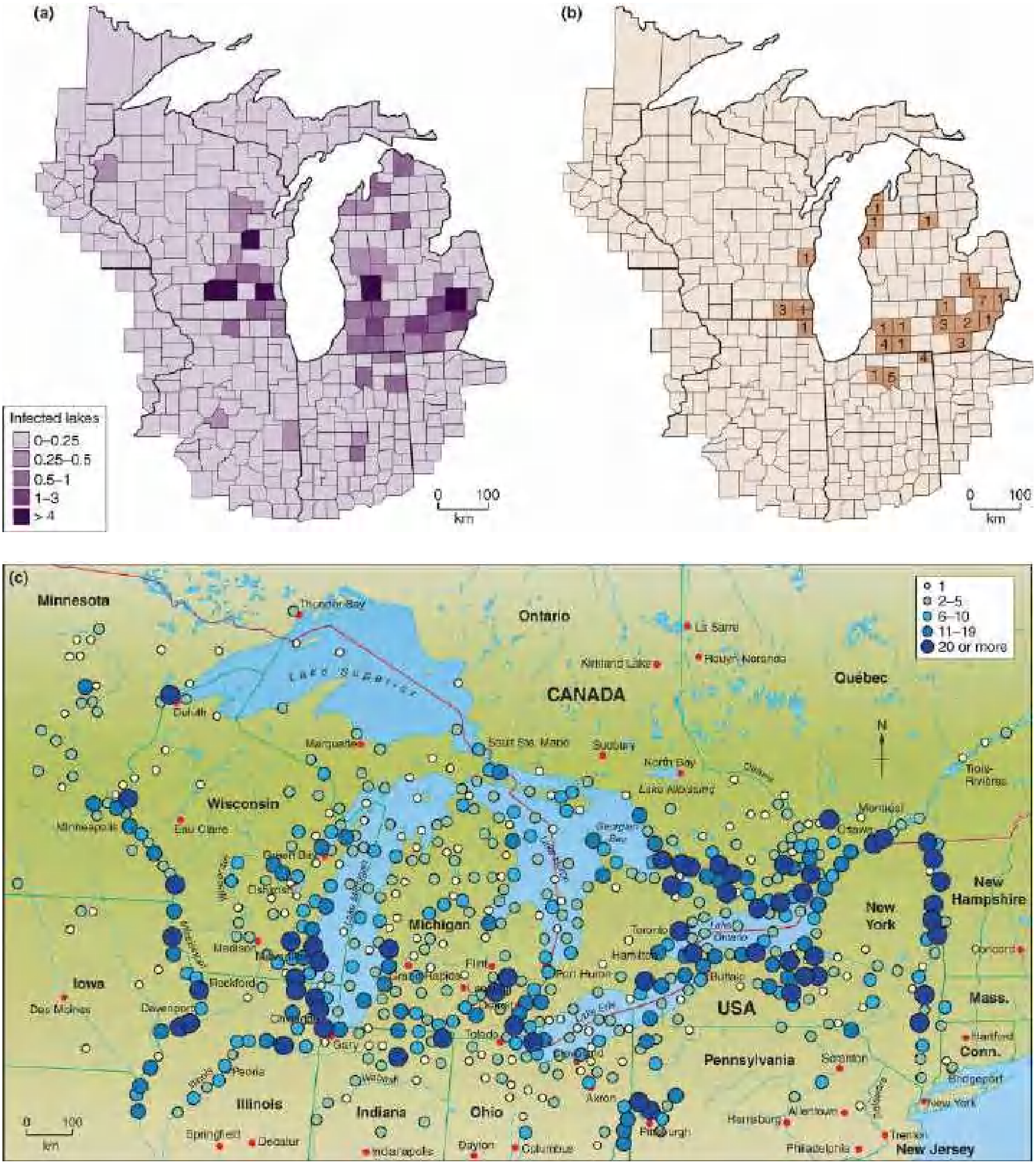
In almost every aspect of life, there is a danger in imagining that what is usual and 'normal' is in fact universal, and that what is unusual or eccentric can safely be dismissed or ignored. Every statistical distribution has a tail, however, and those that occupy the tail are as real as the conformists that outnumber them. So it is with dispersal. For many purposes, it is reasonable to characterise dispersal rates and distances in terms of what is typical. But especially when the focus is on the spread of a species into a habitat it has not previously occupied, those propagules dispersing furthest may be of the greatest importance. Neubert and Caswell (2000), for example, analysed the rate of spread of two species of plants, *Calathea ovandensis* and *Dipsacus sylvestris*. In both cases they found that the rate of spread was strongly dependent on the maximum dispersal distance, whereas variations in the pattern of dispersal at lesser distances had little effect.

APPLICATION 6.6 Invaders of the Great Lakes

In many cases, the tail of the dispersal curve is represented by rare individuals that have dispersed by some means that is itself unusual. For example, the Great Lakes of North America have been invaded by more than 170 alien species (Ricciardi *et al*., 2017), and many of these have arrived as stowaways in ships' ballast water. Ships are filled with ballast water in one location to stabilise them, and suck in organisms with the water, but then transport them to another location where the ballast water and organisms are discharged as heavy cargo is taken aboard. A ballasted ocean freighter, before taking on cargo in the Great Lakes, may discharge three million litres of ballast water that contain various life stages of many plant and animal taxa (and even the cholera bacterium *Vibrio cholerae*) that originate where the ballast water was taken aboard. For example, a whole series of recent invaders (including fish, mussels, amphipods, cladocerans and snails) originated from the other end of an important trade route in the Black and Caspian Seas (Ricciardi & MacIsaac, 2000). One solution is to make the dumping of ballast water in the open ocean compulsory rather than voluntary (this is now the case for the Great Lakes). Other possible methods involve filter systems when loading ballast water, and on-board treatment by ultraviolet irradiation or waste heat from the ship's engines.

The most damaging invaders are not simply those that arrive in a new part of the world; the subsequent pattern and speed of their spread is also significant to managers. Zebra mussels (*Dreissena polymorpha*) have had a devastating effect since arriving in North America via the Caspian Sea/Great Lakes trade route. They threaten native mussels and other fauna, not only by reducing food and oxygen availability but by physically smothering them. The mussels also invade and clog water intake pipes, and millions of dollars need to be spent clearing them from water filtration and hydroelectric generating plants. From the first observations in 1986, range expansion quickly occurred throughout commercially navigable waters, but overland dispersal into inland lakes, mainly attached to recreational boats, has been much slower (Kraft & Johnson, 2000). Geographers have developed so-called 'gravity' models to predict human dispersal patterns based on distance to and attractiveness of destination points, and Bossenbroek *et al*. (2001) adopted the technique to predict the spread of zebra mussels through the inland lakes of Illinois, Indiana, Michigan and Wisconsin (364 counties in all). The model has three steps involving (i) the probability of a boat travelling to a zebra mussel source; (ii) the probability of the same boat making a subsequent outing to an uncolonised lake; and (iii) the probability of zebra mussels becoming established in the uncolonised lake.

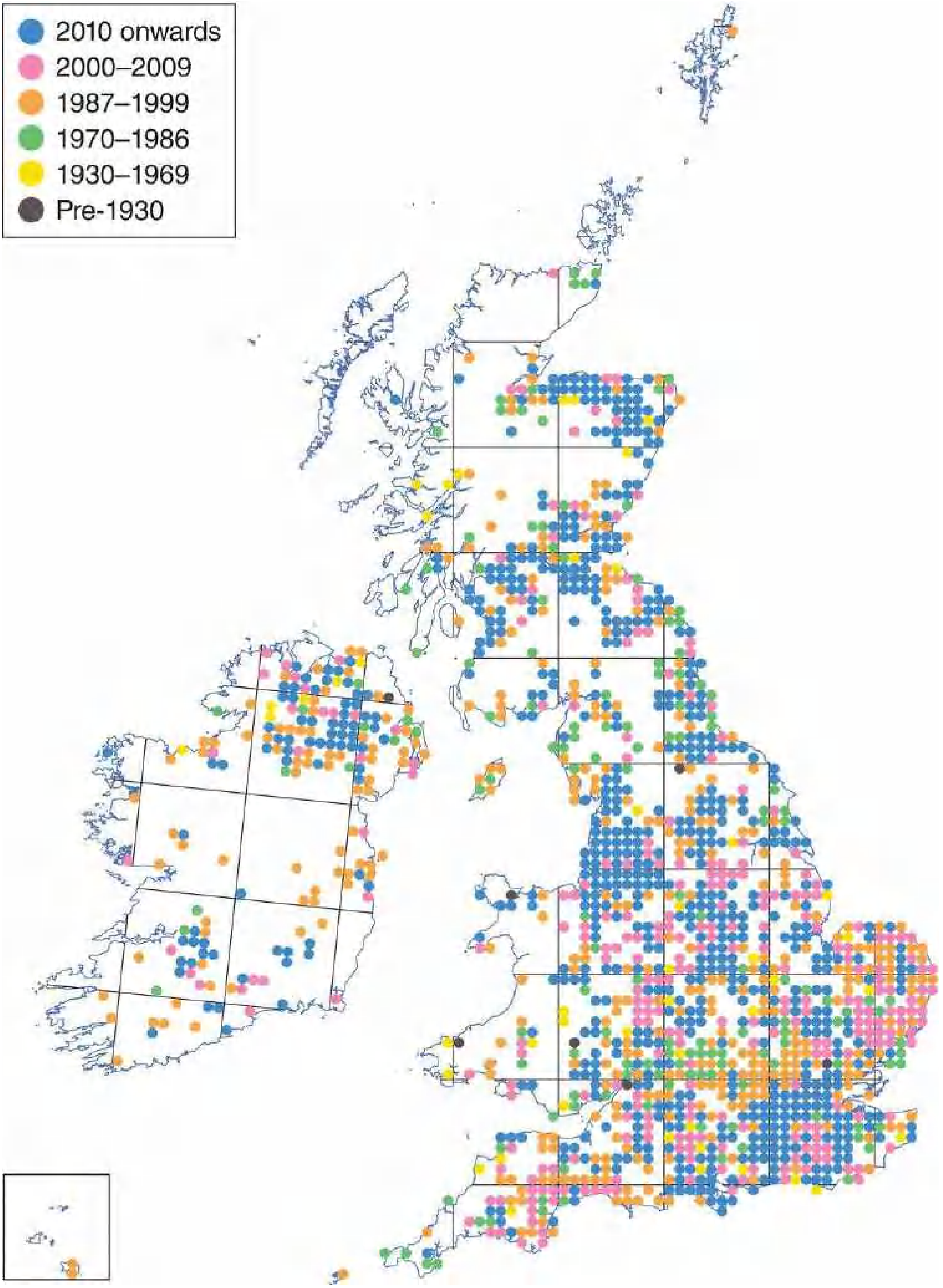
To generate a predicted probabilistic distribution of zebra mussel-colonised lakes, 2000 trials of the model were run for seven years and the number of colonised lakes for each county was estimated by summing the individual colonisation probabilities for each lake in the county. The results, shown in Figure 6.21a, are highly correlated with the pattern of colonisation that actually occurred up to 1997 (Figure 6.21b), giving confidence in the predictions of the model. However, areas of central Wisconsin and western Michigan were predicted to be colonised, but no colonies had at that point been documented. Bossenbroek *et al*. (2001) suggested that invasion may be imminent in these locations, which should therefore be the focus of biosecurity efforts and education campaigns. Nonetheless, by 2017 the mussel had spread to these areas and way beyond (Figure 6.21c).



[Figure 6.21](#) The predictable dispersal of zebra mussels, invading the USA. (a) The predicted distribution (based on 2000 iterations of a stochastic 'gravity' model of dispersal) of inland lakes colonised by zebra mussels in 364 counties in the USA; the large lake in the middle is Lake Michigan, one of the Great Lakes of North America. (b) The actual distribution of colonised lakes as of 1997. (c) The actual distribution as of 1 May 2017.

Source: (a, b) After Bossenbroek *et al*. ([2001](#)). (c) After United States Geological Survey ([2017](#)).

Of course invaders do not all rely on human agency; many disperse by their own devices. We can see the example of the giant hogweed, *Heracleum mantegazzianum*, a noxious weed in the UK and Ireland in [Figure 6.22](#), and we saw something comparable for the western corn rootworm in [Figure 6.18](#).



2010 onwards
2000–2009
1987–1999
1970–1986
1930–1969
Pre-1930

Figure 6.22 Invasion of the giant hogweed. The changing distribution over time, as of 25 May 2017, of the noxious weed, the giant hogweed, *Heracleum mantegazzianum*, in the UK and Ireland as it has invaded an increasing number of areas. The map is compiled and updated by the Biological Records Centre of the UK, where comparable maps for most other UK plant species can be found at http://www.brc.ac.uk/plantatlas/plant/species-genus.

Source: Biological Records Centre (2017).

This dependence of invasion on rare long-distance dispersers means, in turn, that the probability of a species invading a new habitat may have far more to do with the proximity of a source population (and hence the opportunity to invade) than it does on the performance of the species once an initial bridgehead has been established. For example, the violet copper butterfly, *Lycaena helle*, one of the rarest butterfly species in Central Europe, was studied in the Westerwald region in Germany, where it is confined to patches of habitat containing adderwort, *Bistorta officinalis*, its only larval food plant in the area. This was one of several features used to classify habitat patches as 'suitable' for the violet copper, giving 230 patches in total. These were then surveyed to determine whether they were occupied by the butterfly or not, and what features distinguished occupied from vacant patches. Three factors had a significant influence on occupancy. Patches were more likely to be occupied the higher the percentage of the ground covered by adderwort – an indicator of 'patch quality' from the butterflies' point of view. But occupancy was also more likely in larger patches (larger targets for dispersal) and in patches that were less isolated from other occupied sites (Figure 6.23). Abundance was being limited, to a significant degree, not so much by a lack of habitat but by difficulties in reaching it.

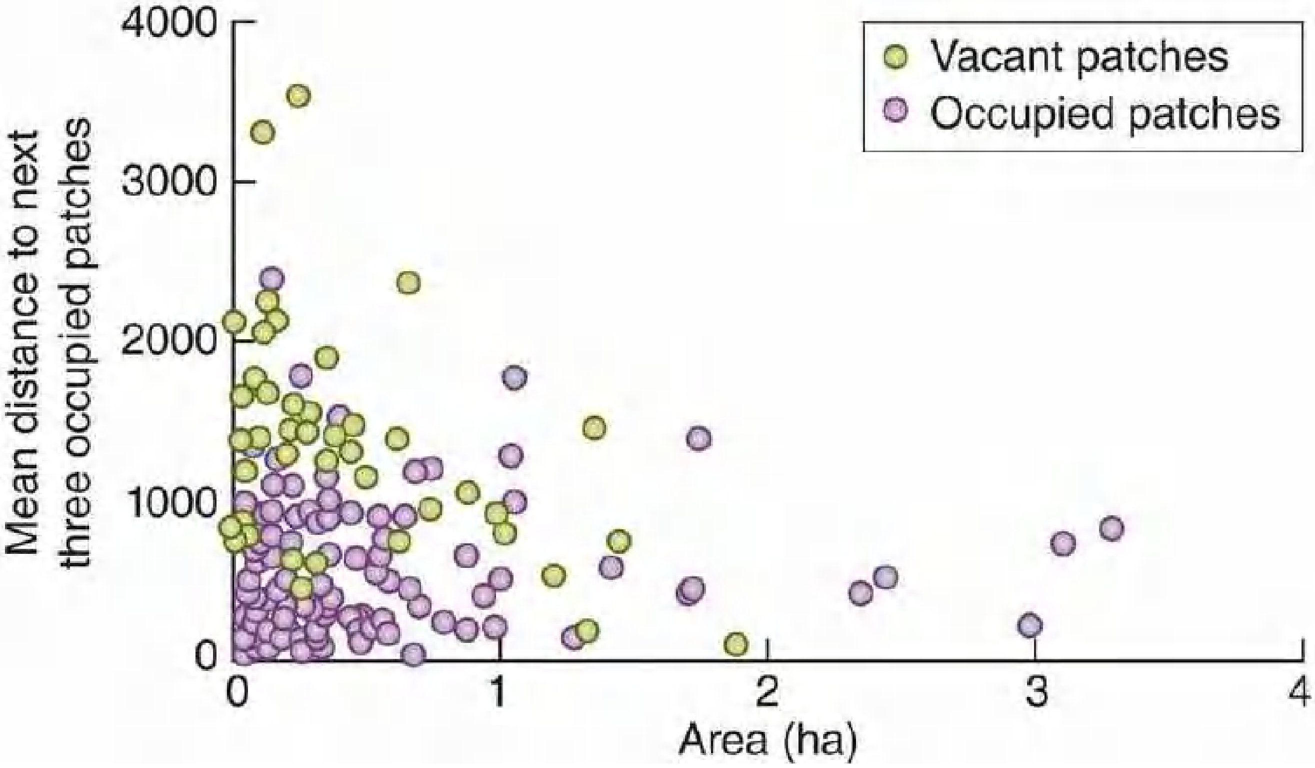


Figure 6.23 Butterflies tend to occupy the largest, least isolated habitat patches. The sizes of vacant patches and those occupied by the violet copper butterfly, *Lycaena helle*, in the Westerwald, Germany, and also their level of isolation from other occupied patches.

Source: After Bauerfeind *et al.* (2002).

6.6.3 Modelling dispersal: the distribution of patches

metapopulations and subpopulations

The nature of the role of dispersal in population dynamics depends on how we think of those populations. The simplest view sees a population as a collection of individuals distributed more or less continuously over a stretch of more or less suitable habitat, such that the population is a single, undivided entity. Dispersal is then a process contributing to either increase (immigration) or decrease (emigration) in the population as a whole, or one that redistributes individuals within the population (see Application 6.7). In many cases, however, we need to remind ourselves of what we noted in Section 6.4.1: the ubiquity of patchiness in ecology and the importance of dispersal in linking patches to one another. From this perspective, many populations are best viewed as *metapopulations*.

A metapopulation is defined as a collection of subpopulations, where a subpopulation occupies a habitable patch in the landscape and corresponds, in isolation, to the simple view of a population described above. However, the dynamics of the metapopulation as a whole is determined, not by dispersal *within* subpopulations, but by the rate of extinction of existing subpopulations and the rate of establishment (colonisation) of new subpopulations by dispersal into habitable but uninhabited patches. Note, however, that just because a species occupies more than one habitable site, each of which supports a population, this does not mean that those populations comprise a metapopulation. As we shall discuss more fully below, ‘classic’ metapopulation status is conferred only when extinction and colonisation play a major role in the overall dynamics.

APPLICATION 6.7 Reaction–diffusion modelling of shifting species' distributions under climate change

Variations in the way that we think of dispersal in the dynamics of populations are reflected in alternative approaches to modelling dispersal. If we treat space not as patchy but as continuous and homogeneous, then we can model dispersal as part of a 'reaction–diffusion' system. Here, the 'reaction' refers to the dynamics at any given location, driven by birth and survival, with dispersal added as a separate 'diffusion' term. The approach has arguably been more useful in other areas of biology (e.g. developmental biology) than it has in ecology. Nonetheless, the mathematical understanding of such systems is strong, and they are particularly good at demonstrating how spatial variation (i.e. patchiness) can be generated, internally, within an intrinsically homogeneous system (Keeling, 1999).

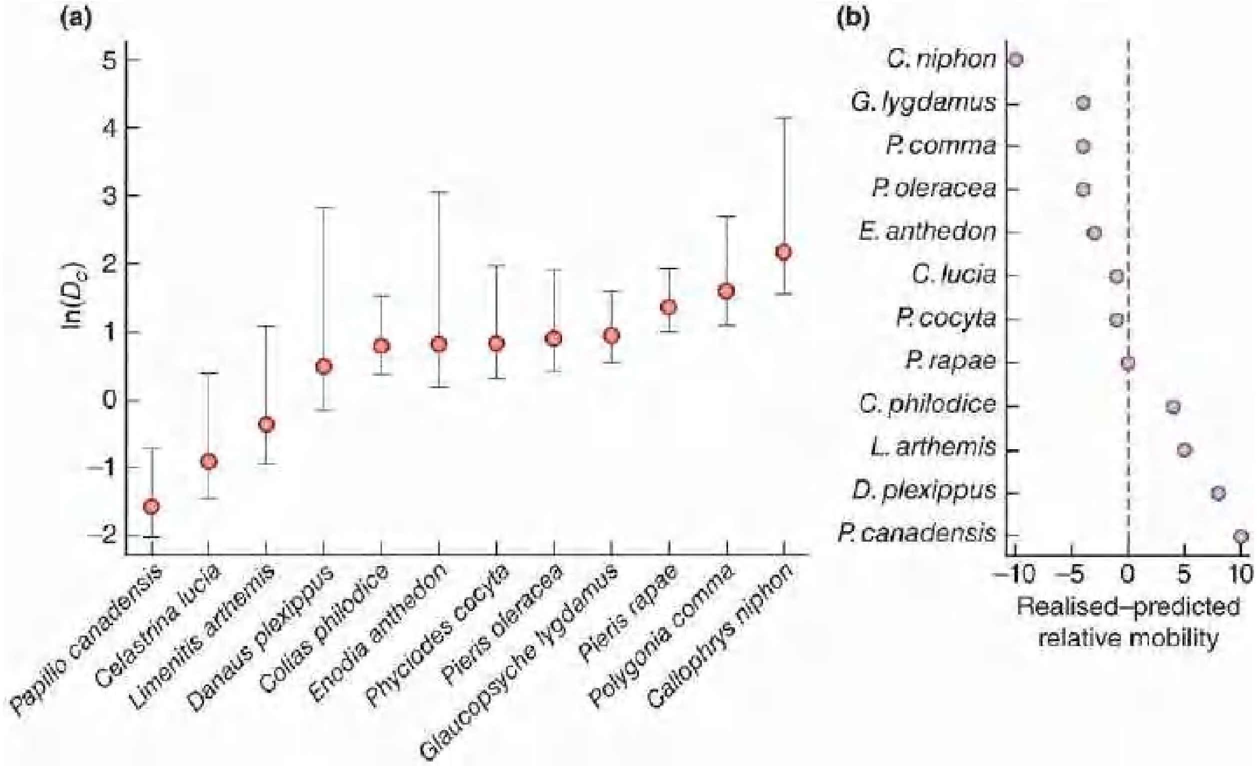
Reaction–diffusion models known as 'moving habitat models' (Harsch *et al*., 2017) can also be valuable in trying to predict how species distributions will respond to climate change. Without going into the details of the models, it is easy to see that as the climate changes, and particularly as the climate warms, the geographic limits of suitable habitat for many species in temperate regions – their 'climate envelope' – will shift polewards. Those species will therefore only survive in the face of global warming if they themselves can shift polewards at a comparable rate, which will require them to be both sufficiently productive at any given location (reflected in the reaction term) and to have sufficient powers of dispersal (the diffusion term). Specifically, a reaction–diffusion model developed by Leroux *et* al. (2013) proposed that a species could only persist if

$$2\sqrt{(Dr)} > q, \tag{6.1}$$

where D is the diffusion rate, r is the per capita rate of increase (Section 4.7.1) and q is the rate of movement of the climate envelope. This in turn allows us to determine a critical diffusion rate, D_c, that must be exceeded if the species is to persist, given by

$$D_c = q^2/4r. \tag{6.2}$$

Leroux *et al*. (2013) then used data on q and r to calculate values of D_c for 12 butterfly species in North America (Figure 6.24a), and compared these with mobility scores for the species derived by seeking the expert opinions of 51 North American lepidopterists (Figure 6.24b). Large negative values in the comparison (mobility much less than the estimated critical value) characterise the species most at risk of failing to keep pace with climate change; large positive values indicate the species least at risk. This ranking of the species is of course not perfect. The methods do not, for example, take account of rare long-distance dispersers. Nonetheless, since conservation ecologists must continuously be setting priorities, and directing their attention to where needs are most urgent or efforts most likely to be rewarded, results from models like these can be a valuable guide.



[**Figure 6.24**]() **Identifying priority North American butterfly species for conservation by assessing their dispersal ability in the face of climate change.** (a) Log-transformed estimates of the critical diffusion rates, D_c, of 12 butterfly species in North America that must be exceeded if those species are to persist in the face of predicted climate change and habitat shifting. The bars are the ranges between the estimates of D_c calculated from the 95% CIs of r, the per capita rate of increase, used to calculate D_c (Equation 6.2). (b) The difference, for these 12 species, between the relative rank of D_c and of estimates of their mobility from consultations with experts. Larger negative differences indicate that species are most at risk of failing to keep pace with climate change, etc.

Source: After Leroux *et al.* (2013).

If we think of dispersal as occurring within a patchy landscape, there are two alternative ways of modelling this (Keeling, 1999). The first is a 'spatially implicit' approach (Hanski, 1999), where the key feature is that a proportion of the individuals leave their home patches and enter a pool of dispersers and are then redistributed amongst patches, usually at random. Thus, these models do not place patches at any specific spatial location. All patches may lose or gain individuals through dispersal, but all are, in a sense, equally distant from all other patches. Many metapopulation models (see later), come into this category, and despite their simplicity (real patches do have a location in space) they have provided important insights, in part because their simplicity makes them easier to analyse.

In contrast, 'spatially explicit models' acknowledge that the distances between patches vary, as do therefore the chances of them exchanging individuals through dispersal. The earliest such models, developed in population genetics, were linear 'stepping stone' models, where dispersal occurred only between adjacent patches in the line. Subsequently, spatially explicit approaches have often involved 'lattice' models in which patches are arranged on a (usually) square grid, and patches exchange dispersing individuals with 'neighbouring' patches – perhaps the four with which they share a side, or the eight with which they make any contact at all, including the diagonals (Keeling, 1999). Further models can bridge the gap between lattices and spatial arrangements in the real world (Figure 6.25a), but they are all still caricatures of nature. They are nonetheless useful in highlighting new dynamic patterns that appear as soon as space is incorporated explicitly, and depending on how it is incorporated. Of course, spatial models generate spatial patterns (see, for example, Section 10.5), but as we can see in Figure 6.25b, for example, they also typically alter temporal dynamics. In that study, the dynamics of four idealised types of metapopulation were compared that differed only in their spatial structure. That is, they were all the same in terms of their overall size (number of patches), the number of connections between patches, and the extinction–colonisation dynamics of individual patches. Nonetheless, there were significant differences in the abundance (proportion of patches occupied) and persistence of the metapopulations, especially where extinction tended to dominate colonisation, with the more complex spatial structures promoting greater abundance and persistence.

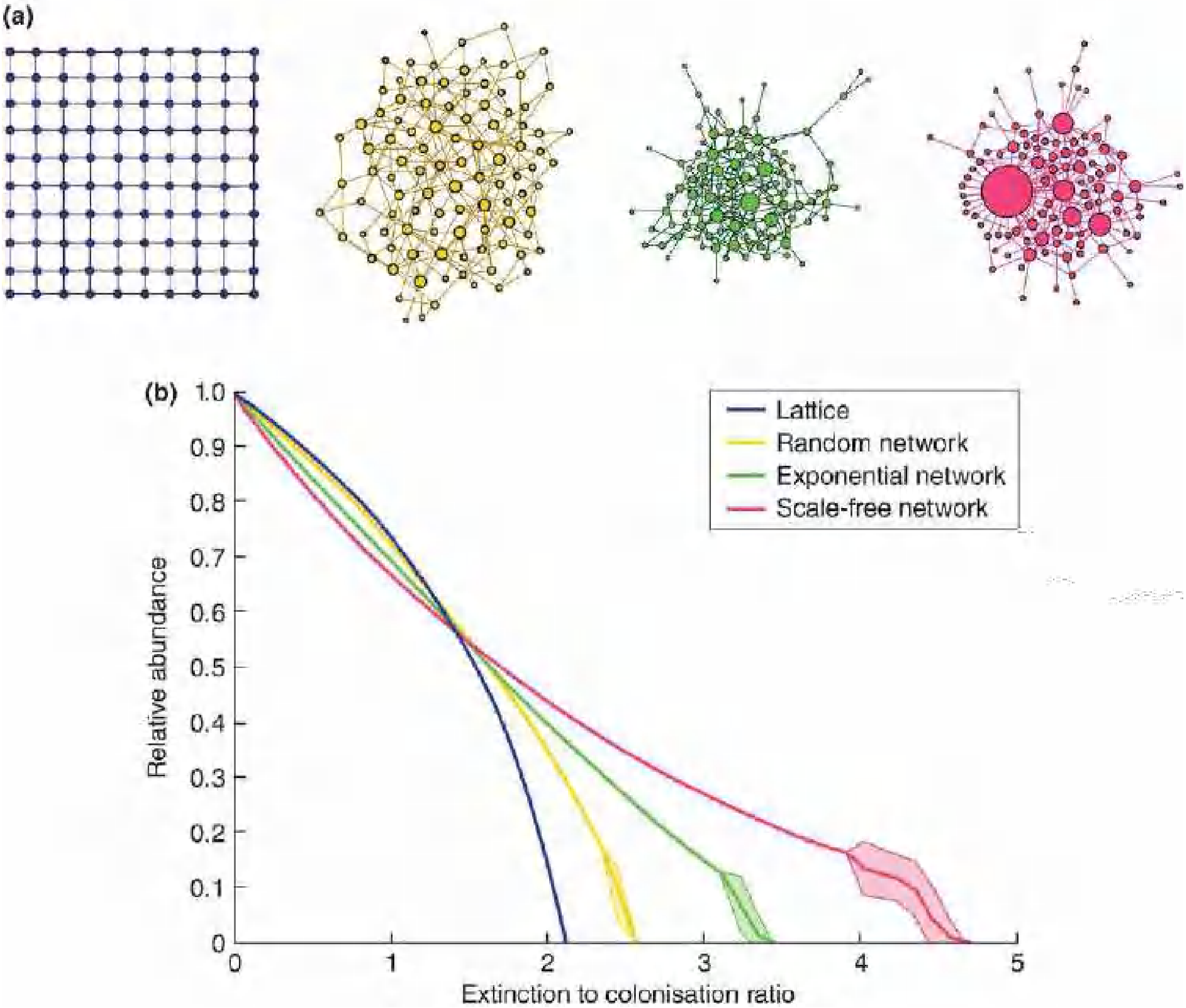


Figure 6.25 The spatial structure of a metapopulation affects its overall abundance and persistence. (a) Four spatial networks (metapopulations) of increasing complexity, compared in a simulation study of the effect of spatial structure on overall population persistence. All networks here have 100 subpopulations (nodes) for purposes of illustration, with the size in the figure of each node proportional to its 'connectivity' (the number of other nodes to which it is connected). In the simulations, all metapopulations had 1024 nodes and 2048 links between nodes. Hence, they differed only in structure. From left to right: a square lattice with each node connected to its neighbours; a network where nodes are connected together at random, so that long distance movement between nodes is possible; a network where connections follow an exponential distribution, again with long distance movement but with greater variation in connectivity; and a scale-free network where the variation in connectivity is greater still. Each node can either be occupied or unoccupied, and so the abundance of a population is given by the proportion of nodes occupied. (b) The abundance of each type (proportion of nodes occupied) at differing extinction–colonisation ratios. Each line is the average of 10 replicate simulations; the shaded areas represent standard deviations.

Source: After Gillaranz & Bascompte (2012).

6.7 The dynamics of metapopulations

6.7.1 Uninhabited habitable patches

Recognition that many populations are in fact metapopulations was firmly established around 1970, but there was a delay of about 20 years before that recognition was translated into action and an increasing number of studies placed metapopulation dynamics prominently on the ecological stage. Now, the danger is not so much one of neglect, but that all populations are thought of as metapopulations, simply because the world is patchy.

Central to the concept of a metapopulation is the idea, emphasised by Andrewartha and Birch back in 1954, that habitable patches might be uninhabited simply because individuals have failed to disperse into them. To establish that this is so, we need to be able to identify habitable sites that are not inhabited. Early attempts included a survey of water voles (*Arvicola terrestris*) in 39 sections of river bank in North Yorkshire, UK, 10 of which contained breeding colonies of voles (core sites), 15 were only visited, and 14 were apparently neither used nor visited. A ‘principle components’ analysis was used to characterise the core sites, and on this basis a further 12 sites were identified that should have been suitable for breeding voles but were uninhabited by voles because they were too isolated to be colonised or in some cases suffered high levels of predation by mink (Lawton & Woodroffe, 1991). In another case, Thomas *et al.* (1992) found that the habitable patches that remained uninhabited of the rare silver-studded blue butterfly, *Plebejus argus*, in North Wales (those containing the plants its larvae feed on) were small and isolated: the butterfly was able to colonise virtually all habitable sites less than 1 km from existing populations. Crucially, the habitability of some of the isolated (previously uninhabited) sites was confirmed when the butterfly was successfully introduced into them (Thomas & Harrison, 1992).

APPLICATION 6.8 Species distribution modelling for (re)introductions and invasions

Recently, there have been many attempts to determine where a species would be able to live just as long as it could disperse there. These come under the umbrella term ‘species distribution modelling’ (Elith & Leathwick, 2009), the most influential variant of which has been ‘ecological niche modelling’, which we discussed in Chapter 2. To recapitulate briefly, as much environmental information as possible is taken from all of the locations where a species is currently found and also from a range of locations where the species is *not* currently found, allowing those locations to be identified that meet the species’ requirements even though the species is currently absent. In the present context, the implication is that it may only be the absence of effective dispersal that is currently preventing the species from living in these locations. The approach can be used to plan introductions or reintroductions of species (dispersal augmented by us) so as to maximise chances of success. We see an example of this for Californian condors, *Gymnogyps californianus*, in California, Oregon and Washington States in the USA in Figure 6.26a. Thirteen predictor variables from locations where they had been observed between 1960 and 2011 were used to ‘train’ the model, which was then applied throughout the region. The results suggest that many parts of the condor’s historical range retain the characteristics that would support it, and they point to sites where reintroductions are most like to be successful.

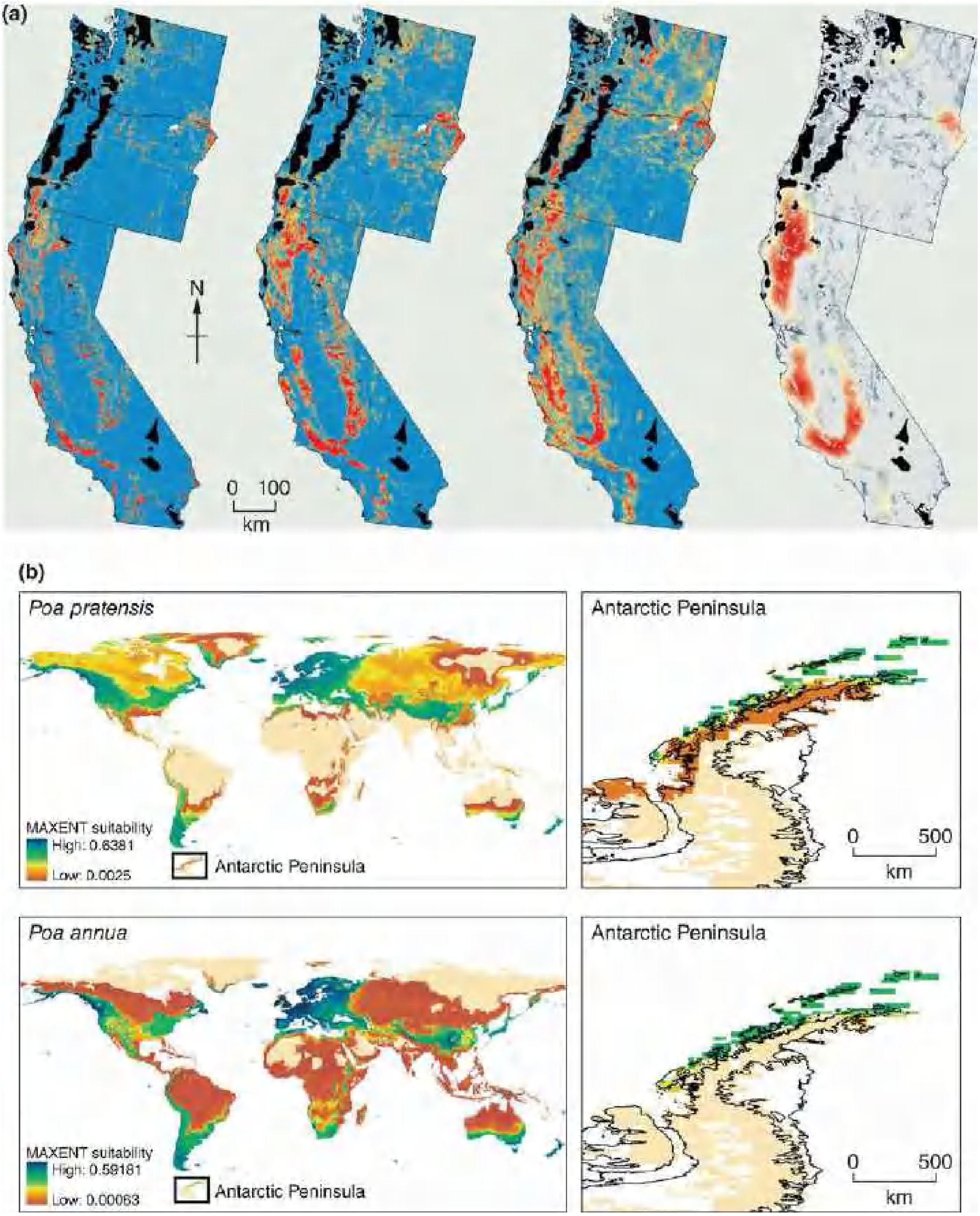

(a)
N
0 100
km
(b)
Poa pratensis
MAXENT suitability
High: 0.6381
Low: 0.0025
Antarctic Peninsula
Antarctic Peninsula
0 500
km
Poa annua
MAXENT suitability
High: 0.59181
Low: 0.00063
Antarctic Peninsula
Antarctic Peninsula
0 500
km

Figure 6.26 Applications of species distribution modelling. (a) Ecological niche modelling, using the algorithm MAXENT, for the critically endangered Californian condor, *Gymnogyps californianus*, in the states of California, Oregon and Washington, USA. From left to right, the first three maps show results for nesting, roosting and feeding, with warmer (redder) colours indicating greater suitability. Black areas represent (presumably unsuitable) areas with environmental conditions not encountered during model training. To the far right is a reintroduction suitability map combining the results to the left. (b) Ecological niche modelling, again using MAXENT, for the grasses *Poa pratensis* and *P. annua* globally and in the Antarctic Peninsula.

Source: (a) After D'Elia *et al.* (2015). (b) After Pertierra *et al.* (2017).

It has also been possible to use species distribution modelling to identify locations most at risk from a pest species if it were able to disperse there. We see an example of this for two species of grass in Antarctica (Figure 6.26b). *Poa pratensis* and *P. annua* are the only non-indigenous vascular plants established long term in the Antarctic Peninsula, but have been limited to only one established site for *P. pratensis* and six for *P. annua*. Results from ecological niche modelling, however, using four selected temperature variables – precipitation is of little importance in Antarctic conditions – show that there are a large number of other sites in Antarctica at threat from further invasion. They point in particular to some sites, in the South Shetland Islands, where the need for biosecurity measures is particularly great.

6.7.2 The development of metapopulation theory: islands and metapopulations

MacArthur and Wilson's (1967) classic book, *The Theory of Island Biogeography* was an important catalyst in radically changing ecological thinking in a whole range of areas. The authors developed their ideas in the context of the dynamics of the animals and plants on real (maritime) islands, which they interpreted as reflecting a balance between the opposing forces of extinctions and colonisations. They emphasised that some species (or local populations) spend most of their time either recovering from past crashes or in phases of invasion of new territories (islands), while others spend much of their time at or around their carrying capacity. These two ends of a continuum are the *r* and *K* species we discuss more fully in Section 7.6.1. At one extreme (*r* species), individuals are good colonisers and have characteristics favouring rapid population growth in an empty habitat. At the other end of the continuum (*K* species) individuals are not such good colonisers but have characteristics favouring long-term persistence in a crowded environment. *K* species therefore have relatively low rates of both colonisation and extinction, whereas *r* species have relatively high rates. These ideas are developed further in the discussion of island biogeography in Chapter 19.

At about the same time as MacArthur and Wilson's book was published, a simple model of metapopulation dynamics was proposed by Levins (1969 , 1970). Like MacArthur and Wilson, he sought to incorporate into ecological thinking the essential patchiness of the world around us. MacArthur and Wilson were more concerned with whole communities of species, and envisaged a 'mainland' that could provide a regular source of colonists for the islands. Levins focused on populations of a single species and awarded none of his patches special mainland status. Levins introduced the variable *p(t)*, the fraction of habitat patches occupied at time *t*, reflecting an acceptance that not all habitable patches are always inhabited.

Levins' model

The rate of change in the fraction of occupied habitat (patches, *p*) is given in Levins' model as:

$$\frac{dp}{dt} = cp(1-p) - ep \tag{6.3}$$

in which e is the rate of local extinction of patches and c is the rate of recolonisation of empty patches. That is, the rate of recolonisations increases both with the fraction of empty patches prone to recolonisation $(1 - p)$ and with the fraction of occupied patches able to provide colonisers, p, whereas the rate of extinctions increases simply with the fraction of patches prone to extinction, p. Rewriting this equation, Hanski (1999) showed that it is structurally identical to the logistic equation (see Section 5.7):

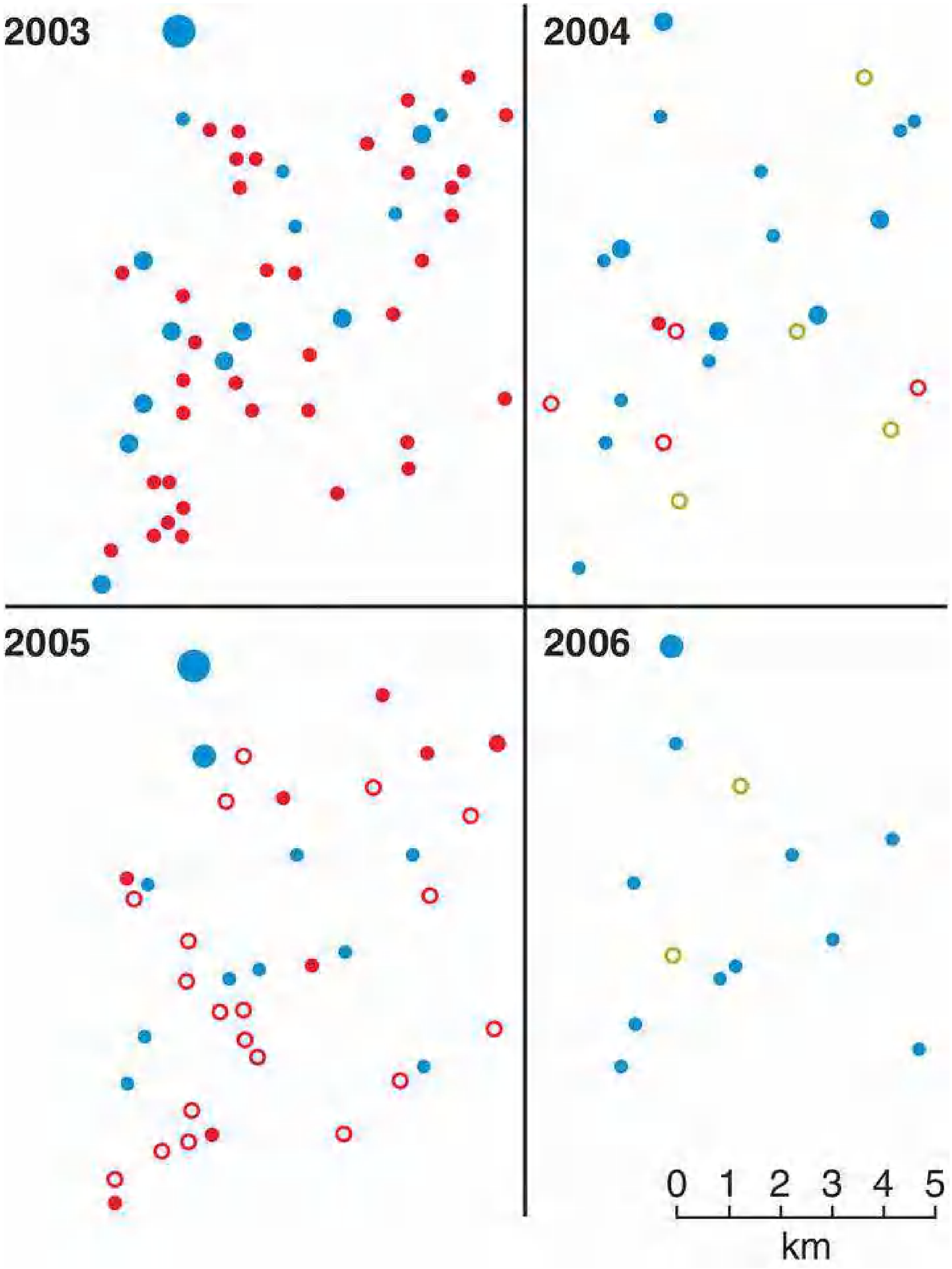
$$\frac{dp}{dt} = (c-e)p\left(1 - \frac{p}{1-\left(\frac{e}{c}\right)}\right). \tag{6.4}$$

Hence, as long as the intrinsic rate of recolonisation exceeds the intrinsic rate of extinction ($c > e$, or $e/c < 1$), the total metapopulation will reach a stable equilibrium, with a fraction, $p^* = 1 - (e/c)$, of the patches occupied (what we would call K in the logistic equation), or if we refer to e/c as δ, the 'extinction threshold', then

$$p^* = 1 - \delta. \tag{6.5}$$

extinctions and colonisations in subpopulations: a stable metapopulation

The most fundamental message from taking a metapopulation perspective, then, which emerges from even the simplest models, is that a metapopulation can persist, stably, as a result of the balance between extinctions and recolonisations even though none of the local populations are stable in their own right. An example of this is shown in Figure 6.27, where a metapopulation of solitary bees, *Andrena hattorfiana*, in southern Sweden was surveyed several times each year from 2003 to 2006. Habitat patches (potential subpopulations) were those containing the bee's main pollen source, *Knautia arvensis*, and it is clear that from one year to the next, the chance of an individual subpopulation going extinct (or a vacant patch being colonised) was extremely high. The metapopulation was stable but the subpopulations were not. Only 16% of the subpopulations remained occupied throughout the study. Those with the smallest bee populations were most likely to go extinct; those covering the largest area were most likely to be colonised. To restate the message another way: if we wish to understand the long-term persistence of a population, or indeed that population's dynamics, then we may need to look beyond the local rates of birth and death (and what determines them), or even the local rates of immigration and emigration. If the population as a whole functions as a metapopulation, then the rates of subpopulation extinction and colonisation may be of at least comparable importance.



2003
2004
2005
2006
0
1
2
3
4
5
km

Figure 6.27 Many subpopulations of a bee metapopulation go extinct from year to year but the metapopulation nonetheless persists. The spatial distribution of subpopulations in a metapopulation of the bee, *Andrena hattorfiana*, in southern Sweden over four years. Filled blue dots represent occupied patches of increasing subpopulation size, in four classes: 1–10, 11–50, 51–100 and >100. Red dots indicate subpopulations that had gone extinct by the subsequent year. Green open dots indicate newly colonised patches. Red open dots indicate either patches that were colonised one year but had gone extinct by the next, or subpopulations that went extinct one year but had been recolonised by the next.

Source: After Franzen & Nilsson (2010).

6.7.3 When is a population a metapopulation?

Two necessary features of a metapopulation have now been established: that individual subpopulations have a realistic chance of experiencing extinction and of recolonisation. To this we should add a third, which has been implicit up to now. The dynamics of the various subpopulations should be largely independent, i.e. not synchronous. There would, after all, be little hope of stability if when one subpopulation went extinct they all did. Rather, asynchrony guarantees that as one goes extinct (or even declines towards extinction), there are likely to be others that are thriving and generating dispersers, promoting a 'rescue effect' (Brown & Kodric-Brown, 1977) of the former by the latter.

sources and sinks: mainland–island metapopulations

Some metapopulations may conform to the 'classic' concept, in which all the subpopulations have a realistic (and roughly equal) chance of extinction. In other cases, however, there may be significant variation in the quality or especially in the size of individual patches. Thus, patches may be divided into 'sources' (donor patches) and 'sinks' (receiver patches) (Pulliam, 1988), and metapopulations with marked variations in patch size are often referred to as *mainland–island metapopulations*, acknowledging the gradations between MacArthur and Wilson's island biogeography (one mainland) and Levins' classic metapopulation. In source patches at equilibrium, the number of births exceeds the number of deaths, whereas in sink patches the reverse is true. Hence, source populations support one or more sink populations within a metapopulation. The persistence of the metapopulation depends not only on the overall balance between extinction and recolonisation, as in the simple model, but also on the balance between sources and sinks. In practice, of course, there is likely to be a continuum of types of metapopulation: from collections of nearly identical local populations, all equally prone to extinction, to metapopulations in which there is great inequality between local populations, some of which are effectively stable in their own right. This contrast is illustrated in Figure 6.28 for the silver-studded blue butterfly (*Plebejus argus*) in North Wales. We saw something similar for the violet copper in Figure 6.23. The smaller patches were more likely to be vacant because they were less likely to be colonised (smaller target) and more likely to go extinct (smaller, more vulnerable subpopulation).

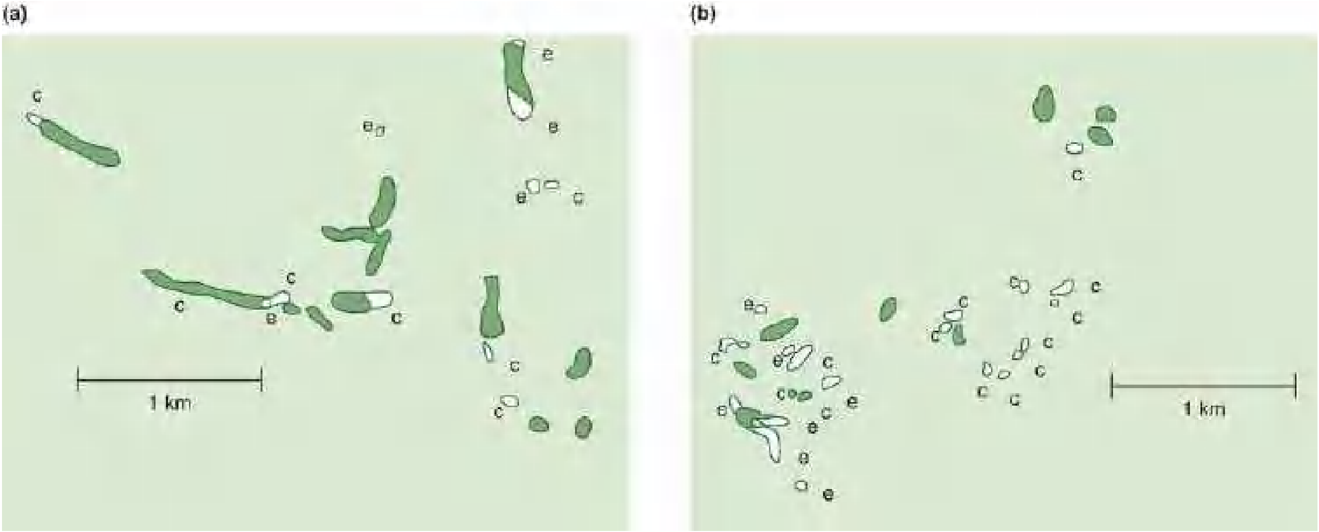


Figure 6.28 Mainland–island metapopulations of a butterfly with contrasting proportions of sources and sinks. Two metapopulations of the silver-studded blue butterfly (*Plebejus argus*) in North Wales: (a) in a limestone habitat in the Dulas Valley, where there was a large number of persistent (often larger) local populations amongst smaller, much more ephemeral local populations; (b) in a heathland habitat at South Stack Cliffs, where the proportion of smaller and ephemeral populations was much greater. Filled outlines, present in both 1983 and 1990; open outlines, not present at both times; e, present only in 1983 (presumed extinction); c, present only in 1990 (presumed colonisation).

Source: After Thomas & Harrison (1992).

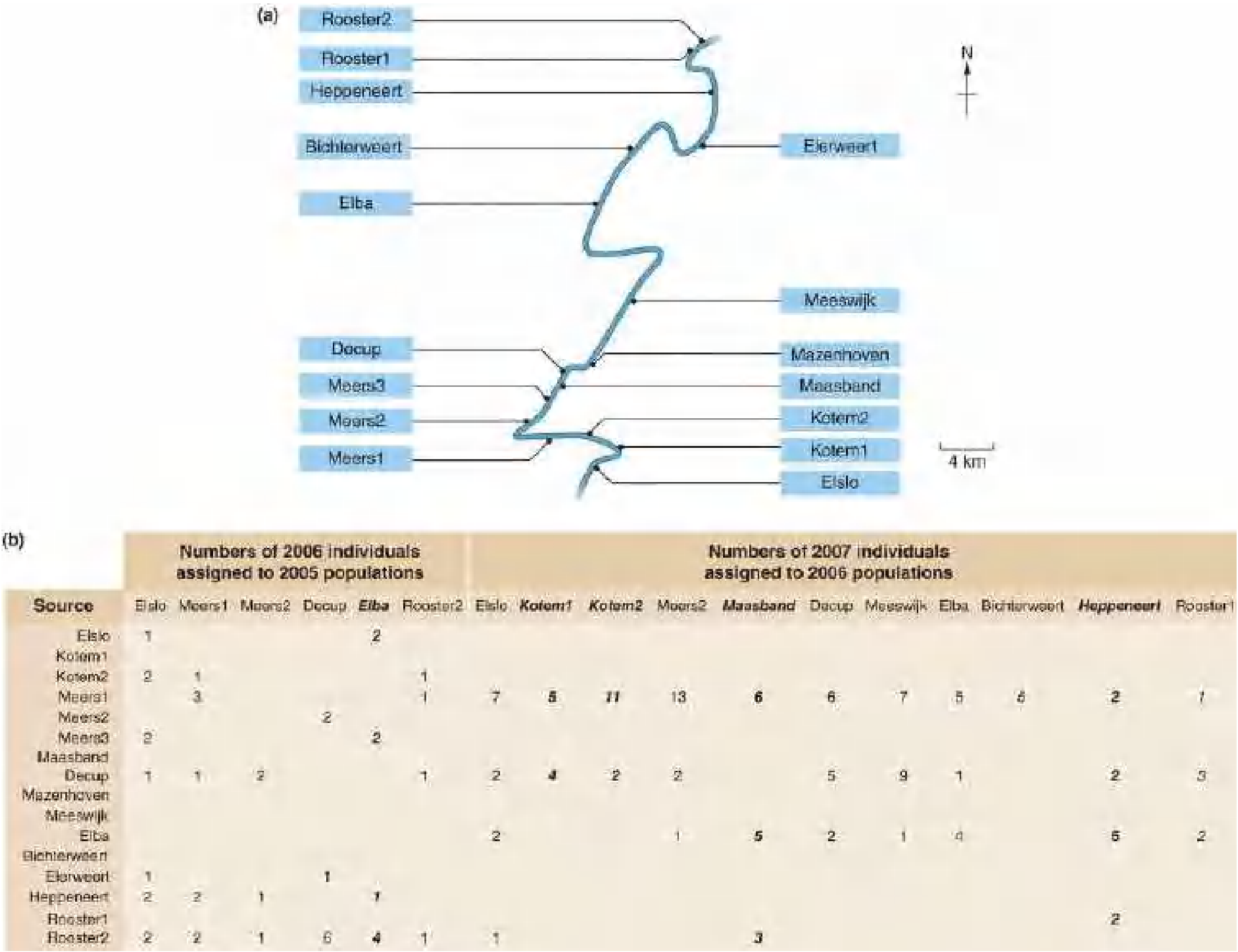
when is a metapopulation not a metapopulation?

Just because a population is patchily distributed, however, this does not necessarily make it a metapopulation (Harrison & Taylor, 1997 ; Bullock *et al*., 2002). Firstly, a population may be patchily distributed, but dispersal between the patches may be so great that the dynamics of the individual patches are no longer independent: this is a single population, albeit occupying a heterogeneous habitat. Alternatively, patches may be so isolated from one another that dispersal between them is negligible: a series of effectively separate populations.

Finally, and perhaps most commonly, all patches may simply have a negligible chance of extinction, at least on observable timescales. Their dynamics may be influenced by birth, death, immigration and emigration – but not to any significant degree by extinction or recolonisation. This last category comes closer to being a 'true' metapopulation, and there can be little doubt that many such patchy populations have been described as metapopulations. Strictly speaking this is wrong, but there can be a danger in being overprotective of the purity of definitions. What harm can there be if, as interest in the metapopulation concept grows, the term itself is extended to a wider variety of ecological scenarios? Perhaps none – and the spread of the term's usage to populations originally beyond its reach may, in any case, be unstoppable. But a word, like any other signal, is only effective if the receiver understands what the sender intends. At the very least, care should be taken by users of the term to confirm whether the extinction and recolonisation of patches has been established.

metapopulations of plants? remember the seed bank

The problem of identifying metapopulations is especially apparent for plants (Husband & Barrett, 1996 ; Bullock *et al*., 2002). There is no doubt that many plants inhabit patchy environments, and apparent extinctions of local populations may be common, but the applicability of the idea of recolonisation following a genuine extinction is nonetheless questionable in any species that has a buried seed bank (see Section 4.4). ‘Recolonisations’ may often simply be the result of the germination, following habitat restoration, of seeds that are already in place. Recolonisation by dispersal, a prerequisite for a true metapopulation, may be extremely rare. There are, though, at least some examples, one of which looked at an annual plant, the treacle mustard, *Erysimum cheiranthoides*, that occupies sites on stony banks of the River Meuse in Belgium (Figure 6.29a). As an annual, there is clearly no continuity of adults at a site from one year to the next, but individual sites are subject to flooding each winter, which may be powerful enough to also flush away all seeds from a site, leading to extinction of a subpopulation. The river, though, may also bring new seeds (i.e. colonists) to a site. Samples were taken from sites, and DNA extracted and analysed, allowing each site to be genetically characterised and differentiated from other sites. This in turn allowed at least a proportion of individuals sampled in 2006 to be assigned to populations from 2005, and likewise for individuals in 2007 assigned to 2006 populations. It was apparent not only that seeds moved between the subpopulations, but also that in some cases they were carried to vacant sites and recolonised them (Figure 6.29b). In this case at least, the plant formed a true metapopulation; its dynamics had at least as much to do with extinction and colonisation as with local birth and death.



	Numbers of 2006 individuals assigned to 2005 populations						Numbers of 2007 individuals assigned to 2006 populations										
Source	Elsio	Meers1	Meers2	Decup	***Elba***	Rooster2	Elsio	***Kotem1***	***Kotem2***	Meers2	***Maasband***	Decup	Meeswijk	Elba	Bichterweert	***Heppeneert***	Rooster1
Elsio	1				2												
Kotem1																	
Kotem2	2	1				1											
Meers1		3				1	7	5	11	13	6	6	7	5	6	2	1
Meers2				2													
Meers3	2				2												
Maasband																	
Decup	1	1	2			1	2	4	2	2		5	9	1		2	3
Mazenhoven																	
Meeswijk																	
Elba							2			1	5	2	1	4		5	2
Bichterweert																	
Elerweert	1			1													
Heppeneert	2	2	1		1												
Rooster1																2	
Rooster2	2	2	1	6	4	1	1				3						

[**Figure 6.29**](#) **A plant metapopulation.** (a) Locations of the subpopulations, on the River Meuse in Belgium, of a metapopulation of the bank-side plant, *Erysimum cheiranthoides*. The river flows from south to north. (b) Table showing the assignment, where possible, of individuals sampled in 2006 and 2007 to populations from 2005 and 2006, respectively. Colonists (individuals reaching a vacant site) and recolonised sites are shown in bold italics.

Source: After Honnay *et al.* (2009).

Moreover, as Bullock *et al.* (2002) point out, of the plant studies that have documented patch extinctions and colonisations, the vast majority have been in recently emerged patches (the early stages of succession, see Chapter 18). Extinctions mostly occur when the vegetation in a patch develops to a state where it is no longer suitable for the plant species in question, and that patch is therefore also not suitable for recolonisation by the same species. This is ‘habitat tracking’ (Harrison & Taylor, 1997) rather than the repeated extinction and recolonisation of the same habitat that is central to the concept of a metapopulation.

6.7.4 Metapopulation dynamics

a case study: the Glanville fritillary on Finnish islands

We can understand many aspects of the dynamics of metapopulations by focusing on studies of one key system. As well as playing a central role in the development of metapopulation theory, Ilkka Hanski initiated in 1993, and then guided, a long-term study of metapopulations of the Glanville fritillary butterfly, *Melitaea cinxia*, on the Åland islands in Finland (Hanski *et al.*, 2017). The butterfly inhabits dry meadows that support at least one of its two larval host plants, the ribwort plantain *Plantago lanceolata* and the spiked speedwell *Veronica spicata*. The whole area, 50 km × 70 km, has been mapped twice, in 1993 and 1998–99, and monitored since, and consists of around 4500 patches (areas containing the host plants). The patches are small. They have a median size of about 0.06 ha and fewer than 1% of them extend to 2 ha. Thus, the distribution of the species is indeed highly fragmented, covering only around 1% of the total area. The team studying them, however, do not consider the butterflies to comprise a single metapopulation. Instead they group them into 125 semi-independent networks of patches (Figure 6.30a) with each network described as a metapopulation. This itself makes an important point. The clustering is based on the typical lifetime dispersal of the butterflies – mostly limited to 2–3 km. The patches within each cluster, each metapopulation, have a good chance of exchanging individuals within a generation, and hence of providing colonisers that re-establish a subpopulation on a patch, following an earlier extinction there. By contrast, nearby metapopulations exchange individuals only occasionally. The point to understand is that there is no single, objective threshold distance that determines whether two patches are members of the same or different metapopulations. Drawing a line around a metapopulation, like drawing a line around a conventional population, is more often than not a matter of pragmatism. But we cannot, and should not, shy away from judgement calls that apply useful structures to natural patterns filled with shades of grey.

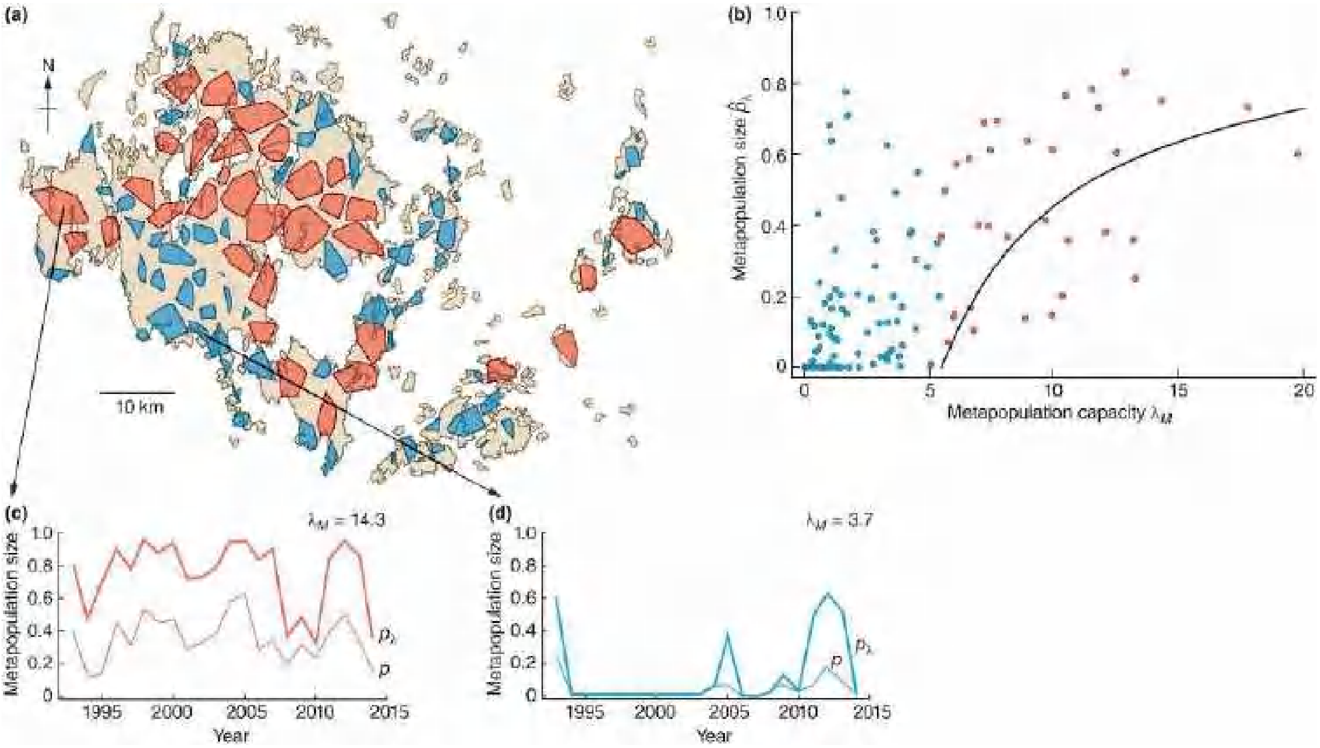


Figure 6.30 **Hanski's metapopulation of the Glanville fritillary.** (a) Map of the Åland islands, off the south-west coast of Finland, with locations and extents of the 125 habitat patches where metapopulations of the Glanville fritillary butterfly, *Melitaea cinxia*, have been identified (smallest not visible): red above the extinction threshold (see text), blue below. (b) The relationship between the size and capacity of the metapopulations, along with data for each of the metapopulations, with the same colour-coding as in (a). The line is the best fit from field data of a variant of Equation 6.6. It crosses the horizontal axis at the extinction threshold, 5.47. (c) and (d) show illustrative examples of the variations in size of two of the metapopulations (in terms of both the fraction of patches occupied, p, and the patch-weighted fraction, p_λ), as indicated, above and below the extinction threshold, respectively. The terms 'extinction threshold' and 'metapopulation capacity' are explained in the main text.

Source: After Hanski *et al.* (2017).

p_λ : metapopulation size accounting for patch variation

The most fundamental aspects of the dynamics of a metapopulation, just like any population, concern its size, variations in its size and its persistence over time. The most obvious measure of the size of a metapopulation, as described above, is p, the fraction of patches occupied. However, subsequent elaborations of the basic theory have acknowledged, as we did previously, that not all patches are equal – some are bigger than others, some are better connected. A more satisfactory alternative measure of metapopulation size therefore is denoted by p_λ, which assigns different weightings to different patches depending on their contribution to the overall dynamics of the metapopulation (a heavier weighting for bigger, more connected patches, etc.). Without going into mathematical details (see Hanski & Ovaskainen, 2000) the equilibrium value of p_λ may be given by:

$$p_\lambda^* = 1 - \delta/\lambda_M. \qquad (6.6)$$

Here, δ is the extinction threshold, the ratio of extinction to colonisation rates, e/c, as previously. This is a property of the particular species concerned, and hence applicable to all of its metapopulations. λ_M is the so-called *metapopulation capacity*, and it is this that aims to integrate the effects of the areas and spatial locations of patches on the capacity of a particular network to sustain a viable metapopulation, as explained next (Hanski & Ovaskainen, 2000).

the metapopulation capacity

The key to the successful application of these ideas is having reliable estimates for the parameters in Equation 6.6. In the case of the Glanville fritillary in the Åland islands, the extinction threshold, δ, could be estimated directly from the colonisation and extinction rates, c and e (monitored through the massive effort invested in surveying all patches annually) or indirectly from data on individual patches. The two methods produced estimates that were highly correlated (0.97). To estimate the metapopulation capacity, it is necessary to make some assumptions about the dependence of extinction and colonisation rates on patch size and location, and then to carry out some simple mathematical matrix calculations (details unimportant). Here it was assumed, for any patch, i:

$$\text{Extinction rate}_i = e/A_i \text{ and,} \qquad (6.7)$$

$$\text{Colonization rate}_i = c\sum_{j\neq i} \exp\left(-d_{ij}/\theta\right) A_j p_j(t). \qquad (6.8)$$

That is, the extinction rate of a patch depends on the species-specific constant e, but also declines with increasing patch area, A_i. The colonisation rate of a patch from all other patches, j, depends on the species-specific constant c, but declines with the distance between patch i and the other patches, d_{ij}, relative to the species' average dispersal distance, θ, while also increasing with the areas of those other patches, A_j, and the fraction of them currently occupied, $p_j\,(t)$, since these codetermine the size of the source of new colonisers. These colonisation and extinction rates then contribute to a matrix, the elements of which describe the net effect of each patch on each other patch. As a final step, a particular property of that matrix, the 'leading eigenvalue', is the metapopulation capacity – largest when patches are themselves large and well connected.

dynamics of the Glanville fritillary

It is apparent now from Equation 6.6 that for a metapopulation to persist, p^*_λ must be positive, and hence the metapopulation capacity must exceed the species' extinction threshold. Equation 6.6 therefore expresses formally what we would expect: that metapopulations are more likely to persist (and will be larger if they do persist) when colonisation rates are high and extinction rates low, and when individual patches are large and well connected. Crucially, though, in the case of the Glanville fritillary it was also possible to test our understanding of their metapopulation dynamics in a quantitative way. We can see, first, that the estimated extinction threshold (5.47) did a good job in predicting metapopulation viability (Figure 6.30b). All networks with an estimated capacity above the threshold supported a viable metapopulation in the sense that its size, the mean patch occupancy, $\hat{p}_\lambda$, was positive. (Note that $\hat{p}_\lambda$ is based on the observed occupancies of individual patches within a network but applies weightings to those patches when computing a mean occupancy, calculated based on their size and connectivity.) What's more, the most important explanatory variable in accounting for variations in population size overall was the metapopulation capacity (around 40% of the total: $F_{1,123} = 84.16$, $P < 10^{-14}$), and it did particularly well in accounting for variations above the threshold, where the connectivity of patches had negligible effect – these viable metapopulations behaved, effectively, like independent entities.

metapopulations of metapopulations

Below the threshold, on the other hand, while simple application of the threshold criterion suggests a non-viable metapopulation, in many cases a positive value for $\hat{p}_\lambda$ was recorded overall. This, in fact, is not unexpected, since many metapopulations that would not be viable in isolation will be 'rescued' at least temporarily by colonisers from nearby metapopulations. This is emphasised by noting that of the 92 'non-viable' networks, 79 (86%) were extinct for at least five years out of the 22 they were observed, compared with only five (15%) of the 33 viable networks. Furthermore, connectivity to nearby networks was highly influential in accounting for variations in $\hat{p}_\lambda$ below the threshold ($t = 3.13$, $P = 0.002$). In this very important sense, the Glanville fritillary on the Åland islands exists, as do many populations, as a metapopulation of metapopulations. These contrasts are further illustrated in Figure 6.30c and d, which show time series from a viable and non-viable network, respectively – in the first case, fluctuating but never coming close to extinction, and in the second case, suffering extinction twice, once for an extended period. Of course, most fundamentally of all perhaps, these results demonstrate again how whole metapopulations can be stable when their individual subpopulations are not, and metapopulations of metapopulations more stable still.

alternative stable equilibria

The Glanville fritillary study also illustrates a tendency for metapopulations to exhibit alternative stable equilibria. As the fraction of patches occupied by the metapopulation, p, increases, there should on average be more migrants, more immigration into patches, and hence larger local subpopulations (confirmed for the Glanville fritillary – Hanski *et al.*, 1995). Thus, the extinction rate, e, of local subpopulations should arguably not be constant as it is in the simplest models, but decline as p increases. Models incorporating this effect (Hanski, 1991 ; Hanski & Gyllenberg, 1993) typically give rise to an intermediate unstable threshold value of p. Above the threshold, the sizes of local populations, and hence the flow of migrants, are sufficiently large for many subpopulations to persist and generate further potential migrants, further sustaining the subpopulations. But below the threshold, the average size of local populations is too low and, with little flow of migrants, their rate of extinction too high. Different metapopulations of the same species might therefore be expected to occupy either a high or a low fraction of their habitable patches (the alternative stable equilibria) but not an intermediate fraction (close to the threshold). Such a bimodal distribution is indeed apparent for the Glanville fritillary (Figure 6.31).

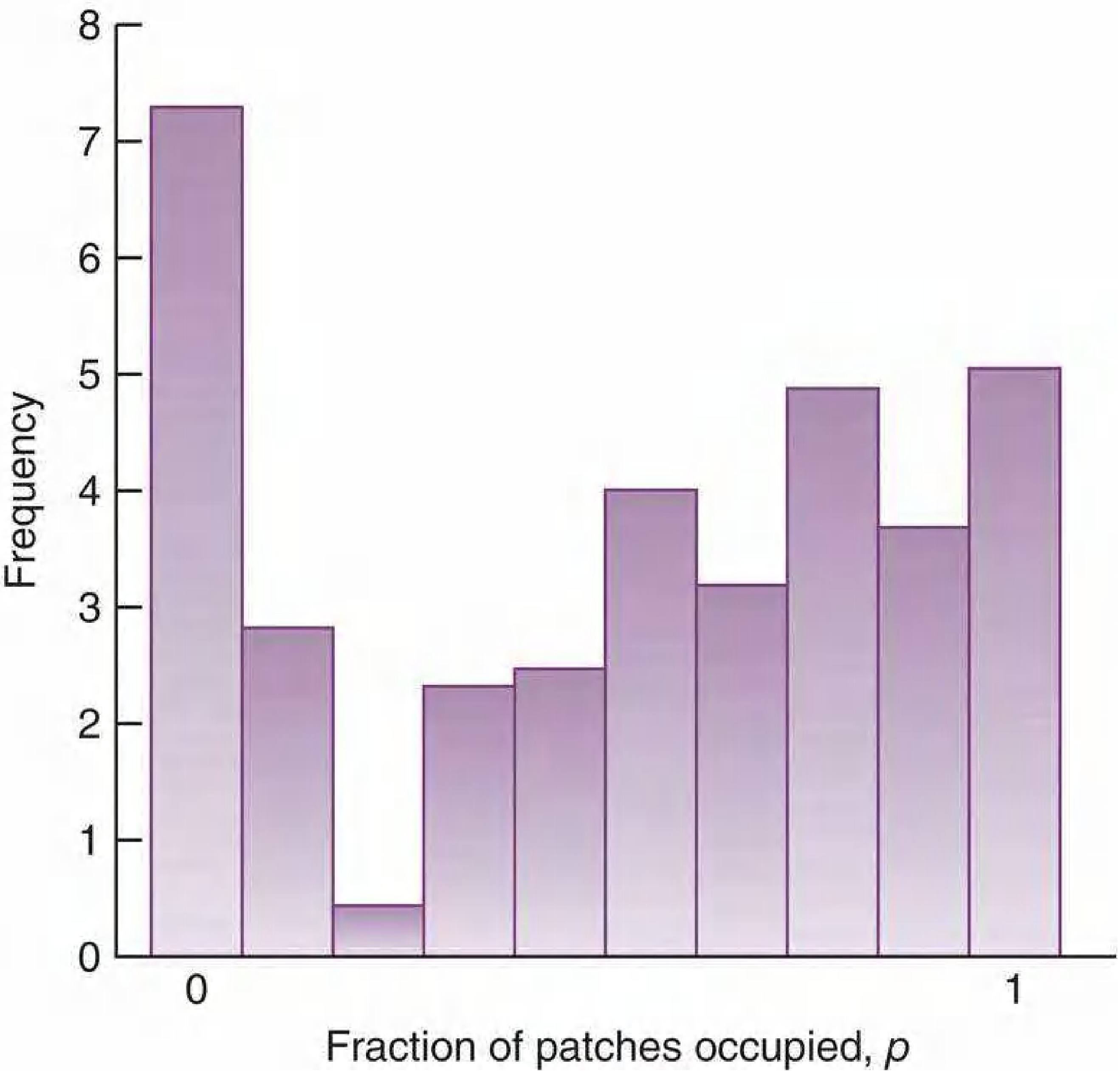


Figure 6.31 Alternative stable states for the Glanville fritillary metapopulation. The bimodal frequency distribution of patch occupancy (proportion of habitable patches occupied, *p*) amongst different metapopulations of the Glanville fritillary (*Melitaea cinxia*) on the Å land islands in Finland.

Source: After Hanski *et al*. (1995).

important genetic effects on ecological dynamics

Finally, the Glanville fritillary study illustrates, in the context of metapopulations, a point of much wider significance, namely that evolutionary genetic dynamics can have a significant effect on the ecological dynamics of populations. That is, ecological and evolutionary time scales are not so different as they are often imagined to be (Saccheri & Hanski, 2006). We can see this in the present case by noting that while, above the threshold, the metapopulation capacity accounted for around 15% of the variation in metapopulation size, a genetic effect – the alleles present at a particular gene – accounted for twice that: 30% of the variation. Specifically, individuals with a C substituted for an A in the genetic code at one particular location in the phosphoglucose isomerase (*Pgi*) gene (whether they are homozygous, CC, or heterozygous, AC) have a higher flight metabolic rate and disperse further than AA homozygotes (though they no doubt pay an energetic cost for this activity). It is the proportion of the population that are either CC or CA that did so well in accounting for variations in metapopulation size. These more dispersive individuals were especially prevalent in newly established subpopulations and in metapopulations with higher turnover rates (many colonisations and extinctions). A very plausible explanation for these patterns, therefore, is that natural selection favours CC/AC over AA in metapopulations with small patches, since these require high rates of colonisation both to persist and to be larger if they do persist. On the other hand, in metapopulations with large patches, the premium on being able to disperse is not so high (less advantage in colonising or preventing extinction through a rescue effect) and the energetic costs of the strategy take on greater importance. Hence, when a given metapopulation capacity is arrived at by having larger numbers of smaller patches, the C-variants are favoured, there are high rates of dispersal and colonisation, and a relatively high metapopulation size. But when that same metapopulation capacity is arrived at by having a smaller number of larger patches, AA homozygotes are favoured, there is less dispersal, and metapopulation sizes are smaller despite individual patches being more stable (Figure 6.32). We see then that when the whole concept of a population is spatially explicit, as it is in a metapopulation, then not only dispersal, but the evolutionary dynamics of dispersal, can play a key role in driving variations in population size.

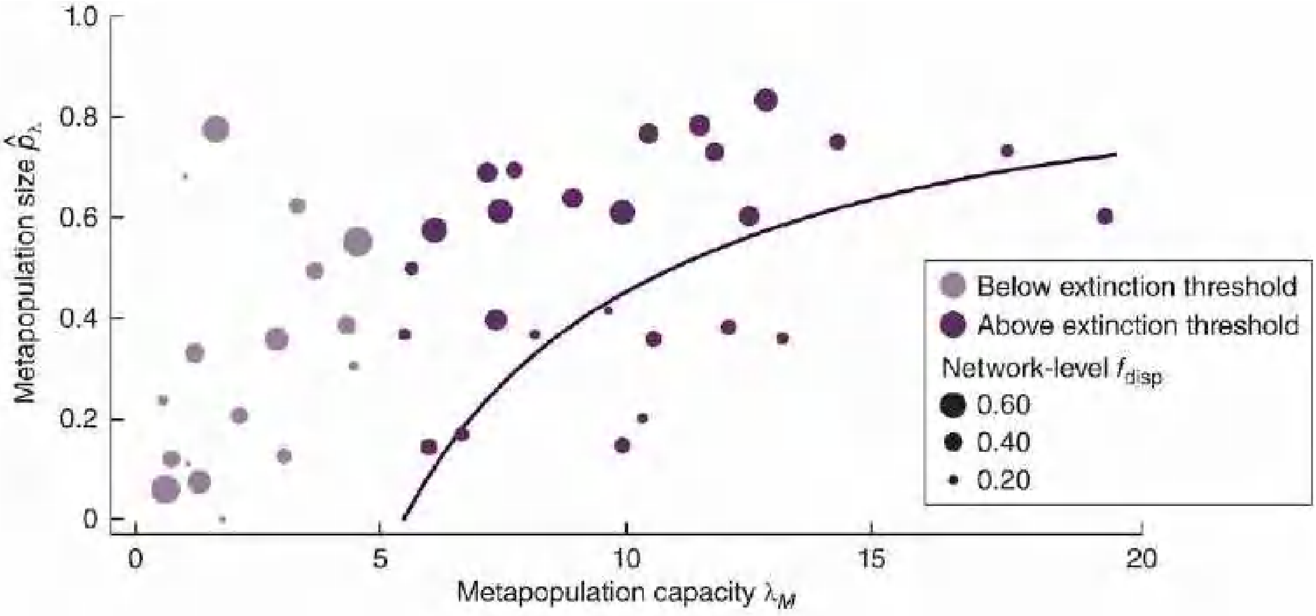


Figure 6.32 Genetic effects on the dynamics of the Glanville fritillary metapopulation. The relationship between metapopulation size and capacity for the networks of the Glanville fritillary on the Åland islands where sufficient genetic data were available. The fitted line is the same as in Figure 6.30. Pale and dark purple dots distinguish those below and above the extinction threshold, respectively. Larger dots have a higher proportion (f_{disp}) of the dispersive genotypes at the *Pgi* locus (CC and AC), as indicated.

Source: After Hanski *et al.* (2017).

APPLICATION 6.9 Metapopulation capacities for birds and the giant panda

Recall that the concept of a metapopulation capacity integrates the effects of the areas and spatial locations of patches on the capacity of a network to sustain a viable metapopulation. This has obvious applicability, if, for conservation purposes, we wish to assess the risks of extinction posed to a species in a fragmented landscape. This can be useful both in assessing the current conservation status of a species and in planning future conservation measures. One study, for example, examined recent contractions in the distributions of four species of birds in the central American highlands (Figure 6.33). The species were the azure-rumped tanager (*Tangara cabanisi*), classified as 'endangered' on the International Union for Conservation of Nature's (IUCN) Red List (see Section 15.4.2), the pink-headed warbler (*Ergaticus versicolor*) classified as 'vulnerable', and the rufous-browed wren (*Troglodytes rufociliatus*) and rufous-collared robin (*Turdus rufitorques*), both classified as 'least concern'. The comparisons were between their distributions assessed up to the 1990s in order to determine their conservation status, and a reassessment by Harris and Pimm (2008) based on satellite imagery analysis of the distribution of the forest habitats within their known altitudinal ranges. It is clear from the figure that the reassessment has led to a significant contraction in the areas of their available habitat – down to between 15% and 30% of what they were previously. But when the metapopulation capacity was calculated in each case, the reductions were down to 1% and 6% of the original. Indeed, the robin, officially of least concern, had a metapopulation capacity almost exactly the same as that of the 'vulnerable' warbler. Its habitat in the area now consists almost entirely of small fragments. Thus, by taking fragmentation into account in an informed and quantitative way, the metapopulation capacity guards against potential complacency and gives us an arguably more realistic measure of the risks that species face.

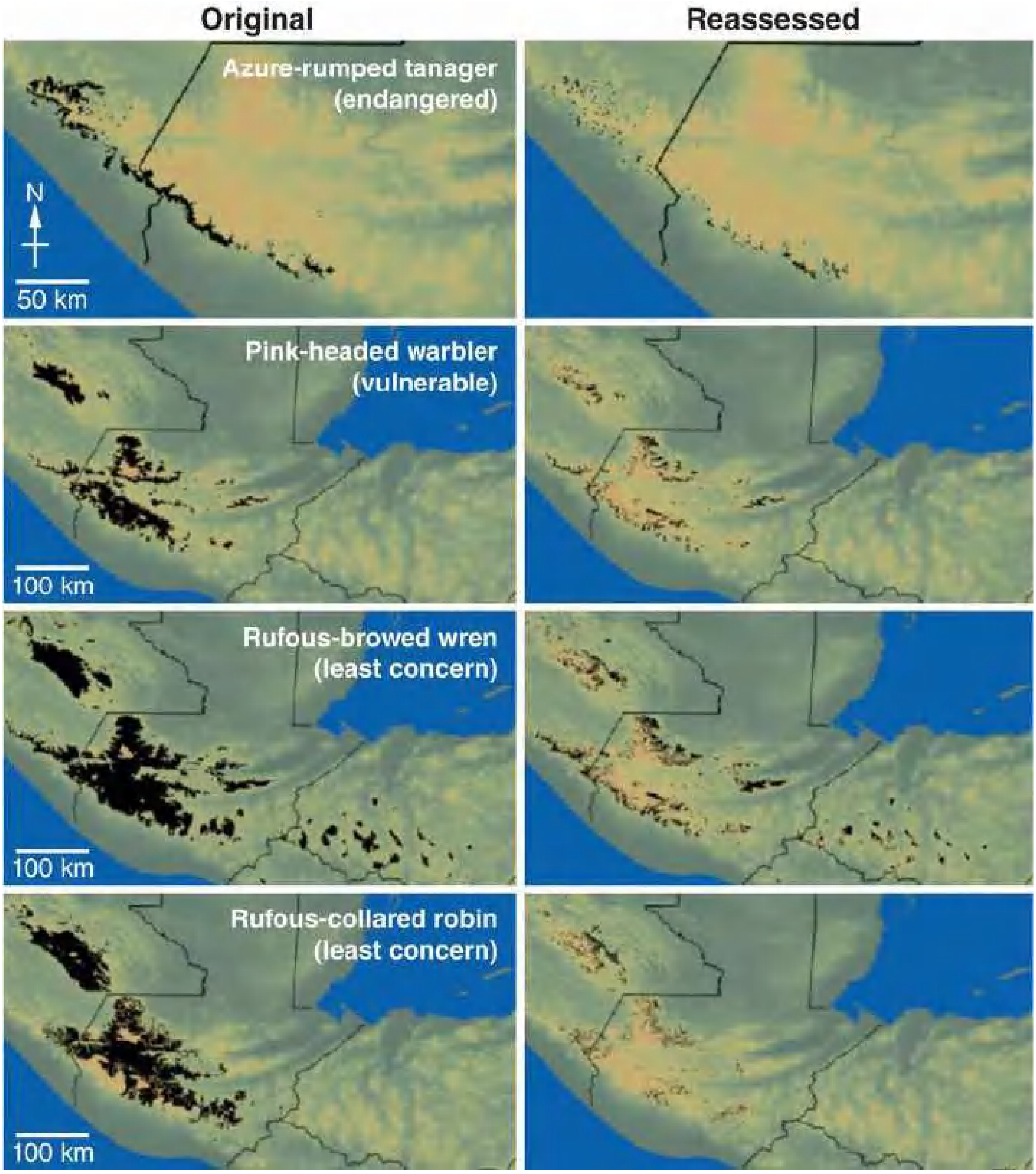


Figure 6.33 Contractions in the ranges of four bird species that have especially profound effects on their metapopulation capacities. Maps of the distributions of available habitat (marked in black) of four species of birds, as indicated, in the Central American highlands, historically and following a reassessment of available habitat. The border that appears in all maps is between Mexico to the north-west and Guatemala to the south-east.

Source: After Schnell *et al.* (2013).

A second example has looked to the future in predicting the changing distribution of the giant panda, *Ailuropoda melanoleuca*, in the Minshan Mountains in China between 2011 and 2100, and expressing this, too, in terms of metapopulation capacity. These projections estimate around a 16% loss in habitable area, with around 18% of this occurring in locations currently covered by a network of nature reserves (Figure 6.34). However, these same projections suggest substantial fragmentation of panda habitat, with mean patch size dropping from 3859 to 1568 ha, and the percentage of patches below the estimated minimum size for panda persistence (400 ha) rising from 3% to more than 11%. Translating these into estimates of metapopulation capacity again paints a bleaker picture than area alone. Currently, the metapopulation capacity of the protected areas is 9% lower than areas outside. But the projections for 2100 suggest that there will be a 35% decline in the capacity of the protected areas, making them 40% lower than areas outside. Given that nature reserves are of the greatest importance, compared with areas outside where panda conservation has to compete with other interests, especially farming, these figures are worrying.

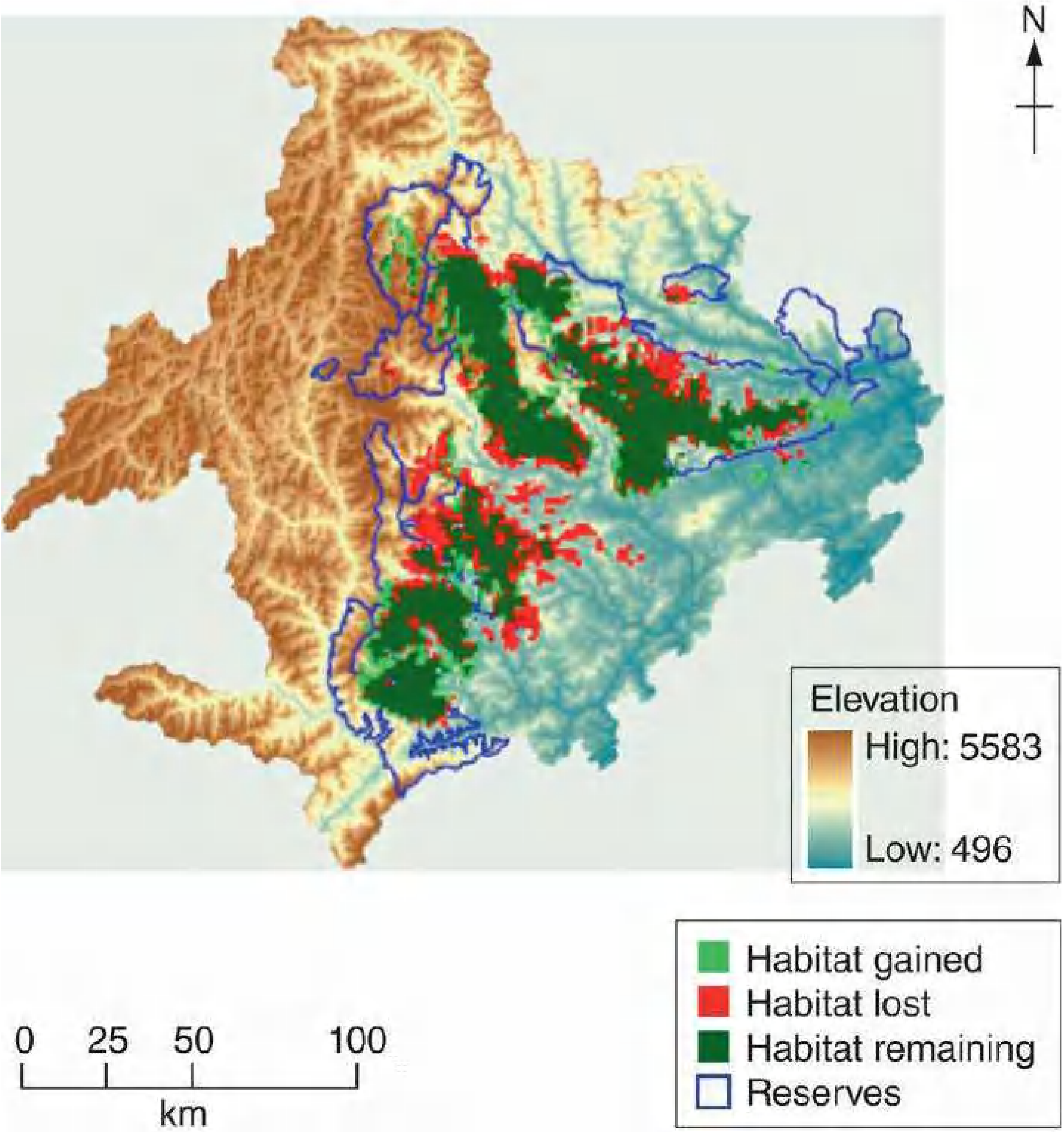


Figure 6.34 The changing distribution of the giant panda in China that will affect their metapopulation capacity. Projected gains and losses of the giant panda, *Ailuropoda melanoleuca*, habitat in the Minshan Mountains between 2011 and 2100 based on likely climate change scenarios (for location in China, see Figure 6.3).

Source: After Shen *et al.* (2015).

We conclude this chapter, therefore, with confirmation that the movement of individuals often does far more than redistribute them. It may alter local abundances significantly, but when populations are distributed patchily – and at some scale, all populations are distributed patchily – movement may be the glue that binds the dynamics of subpopulations together and determines the persistence and dynamics of the whole.

Chapter 7

Life History Ecology and Evolution

7.1 Introduction

An organism's life history is its lifetime pattern of growth, differentiation, storage and reproduction. In Chapter 4, we saw something of the variety in these patterns and what their consequences may be in terms of population rates of increase. We can simply accept this variation. But in the spirit of nothing making sense except in the light of evolution (Chapter 1), we can go on to ask what evolutionary pressures have given rise to this variety, and what sustains it. In turn, this feeds back to an understanding of patterns in life history ecology: what types of life history are found where, and in what types of organism.

three types of question

There are at least three different types of question that are commonly asked. The first concerns individual life history traits. How is it that swifts, for example, usually produce clutches of three eggs, when other birds produce larger clutches and the swifts themselves are physiologically capable of doing so? Can we establish that *this* clutch size is ultimately the most productive – the fittest in evolutionary terms – and what is it about this particular clutch size that makes it so?

The second question concerns links between life history traits. How is it, for example, that the ratio between age at maturity and average lifespan is often roughly constant within a group of organisms but markedly different between groups (e.g. mammals 1.3, fish 0.45)? What is the basis for the constancy within a group of related organisms? What is the basis for differences between groups?

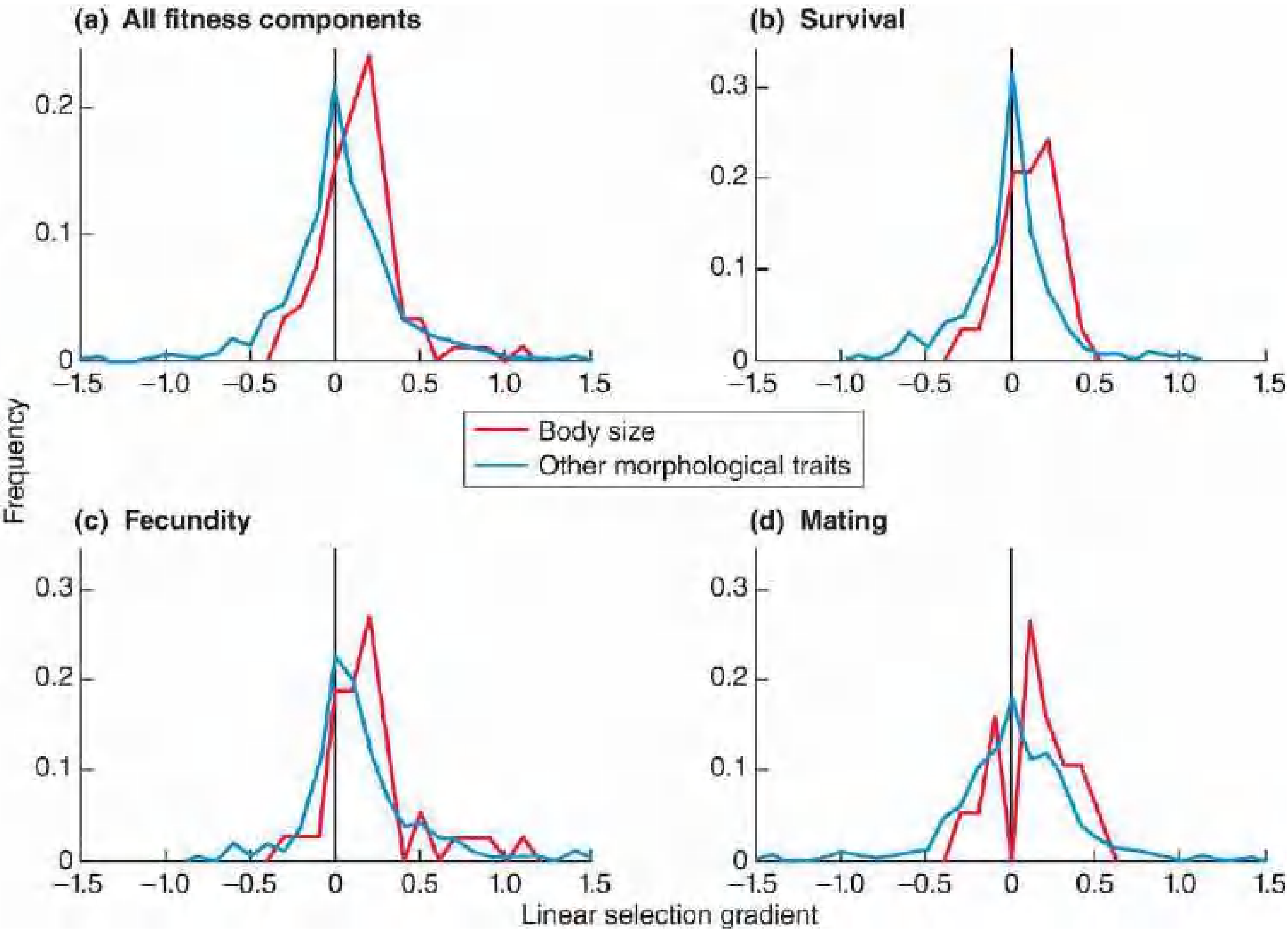
Thirdly, we can ask questions about links between life histories and habitats, where 'habitat', of course, includes other species with which a focal species coexists. Orchids, for example, produce vast numbers of tiny seeds, whereas tropical *Mora* trees produce just a few enormous ones. Can the difference be related directly to differences in the habitats that they occupy, or to any other differences between them?

The study of the evolution of life histories is thus a search for patterns – and for explanations for those patterns. Remember, however, that every life history, and every habitat, is unique. In order to find ways in which life histories might be grouped, classified and compared, we must find ways of describing them that apply to all life histories and all habitats. It is also important to realise that the possession of one life history trait may limit the possible range of some other trait, and the morphology and physiology of an organism may limit the possible range of all its life history traits. The most that natural selection can do is to favour, in a particular environment with its many, often conflicting demands, the life history that has been most (not ‘perfectly’) successful, overall, at leaving descendants in the past. Nonetheless, most of the successes in the search for an understanding of life history evolution have been based on the idea of optimisation: establishing that observed combinations of life history traits are those with the highest fitness (Stearns, 2000).

7.2 The components of life histories

growth and size – is bigger better?

Of the important components of any organism’s life history, the size an individual attains is perhaps the most apparent. Large size may increase an organism’s competitive ability, or increase its success as a predator, or decrease its vulnerability to predation, and hence increase its survival. Stored energy and/or resources will also be of benefit to organisms that pass through periods of reduced or irregular nutrient supply (probably true of most species at some time). Finally, larger individuals within a species usually produce more offspring. On the other hand, size can increase some risks: a larger tree is more likely to be felled in a gale, many predators exhibit a preference for larger prey, and larger individuals typically require more resources and may therefore be more prone to a shortage of them. Perhaps most important: becoming large itself takes time, during which an individual may be prone to a variety of mortality risks. An intermediate, not a maximum, size may therefore be optimal. In practice, however, most studies have found a larger rather than an intermediate size to be favoured (Figure 7.1). This provides an explanation for ‘Cope’s rule’ (Hone & Benton, 2004) – the tendency for the species in a lineage to increase in size over evolutionary time – but does not really explain why more contemporary species are not at or close to their maximum size (Kingsolver & Pfennig, 2008 .)



[Figure 7.1](#) Studies of selection show a tendency for larger body size to be favoured but intermediate values of other morphological traits. Results of a survey from the published literature of studies measuring the strength of selection on body size (red line; a positive value means selection for increased body size) and other morphological traits (blue line) when this was determined by reference to (a) all fitness components combined, (b) survival, (c) fecundity and (d) mating success. In all cases, there was an overall tendency for larger size to be favoured (giving rise to higher fitness), whereas for traits other than body size the distribution was symmetrical around zero (no change). The selection gradient is the change in relative fitness that results from 1 standard deviation of change in a trait, such that a value of 0.5 means that a change of one standard deviation increases fitness by 50%.

Source: After Kingsolver & Pfennig (2008).

development, differentiation and reproductive allocation

Development is the progressive differentiation of parts, enabling an organism to do different things at different stages in its life history. Hence, rapid development can increase fitness because it leads to the rapid initiation of reproduction. As we have seen, reproduction itself may occur in one terminal burst (semelparity) or as a series of repeated events (iteroparity). Amongst iteroparous organisms, variation is possible in the number of separate clutches of offspring, and all organisms can vary in the number of offspring in a clutch.

The individual offspring can themselves vary in size. Large newly emerged or newly germinated offspring are often better competitors, better at obtaining nutrients and better at surviving in extreme environments. Hence, they often have a better chance themselves of surviving to reproduce.

Combining all of this detail, life histories are often described in terms of a composite measure of reproductive activity known as 'reproductive allocation' (also often called 'reproductive effort'). This is best defined as the proportion of the available resource input that is allocated to reproduction over a defined period of time, but it is often not clear which resource's allocation is the most appropriate to measure – that is, which resource is most limiting and therefore best reflects the effort invested by the organism concerned. Figure 7.2 shows an example comparing the allocation of dry matter on the one hand, and nitrogen on the other, to the various parts of garden tulips, *Tulipa gesneriana*, over their growing season in North Carolina, USA. Broad patterns are the same, but the details differ. In practice, even the better studies usually monitor only the allocation of energy, or just dry weight, to various structures at a number of stages in the organism's life cycle.

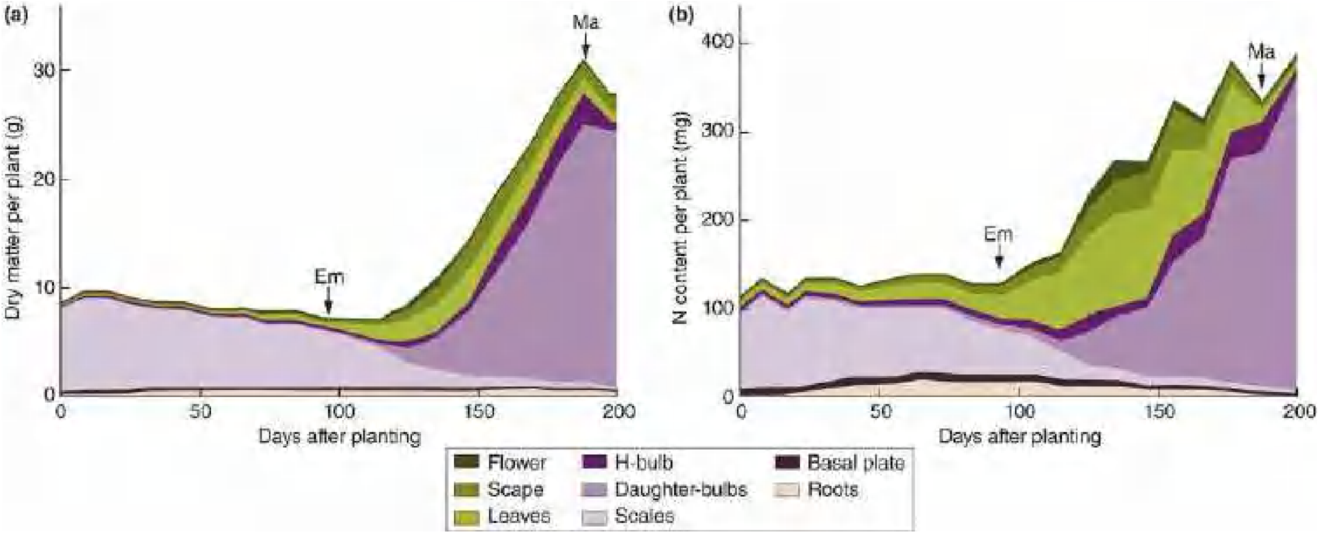


Figure 7.2 Allocations of dry matter and nitrogen to the parts of a plant vary over its lifetime. The allocation of (a) dry matter and (b) nitrogen to the various organs (see key in figure) of the garden tulip, *Tulipia gesneriana*, over the course of the growing season. Em, shoot emergence; Ma, bulb maturity. The scape is the flower-supporting stem; the H-bulb is the outermost, 'tunic' bulb amongst the daughter bulbs.

Source: After Niedziela *et al.* (2015).

7.2.1 Reproductive value

Natural selection favours those individuals that make the greatest proportionate contribution to the future of the population to which they belong. The life history components that we have described affect this contribution, ultimately through their effects on fecundity and survival. It is necessary, though, to combine these effects into a single currency so different life histories may be judged and compared on a common scale. A number of measures of fitness have been used. All the better ones have made use of both fecundity and survival schedules. The intrinsic rate of natural increase, r, the basic reproductive rate, R_0 (see Chapter 4) and lifetime reproductive success have all had their advocates. Here, we will concentrate on 'reproductive value', because it is readily decomposed into current and future contributions to fitness (Fisher, 1930 ; Williams, 1966 ; Kozlowski, 1993). However, for an exploration of the basic patterns in life histories, the similarities between these various measures are far more important than the minor differences between them.

reproductive value described in words

Reproductive value is described in some detail next. For most purposes though, these details can be ignored as long as we remember that: (i) reproductive value at a given age or stage is the sum of the contemporary reproductive output and the residual (i.e. future) reproductive value (RRV); (ii) RRV combines expected future survival and expected future fecundity; (iii) survival and fecundity are combined in a way that takes account of the contribution of an individual to future generations, relative to the contributions of other individuals; and (iv) the life history favoured by natural selection from amongst those available in the population will be the one for which the sum of contemporary output and RRV is highest.

The reproductive value of an individual of age x (RV_x) is the currency by which the worth of a life history in the hands of natural selection may be judged. It is defined in terms of the life-table statistics discussed in Chapter 4. Specifically:

$$RV_x = \sum_{y=x}^{y=y_{\max}} \left(\frac{l_y}{l_x} \cdot m_y \cdot R^{x-y} \right) \tag{7.1}$$

where m_x is the birth rate of the individual in age-class x; l_x is the probability that the individual will survive to age x; R is the net reproductive rate of the whole population per unit time (the time unit here being the age interval); and $\sum$ means 'the sum of'.

To understand this equation, it is easiest to split RV_x into its two components:

$$RV_x = m_x + \sum_{y=x+1}^{y=y_{\max}} \left(\frac{l_y}{l_x} \cdot m_y \cdot R^{x-y} \right). \tag{7.2}$$

Here, m_x, the individual's birth rate at its current age, can be thought of as its *contemporary reproductive output*. What remains is then the *residual reproductive value* (Williams, 1966): the sum of the 'expectations of reproduction' at all subsequent ages, modified in each case by R^{x-y} for reasons described later. The 'expectation of reproduction' for age class y is ($l_y / l_x \cdot (m_y)$), i.e. it is the birth rate of the individual should it reach that age (m_y), discounted by the probability of it doing so given that it has already reached stage x (l_y / l_x).

Reproductive value takes on its simplest form where the overall population size remains approximately constant. In such cases, $R = 1$ and can be ignored. The reproductive value of an individual is then simply its total lifetime expectation of reproductive output (from its current age class and from all subsequent age classes).

However, when the population consistently increases or decreases, this must be taken into account. If the population increases, then $R > 1$ and $R^{x-y} < 1$ (because $x < y$). Hence, the terms in the equation are reduced by R^{x-y} the larger the value of y (the further into the future we go), signifying that future (i.e. 'residual') reproduction adds relatively little to RV_x, because the proportionate contribution to a growing population made by a given reproductive output in the future is relatively small – whereas the offspring from present or early reproduction themselves have an early opportunity to contribute to the growing population. Conversely, if the population decreases, then $R < 1$ and $R^{x-y} > 1$, and the terms in the equation are successively increased, reflecting the greater proportionate contribution of future reproduction.

In any life history, the reproductive values at different ages are intimately connected, in the sense that when natural selection acts to maximise reproductive value at one age, it constrains the values of the life table parameters – and thus reproductive value itself – for subsequent ages. Hence, strictly speaking, natural selection acts ultimately to maximise reproductive value *at birth*, RV_0 (Kozlowski, 1993).

The way in which reproductive value changes with age in two contrasting populations is illustrated in Figure 7.3. It tends to be low for young individuals, since these, individually, have only a low probability of surviving to reproductive maturity. It then rises and eventually peaks as the age of first reproduction is approached, as it becomes increasingly likely that surviving individuals will reproduce, and then declines again to low values for old individuals, since their reproductive output is likely to have declined and their expectation of future reproduction is even lower. The detailed rise and fall, of course, varies with the detailed age- or stage-specific birth or mortality schedules of the species concerned. Note also that there is no contradiction between this pattern (reproductive value low for young individuals) and the fact that natural selection acts ultimately to maximise reproductive value at birth. Natural selection can discriminate only between those options available at that stage.

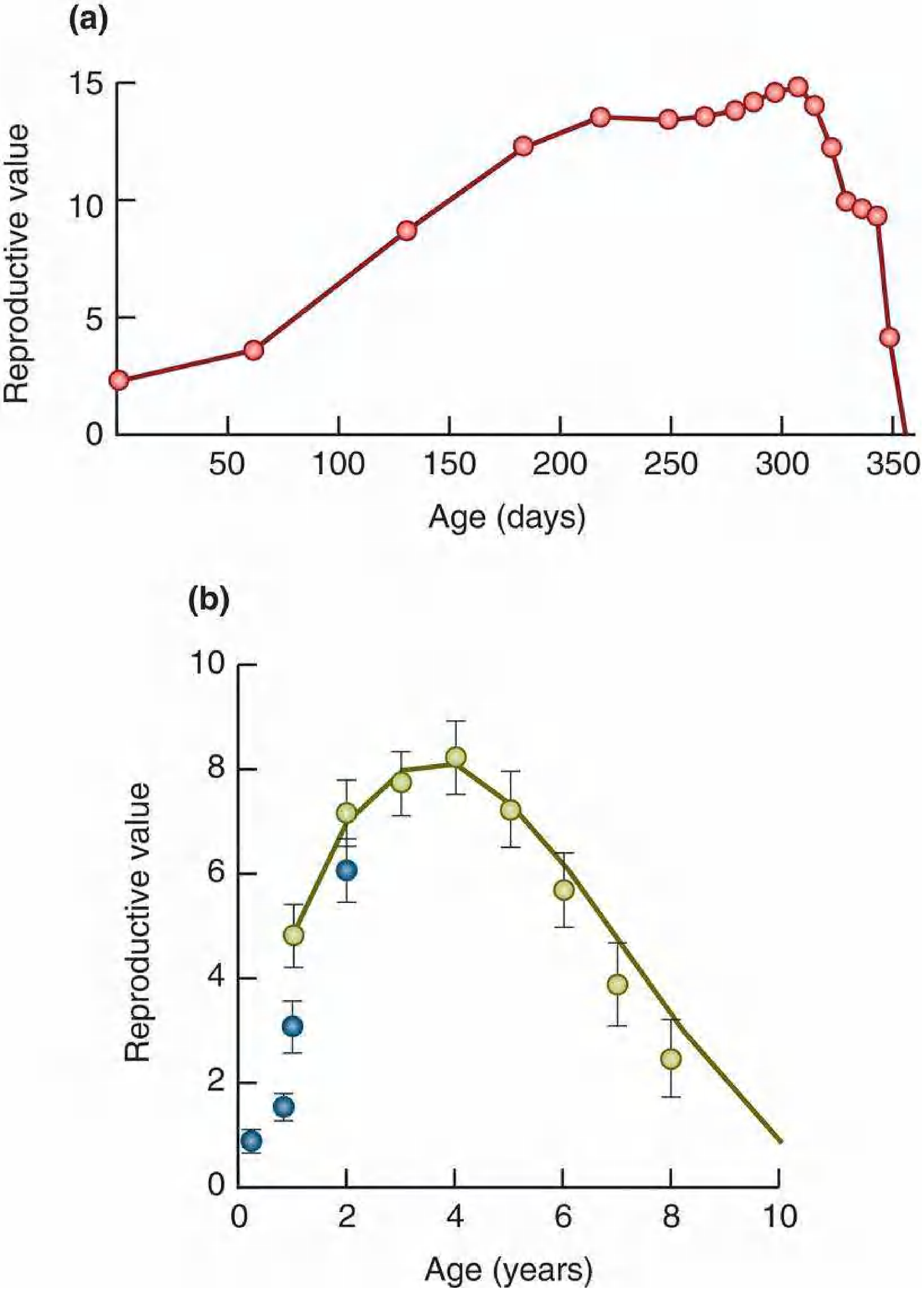

(a)
Reproductive value
15
10
5
0
50
100
150
200
250
300
350
Age (days)
(b)
Reproductive value
10
8
6
4
2
0
0
2
4
6
8
10
Age (years)

Figure 7.3 Reproductive value generally rises and then falls with age. (a) The annual plant *Phlox drummondii*. (b) The sparrowhawk, *Accipiter nisus*, in southern Scotland. Green symbols (±1 SE) refer to breeders only; blue symbols include non-breeders. Note that in both cases the vertical scale is arbitrary, in the sense that the rate of increase (*R*) for the whole population was not known, and a value therefore had to be assumed.

Source: (a) After Leverich & Levin (1979). (b) After Newton & Rothery (1997).

7.3 Trade-offs

In addressing one of our key questions – are there patterns linking life history traits? – we should note immediately that any organism's life history must be a compromise allocation of the resources available to it. Resources devoted to one trait cannot then be invested in other traits. We might therefore expect to see negative relationships between two life history characteristics in which increases in one are associated with decreases in the other as a result of such compromises, and there are certainly examples of this. For instance, trees benefit both from reproducing and from growing (which, amongst other things, enhances future reproduction). But the more seeds trees produce, the less they grow, and the more they grow the less they reproduce (Figure 7.4a). Similarly, squirrels benefit from both surviving and reproducing, but the more they reproduce, the less well they survive (Figure 7.4b). And mountain goats benefit from reproducing both now and in the future, but the more they reproduce now the less likely they are to reproduce the following year (Figure 7.4c).

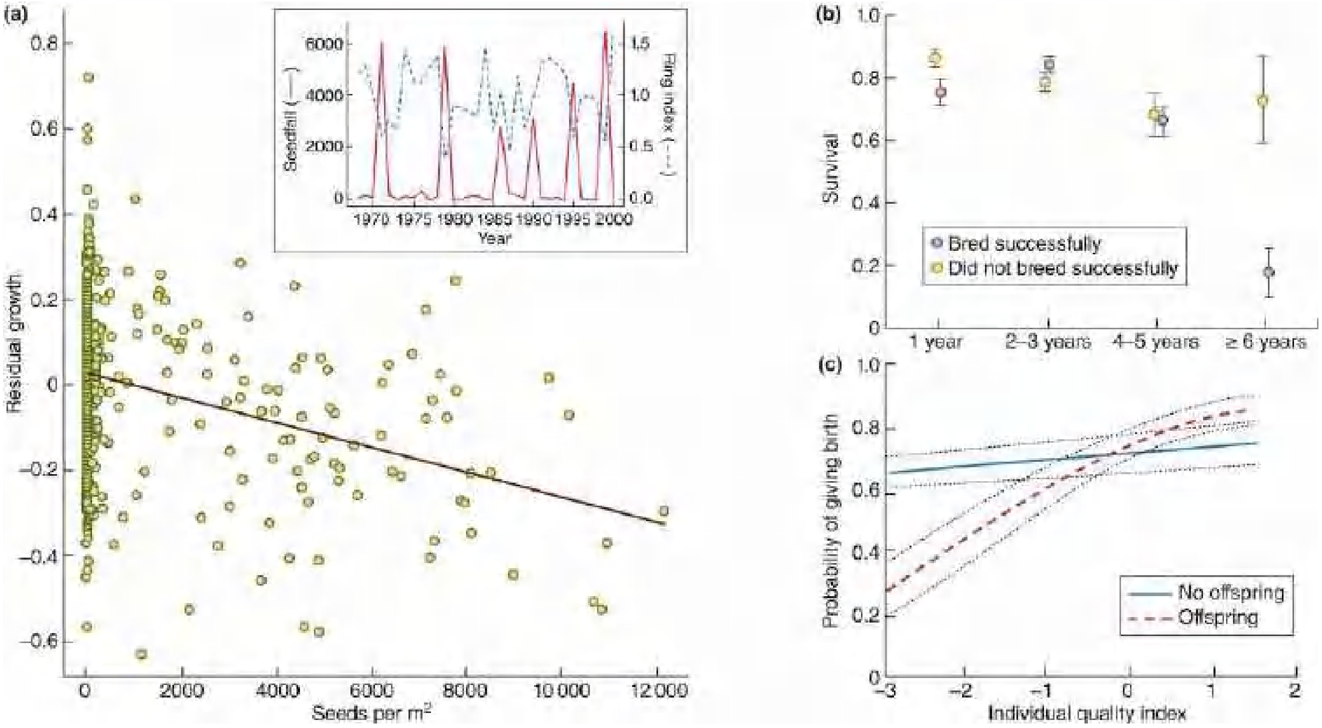


Figure 7.4 Life history trade-offs demonstrate the costs of reproduction. (a) The tree *Nothofagus truncata*, growing in the Orongorongo Valley, New Zealand exhibits ‘mast seeding’ (inset): years of massively increased seed production (i.e. reproduction) exhibited by whole groups of trees growing together. Much of this growth variation was attributable to prevailing weather conditions, but once this had been accounted for in a statistical model, the remaining ‘residual growth’, higher or lower than the weather would predict, was significantly negatively correlated with seed output ($P < 0.0001$): trees that reproduced most grew least; trees that grew most reproduced least. (b) Among North American red squirrels, *Tamiasciurus hudsonicus*, from the southern Yukon, Canada, survival was lower in females that bred successfully in the previous season (means ± SEs) than in those that did not, but only among the youngest and oldest: those in their first year or more than five years old. (c) In mountain goats, *Oreamnos americanus*, from Alberta, Canada, the probability of giving birth one year was reduced among females that had given birth the previous year, but only for females in relatively poor condition, measured on a scale according to which females were either relatively light and subordinate (negative values) or heavy and dominant (positive values). The lines are not the original data but the output of best-fit statistical models, with associated standard errors.

Source: (a) After Monks & Kelly (2006). (b) After Descamps *et al.* (2009). (c) After Hamel *et al.* (2009).

7.3.1 Observing trade-offs

These statistical associations (negative correlations) are conventionally said to represent ‘trade-offs’ between the two traits, though we also use the term trade-off to describe the functional relationship between two traits: the idea that increased investment in one trait *gives rise* to decreased investment in other traits (Roff & Fairbairn, 2007). However, the negative correlations – evidence of an underlying trade-off – are not always easy to obtain. They may often be hidden within other patterns or observable only in some circumstances. For example, Figure 7.4 shows that only the youngest and oldest squirrels paid a survival cost of reproduction (Figure 7.4b), and only mountain goats in the poorest condition had their chances of future reproduction reduced when they produced offspring (Figure 7.4c).

hidden trade-offs and the Y model

More generally, a trade-off may be hidden, or there may be a positive, not a negative correlation between two apparently alternative processes, whenever there is variation between individuals in the amount of resource that they acquire and then have at their disposal. The simple idea behind this is most often referred to as the 'Y model' (de Jong & van Noordwijk, 1992 ; Roff & Fairbairn, 2007), for reasons that are readily apparent in Figure 7.5. The resources acquired are represented by the stem of the Y, and if a proportion of these, C, are allocated to, for example, fecundity, then that leaves a remaining proportion, $1 - C$, to be allocated to, say, growth. However, if two individuals (or populations, or species) differ in the amount of resource they have acquired – the width of the stem of the Y – then any association between fecundity and growth depends on the differentials in both allocation *and* acquisition.

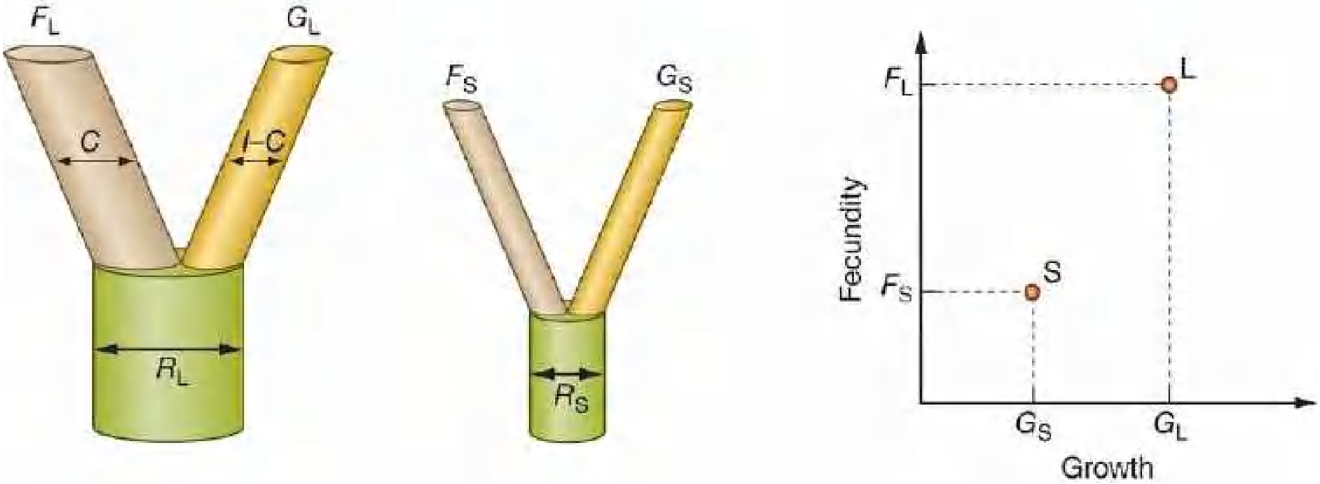


Figure 7.5 The 'Y model' of de Jong and van Noordwijk (1992). The two letter Ys (to the left) represent two individuals (or populations, or species), L and S, that acquire a large or a small amount of resource (R_L and R_S, respectively). They allocate this to either fecundity, F, or growth, G, according to a trade-off in which the proportion that is allocated to fecundity, C, leaves a remainder for allocation to growth, $1 - C$. In this case, there is indeed an underlying trade-off. Individual L, by allocating a higher proportion of its available resource to F than individual S, allocates a smaller proportion to G. However, because of the differential in resource acquired, $F_L > F_S$ and $G_L > G_S$, there is a positive, not a negative association between F and G when individuals are compared, shown to the right.

The most obvious potential outcome is that those that have acquired the most resource will allocate the most to everything, while those that have acquired least will allocate least, even if there is an underlying trade-off, leading to a positive rather than a negative association. We see the principle behind this in Figure 7.5, and an example of it in Figure 7.6, where the snakes that produced larger litters also recovered from breeding more rapidly, ready to breed again. Moreover, the pattern of allocation, the trade-off, may itself depend on the level of acquisition. These complexities, and possible ways of disentangling them, are discussed, for example, by Roff and Fairbairn (2007) and by Saeki *et al.* (2014).

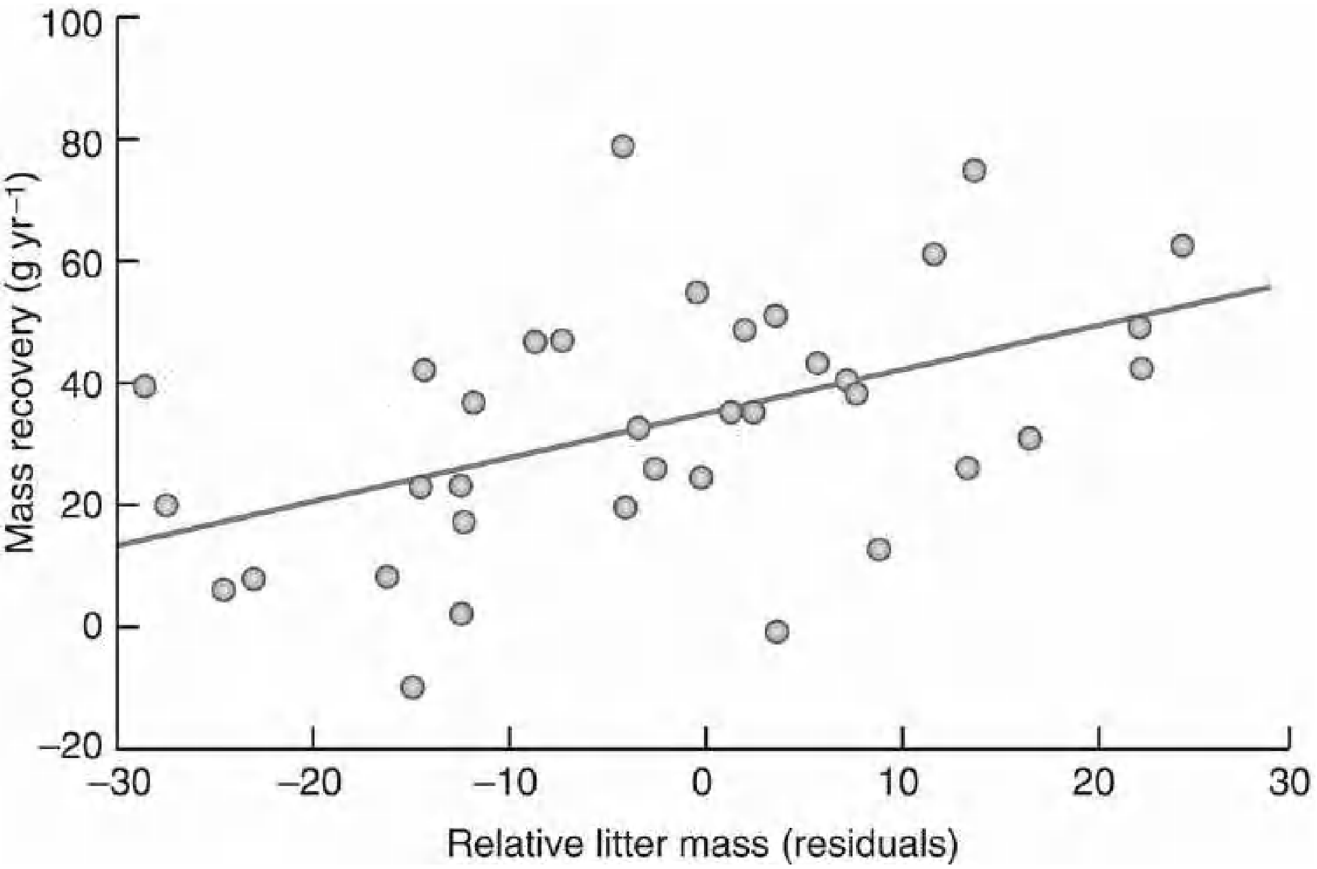


Figure 7.6 Snakes that produced larger litters also recovered quicker from reproduction. Female aspic vipers (*Vipera aspis*) that produced larger litters ('relative' litter mass because total female mass was taken into account) also recovered in mass more rapidly from reproduction (not 'relative' because mass recovery was not affected by size) ($r = 0.43$; $P = 0.01$).

Source: After Bonnet *et al.* (2002).

We should also note that the words 'investment' and 'allocation' have sometimes been used synonymously, but sometimes a distinction is drawn between the *absolute* amount of resource allocated, on the one hand, and investment as a *proportion* of available resource, on the other. In this sense, if two animals allocate the same amount of resource to reproduction, but the first has less total resource at its disposal, then it is making a greater investment in reproduction.

experimental manipulations

One approach to overcoming these difficulties is to use experimental manipulation to reveal a trade-off. The great advantage of this over simple observation is that individuals are assigned treatments by the experimenter rather than differing from one another, for instance, in the quantity of resource that they have at their disposal. In one example, juvenile three-spined sticklebacks, *Gasterosteus aculeatus*, were subjected to a brief alteration in their rearing temperature, either boosting their early growth rate or holding it back. In all cases, though, either by investing energy in catching up, or by saving energy that was not required, the fish achieved the same final weight by around week 19 of their life. This revealed a trade-off between their growth rate and their subsequent survival (Figure 7.7). Those that had been forced into investing more in growth experienced a 15% reduction in median lifespan. Those that had needed to invest less experienced a 31% increase.

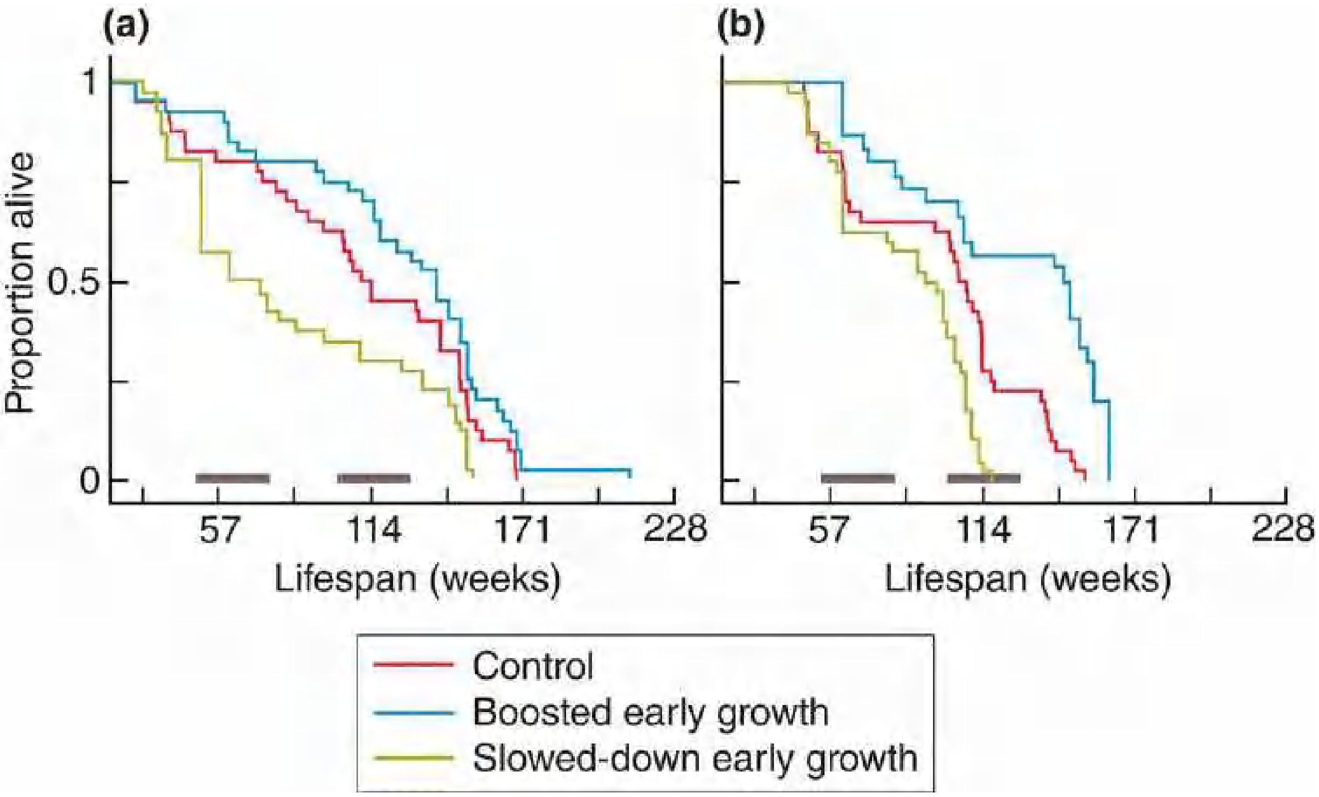


Figure 7.7 **A trade-off between growth and survival in sticklebacks.** Survival curves of three-spined sticklebacks in experiments in (a) winter and (b) spring showing control, boosted early growth and slowed down early growth data. Those with boosted early growth survived better. The solid grey horizontal bars indicate the times of first and second breeding periods, when survival was worse.

Source: After Lee *et al.* (2013).

However, this contrast between the potential advantages of experimental manipulation ('good') as opposed to simple observation ('bad') is not always straightforward. Some manipulations suffer from much the same problems as simple observations. For instance, if clutch size is manipulated by giving supplementary food, then improvements in other traits are to be expected as well. It is important that the manipulation should alter the target trait and nothing else. On the other hand, simple observation may be acceptable if based on the results of a 'natural' experiment. For example, it is likely that as a result of 'mast seeding' (see Section 10.2.4) the population of trees in Figure 7.4a produced large and small crops of seeds in response to factors other than resource availability, and that the negative correlation therefore genuinely represented an underlying trade-off.

genetic comparisons

Ultimately, since life histories can only evolve if they have a genetic basis, the concept of a trade-off is itself a genetic one – the idea that an individual with a gene or genotype favouring an increase in one life history trait must be suffering a reduction in one or more other traits. In practice, as we have already seen, most studies have compared phenotypes with a genetic basis implied. Some, however, have been more explicitly genetic. Genotypes can be compared, first, by a selection experiment, in which a population is subjected to a selection pressure to alter one trait, and associated changes in other traits are then monitored. For example, populations of the Indian meal moth, *Plodia interpunctella*, that had been selected for increased resistance to a virus by infecting them with the virus for a number of generations, exhibited an associated decrease (negative correlation) in their rate of development (Boots & Begon, 1993).

In other cases, a still more explicitly genetic approach has been possible. For example, modern genetic techniques (genome-wide association mapping) have allowed identification of the gene responsible for variation in horn size in the population of Soay sheep, *Ovis aries*, on the island of St Kilda in Scotland – the gene relaxin-like receptor 2 (RXFP2). Male horns are important weapons in the battle for females, and the allele Ho^+ at the RXFP2 locus gives rise to larger horns than the allele Ho^P. As a result, Ho^PHo^P homozygotes have significantly lower reproductive success (fathering offspring) than both Ho^+Ho^+ homozygotes and Ho^+Ho^P heterozygotes. But there is a trade-off. Ho^+Ho^+ homozygotes have significantly lower survival than both Ho^PHo^P homozygotes and Ho^+Ho^P heterozygotes (Johnston *et al*., 2013). As a result it is the heterozygotes that have the highest fitness overall, accounting for the maintenance of both alleles in the population.

However, this simple genetic basis for a trade-off – just one gene locus – is almost certainly unusual. It is far more likely that life history traits are controlled by multiple genes, often referred to as quantitative trait loci (QTL). For example, flower size in the monkeyflower, *Mimulus guttatus*, is under QTL control, and Mojica *et al*. (2012) studied two of these loci in detail using material from Oregon, USA. The results were very similar to those for the Soay sheep. Alleles at both loci that increased flower size unsurprisingly increased fecundity, but they also reduced the plants' viability – again, there was a genetic trade-off. In this case, field trials over three years did not provide clear evidence of heterozygotes having consistently higher fitness overall. But the fittest genotype (the best point on the trade-off curve) varied from year to year and between wet and dry sites only short distances apart. This, too, would lead to the maintenance of a variety of alleles in the population.

Elaborating the details of trade-offs is not easy. In the final analysis, though, it is generally agreed that trade-offs are widespread and important. The problems arise in revealing and hence quantifying them.

7.3.2 The cost of reproduction

Of all the possible life history trade-offs, most attention has been directed at trade-offs that reveal an apparent 'cost of reproduction' (CR). Here, 'cost' is used in a particular way to indicate that an individual, by increasing its current investment in reproduction, is likely to decrease its survival and/or its rate of growth, and therefore decrease its potential for reproduction in future. We saw this for the trees, squirrels and goats in Figure 7.4. Thus, individuals that delay reproduction, or restrain their reproduction to a level less than the maximum, may grow faster, grow larger or have an increased quantity of resources available for maintenance, storage and, ultimately, future reproduction. Any 'cost' incurred by contemporary reproduction, therefore, is likely to contribute to a decrease in residual reproductive value (RRV). Yet, as we have noted, natural selection favours the life history with the highest available *total* reproductive value: the sum of two quantities, one of which (contemporary reproductive output) tends to go up as the other (RRV) goes down. Trade-offs involving the cost of reproduction are at the heart of the evolution of any life history.

7.3.3 The number and fitness of offspring

A second key trade-off is that between the number of offspring and their individual fitness. At its simplest, this is a trade-off between the size and number of offspring, within a given total reproductive investment. That is, a reproductive allocation can be divided into fewer, larger offspring or more, smaller offspring. We can see evidence in support of this for snakes in Australia (Figure 7.8a), trees in Brazil (Figure 7.8b) and from a compilation of data on many species of plants (Figure 7.8c). However, the size of an egg or seed is only an index of its likely fitness. It may be more appropriate to look for a trade-off between the number of offspring and, say, their individual survivorship or developmental rate. It is therefore worth noting that for the seeds in

Figure 7.8c, for example, seed mass was positively correlated with both seedling and subsequent sapling establishment (Moles & Westoby, 2006).

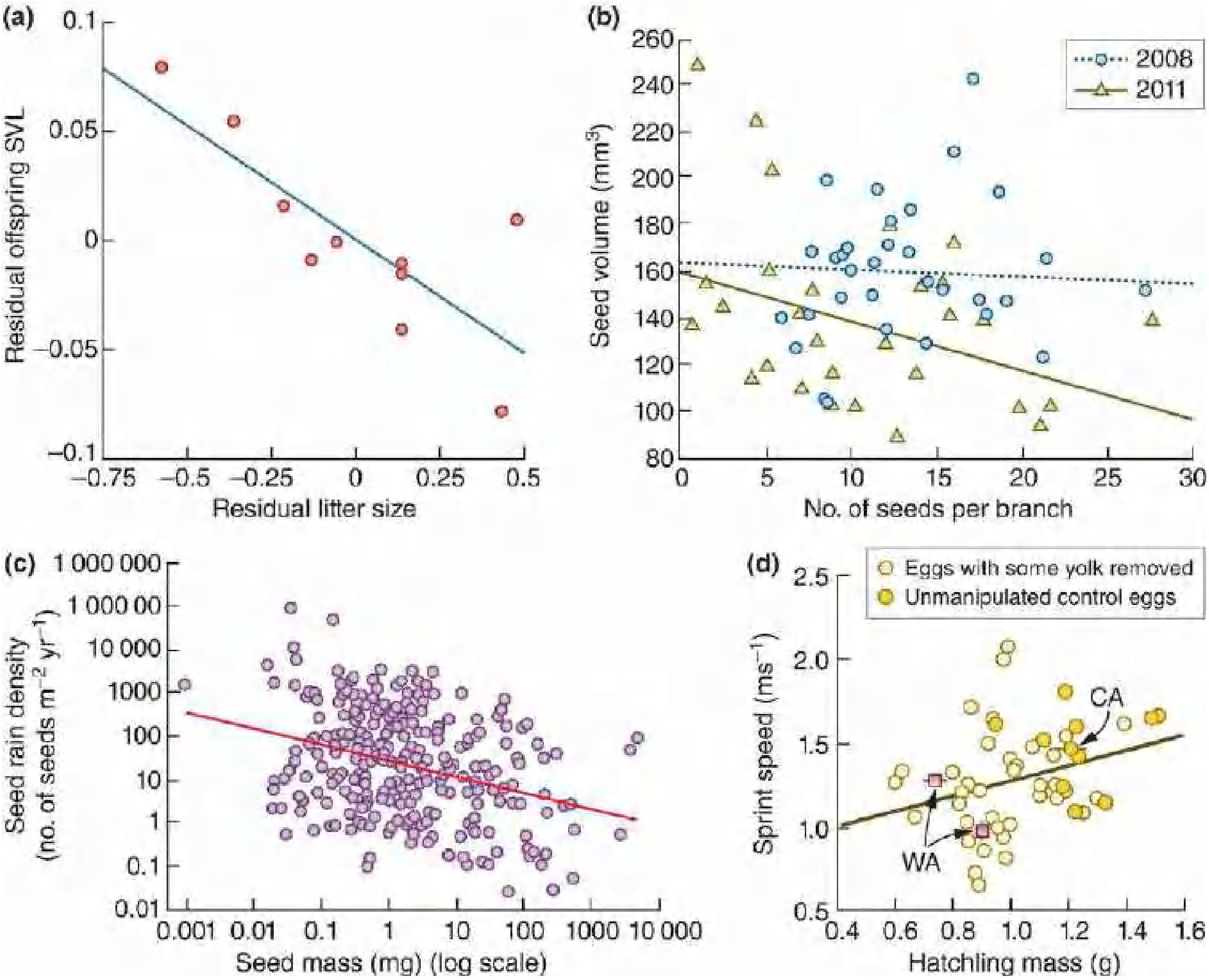


Figure 7.8 Trade-offs between the number of offspring produced in a clutch by a parent and the individual fitness of those offspring. (a) A negative correlation between residual offspring size (measured as snout–vent length, SVL) and the residual number of them in a litter of Australian highland copperhead snakes, *Austrelap ramsayi* ($r^2 = 0.63$, $P = 0.006$) – ‘residual’ because both are values arrived at having taken into account maternal size, with which both increase. (b) Negative correlation between the size and number of seeds produced by the tropical tree, *Copaifera langsdorfii*, in Minas Gerais, Brazil, in years of high (2008) and low (2011) reproductive investment. (Size effect $F = 10.1$, $P < 0.05$; year effect $F = 7.1$, $P < 0.01$; interaction $F = 5.1$, $P < 0.05$.) (c) Negative correlation between seed mass and the number of seeds falling in the seed rain per square metre per year for 303 species of plants ($r^2 = 0.11$, $P < 0.001$.) (d) The mass and sprint speed of Californian (USA) hatchlings of the lizard, *Sceloporus occidentalis*, are lower from eggs with some yolk removed than from unmanipulated control eggs. Also shown are the means for the control Californian hatchlings (CA) (fewer, larger eggs) and those for two samples from Washington, USA (WA) (more, smaller eggs) ($F = 5.2$, $P < 0.03$).

Source: (a) After Rohr (2001). (b) After Souza *et al.* (2015). (c) After Moles & Westoby (2006). (d) After Sinervo (1990).

Once again, however, these trade-offs are by no means universally apparent, and those that are, are often accompanied by caveats. The one for snakes needs to account for variations in maternal size, since both egg size and number increase with this. The one for trees is only really apparent when plants' overall reproductive allocation is relatively limited. And for the one aggregating many species of plants, it is certain that the various species were not allocating the same amount of resource, contrary to a strict interpretation of the trade-off concept.

Moreover, this type of trade-off is especially difficult to observe through experimental manipulation. To see why, note that we need to ask the following type of question. Given that a plant, say, produces 100 seeds each weighing 10 mg and each with a 5% chance of developing to reproductive maturity, what would be the seed size, and what would be the chance of developing to maturity, if an identical plant receiving identical resources produced only 80 seeds? Clearly, it would be invalid to manipulate seed number by altering the provision of resources; and even if 20 seeds were removed at or close to their point of production, the plant would be limited in its ability to alter the size of the remaining seeds, and so their subsequent survivorship would not really address the question posed. Sinervo (1990), though, did manipulate the size of the eggs of an iguanid lizard (*Sceloporus occidentalis*) by removing yolk from them after they have been produced, giving rise to healthy but smaller offspring than unmanipulated eggs. These smaller hatchlings had a slower sprint speed (Figure 7.8d) – probably an indication of a reduced ability to avoid predators – and hence a lower fitness. Crucially, within natural populations, this species produces smaller clutches of larger eggs in California than in Washington (Figure 7.8d). Thus, in the light of the experimental manipulations, the comparison between the two natural populations does indeed appear to reflect a trade-off between the number of offspring and their individual fitness.

APPLICATION 7.1 Grain size and number in wheat

The trade-off between offspring size and number – in this case, seed or grain size and number – is of critical importance in the breeding and production of crops. For all grain crops, the natural measure of yield is the product of grain size and number. It is therefore important to understand the underlying trade-off – how changes in one, brought about either by growing conditions or breeding programmes, will impact on the size or number of the other. In plants generally, including many crops, there is greater variation in seed number than in seed weight in response to variations in growing conditions (Harper, 1977). This, though, is not an invariable rule. Crops like wheat and soybean certainly respond to improved growing conditions mostly by increasing the number of seeds per plant, not their size. But in other cases, for example sunflowers and maize, domestication has led to plants with only one or very few inflorescences per plant, limiting the opportunities for seed number variation, and here responses are much more evenly spread between number and size (Figure 7.9).

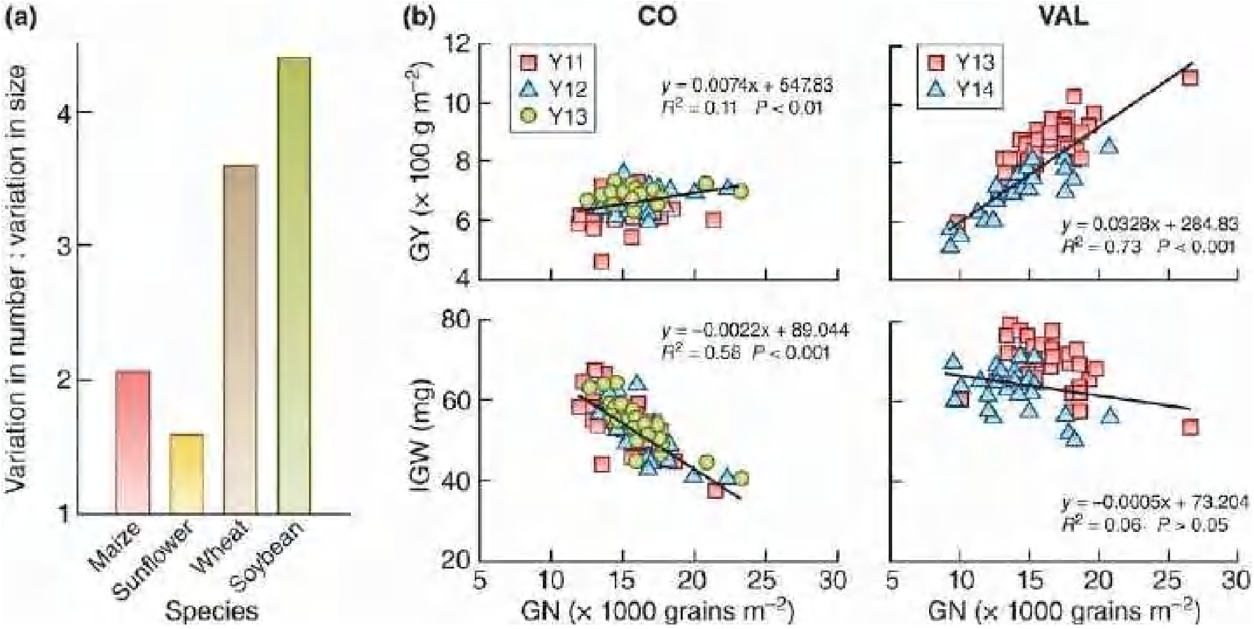


Figure 7.9 Crop plant species differ in the relative responses of seed number and seed weight to changes in growing conditions, and even within species the seed weight–seed number trade-off can vary. (a) The contrast between the median ratio of number- to size-variation between strains in seeds of species that have been bred to reduce inflorescence number – maize and sunflower – and species where this has not been the case: for example, wheat and soybean. (b) The relationship between overall grain yield (GY) and grain number (GN) (top two panels), and between individual grain weight (IGW) and GN (bottom two panels) for 27 wheat genotypes grown in different years, as indicated (Y11–Y14), at two sites: a less favourable site in Mexico (CO) and a more favourable site in Chile (VAL). Statistical relationships are shown in each panel.

Source: (a) After Sadras (2007). (b) After Quintero *et al.* (2018).

This pattern of contrasting trade-offs itself prompts an interesting question: since seed size is more genetically fixed, and seed number more responsive to growing conditions, should breeding programmes focus on seed size so as to realign this fixed point around which seed number can vary? Or should the focus be on seed number because there is little genetic variation in seed size on which breeding programmes can act? This in turn raises questions about the underlying genetic architecture behind any trade-off – whether, for example, alleles that lead to an increase in seed weight, and could be selected for, necessarily lead to a decrease in seed number.

We can see some of the challenges in addressing these questions by examining studies on varieties of wheat (*Triticum aestivum* and *T. turgidum* var. *durum*) held by the International Center for the Improvement of Maize and Wheat (CIMMYT) in Mexico. Firstly, 27 'elite' wheat genotypes were grown both at sites in Chile and in Mexico. Growing conditions, especially the amount of solar radiation, were better in Chile than Mexico and overall yields were higher there (Figure 7.9b). But whereas in Mexico, a clear trade-off between grain size and number was apparent and overall yield was only weakly related to grain number, in Chile the trade-off was totally absent and yield increased steeply with grain number (Figure 7.9b). This suggests that breeding strategies designed to improve yield may themselves have to be tailored to the growing location. Attempts to increase grain number can be proposed where conditions are favourable, as in Chile, and such increases seem likely to be translated into a rise in overall yield. By contrast, where conditions are less favourable, as in Mexico, breeding to increase grain weight may be proposed, since there is much less chance of influencing yield through changes in grain number.

More generally, it is well known that in wheat breeding, direct selection for increased yield mostly gives rise to varieties with increased grain number per unit area, because, as we've seen, this is the more variable of the two yield components. This, though, makes it all the more important to understand the trade-off – the consequences for grain size of selecting for any allele that increases grain number, and ideally to identify alleles that, for example, increase grain weight with little or no associated decrease in grain number. The genetic architecture underlying the grain size–number trade-off in wheat has been examined by combining two of the varieties held at CIMMYT with similar yields overall (Griffiths *et al.*, 2015). One was chosen for its high grain weight and the other for its high grain number, and their recombinant offspring were grown at a variety of sites in Chile, Mexico, Argentina and the UK. Encouragingly, a QTL (see previously) was identified in the high grain-weight variety that did indeed have little or no effect on grain number. Clearly, this makes it an attractive target for further attempts to increase wheat yield.

7.4 Life histories and habitats

7.4.1 Options sets and fitness contours

options sets

Having recognised the ubiquity, and inevitability, of life history trade-offs, one of the most fundamental questions in life history evolution immediately presents itself: what point on a trade-off curve will be favoured by natural selection, and under what circumstances? To address this question, we first introduce the concept of an *options set*. We pursue this initially in the context of the cost of reproduction, and hence the trade-off between present reproduction and the various contributors to residual reproductive value, RRV, but the same principles apply to any trade-off. Here, we focus on a trade-off between present reproduction and present growth, as a contributor to RRV (Figure 7.10). The options sets in the figure reflect the organism's underlying physiology. Hence, they describe the whole range of combinations of levels of present reproduction on the one hand, and growth on the other, that the organism can achieve. The outer boundary of the options set is a trade-off curve and is therefore the only part of the options set that really concerns us, since options below this, 'inside' the options set, can always be bettered by nearby combinations on the boundary.

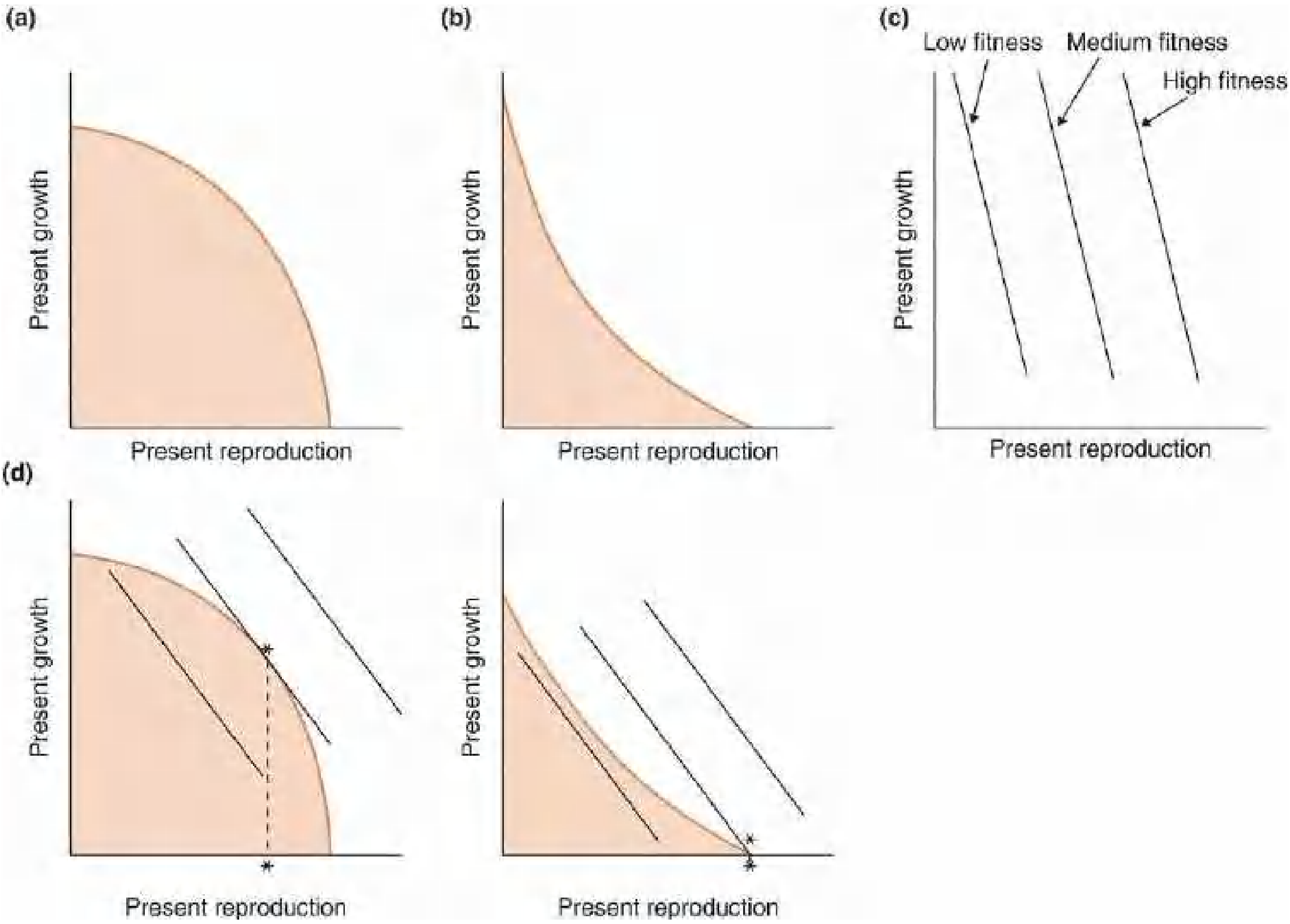


Figure 7.10 Options sets and fitness contours together determining optimal life histories. (a, b) Options sets – the combinations available to an organism of, in this case, present reproduction and present growth. As explained in the text, the outer boundary of the options set is a trade-off curve: (a) is convex-outwards, and (b) concave-outwards. (c) Fitness contours linking combinations of present reproduction and present growth that have equal fitness in a given habitat. Hence, contours further from the origin have greater fitness. (d) The point in an options set with the greatest fitness is the one that reaches the highest fitness contour. This point, and the (optimal) value for present reproduction giving rise to it, are marked with an asterisk. These differ depending on the shape of the trade-off curve.

Source: After Sibly & Calow (1983).

An options set may be convex-outwards (Figure 7.10a), implying in the present case that a level of present reproduction only slightly less than the maximum allows a substantial amount of growth. Alternatively, the set may be concave-outwards (Figure 7.10b), implying that substantial growth can only be achieved with a level of present reproduction considerably less than the maximum.

fitness contours

To address the question of which point on a trade-off curve should be favoured by natural selection, we now combine options sets with *fitness contours*. These join combinations of present reproduction and (in this case) present growth for which overall fitness is constant (Figure 7.10c). Contours further away from the origin therefore represent combinations with greater fitness. Natural selection favours the point in the options set that reaches the highest fitness (touches the highest fitness contour, indicated by the asterisks in Figure 7.10d), which is clearly on its edge, the trade-off curve. The shapes of fitness contours reflect not the organism's underlying physiology but its interaction with the habitat in which it lives, as elaborated later. Thus, options sets and fitness contours can be used together as a guide to where and when different types of life history might be found.

7.4.2 High and low CR habitats: a *comparative* classification

Each organism's interaction with its habitat is unique, but fitness contours allow the effects of habitats on organisms to be classified simply in ways that apply to all organism–habitat interactions. For established individuals (i.e. not newly born offspring) it is useful to draw a distinction between habitats that extract a high cost of reproduction (high CR habitats) and those in which there is only a small price to be paid (low CR habitats). In high CR habitats, present reproduction has a significant negative effect on growth (and/or survival and/or future reproduction), and hence on RRV. Thus, similar fitness can be achieved by combining high reproduction with low RRV or low reproduction with high RRV, and fitness contours therefore run diagonally with a negative slope (Figure 7.11a). By contrast, in low CR habitats, RRV is little affected by the level of present reproduction. Thus, fitness is largely determined by the level of present reproduction alone, and the fitness contours therefore run approximately vertically (Figure 7.11a). This classification is comparative. An organism's interaction with its habitat can only be described as 'high CR' relative to some other interaction that is, comparatively, low CR. The purpose of the classification is to contrast habitats with one another.

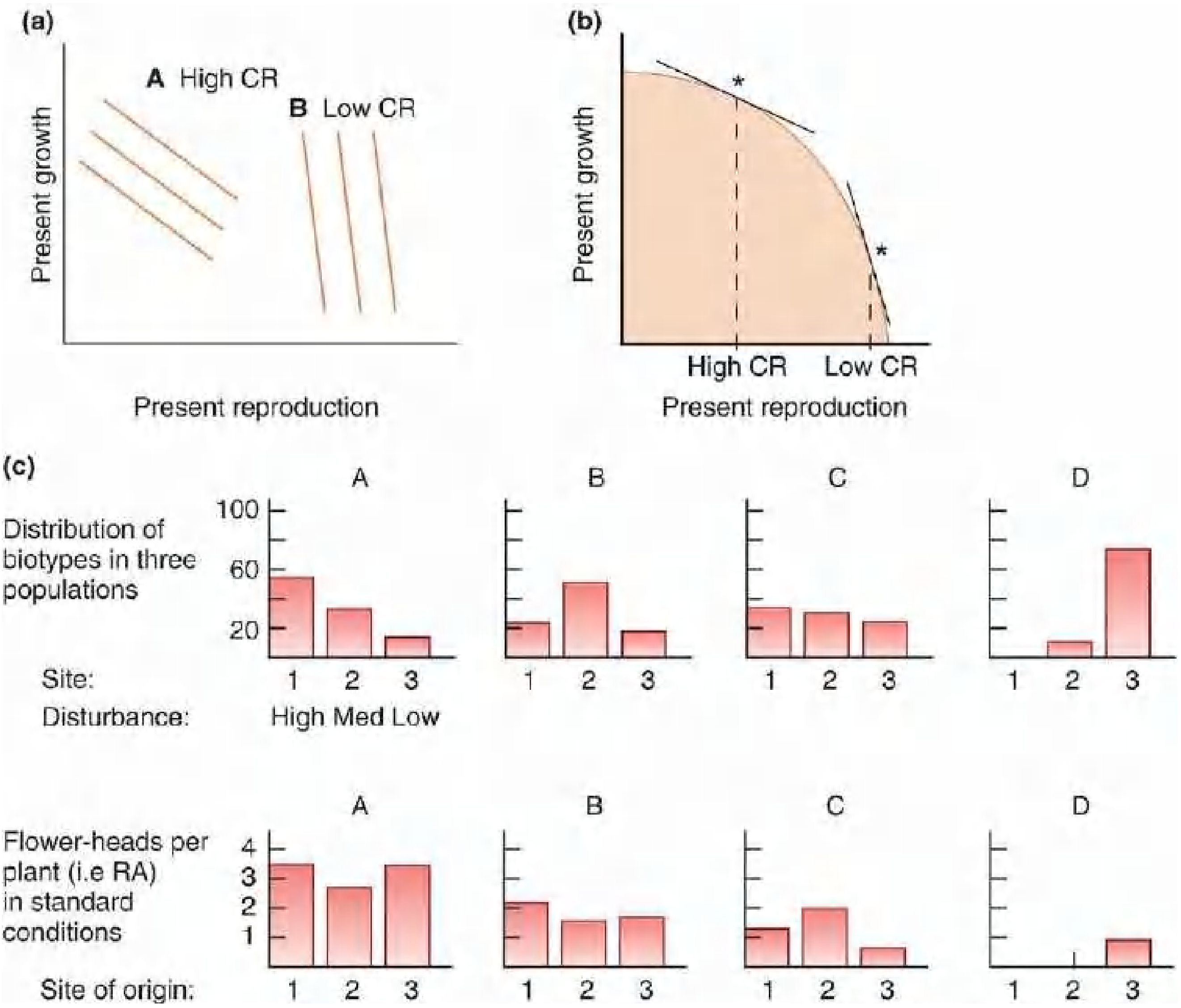


Figure 7.11 Optimal life histories in high and low cost of reproduction habitats. (a) Habitats of established individuals can be either: (A) relatively high cost of reproduction (CR) (fitness contours reflect significant contributions to overall fitness from both present reproduction and increased growth) or (B) relatively low CR (fitness contours largely reflect the level of present reproduction). (b) Options sets and fitness contours combined suggest that relatively high CR habitats should favour relatively small reproductive allocations, whereas low CR habitats should favour large reproductive allocations or even semelparity (maximum present reproduction). (c) Upper panel: the distribution of four biotypes (A–D) of the dandelion *Taraxacum officinale*, amongst three populations subject to high, medium and low levels of disturbance (i.e. habitats ranging from relatively low CR to relatively high CR). Lower panel: the reproductive allocations (RA) of the different biotypes from the different sites of origin.

Source: After Solbrig & Simpson (1974).

high or low CR for a variety of reasons

Habitats can be relatively high CR for an organism for at least two reasons. Firstly, there may be intense competition amongst established individuals with only the best competitors surviving and reproducing, such that present reproduction may be costly because it reduces growth and hence substantially reduces competitive ability in the future, and thus reduces RRV. Species like red deer, where only the biggest males get to mate, are a good example. Alternatively, habitats can be relatively high CR if smaller adults are more susceptible to an important source of mortality from a predator or some abiotic factor. Thus, present reproduction may be costly because it maintains adults in these vulnerable size classes.

Similarly, habitats can be relatively low CR for a number of reasons. Firstly, much mortality may be indiscriminate and unavoidable, so that any increase in size caused by reproductive restraint is likely to be worthless in future. For instance, when temporary ponds dry out, most individuals die irrespective of their size or condition. Second, the habitat may be so benign and competition-free for established individuals that all of them have a high probability of surviving and a large future reproductive output, irrespective of present reproductive activity. This is true, at least temporarily, for the first colonists to arrive in a newly arisen habitat. And third, a habitat may be low CR simply because there are important sources of mortality to which the largest individuals are especially prone, such that restrained present reproduction, by leading to greater size, may give rise to *lowered* survival in future.

7.4.3 Reproductive investment and its timing

We can now focus on overall levels of investment in reproduction, and the timing of that investment. Here, as with other aspects of life history, the same option (for example a large reproductive allocation) can be favoured by a variety of effects of habitat on the organism concerned. We examine examples of these rather than review them exhaustively.

overall reproductive investment

If we assume initially that all options sets are convex-outwards, then relatively low CR habitats should favour a higher reproductive investment, while relatively high CR habitats favour a lower reproductive investment (Figure 7.11b). This pattern can be seen, for example, in three populations of the dandelion *Taraxacum officinale* (Figure 7.11c) The populations were composed of a number of distinct clones that belonged to four biotypes (A–D). The habitats of the populations varied from a footpath (the habitat in which adult mortality was most indiscriminate – 'lowest CR') to an old, stable pasture (the habitat with most adult competition – 'highest CR'); the third site was intermediate between the other two. In line with predictions, the biotype that predominated in the footpath site (A) made the greatest reproductive allocation (whichever site it was obtained from), while the biotype that predominated in the old pasture (D) made the lowest reproductive allocation. Biotypes B and C were appropriately intermediate with respect both to their site occupancies and their reproductive allocations.

lifetime profiles of reproductive allocation

Inevitably, over the course of an organism's life, the price it pays for reproduction changes, irrespective of any external, environmental alterations – as it grows larger, as it gets closer to the expected end of its life, and so on. This in turn can generate predictable patterns in its lifetime profile of reproductive investment. One particularly influential prediction has been the 'terminal investment' hypothesis, going back to Williams (1966). This proposes that as organisms approach the end of their life, their residual reproductive value inevitably declines, their cost of reproduction therefore also declines (since it must be paid in a currency of reduced future reproduction, for which there is little time), their fitness contours consequently become steeper, and their reproductive investment should therefore increase towards the end of life (Figure 7.12a). On the other hand, it might equally be expected that as organisms age, they senesce, their options sets contract (they can no longer do what they used to be able to do), and their reproductive allocation (in absolute terms) will therefore *de*crease (Figure 7.12a). It is not surprising, then, that studies of lifetime profiles, including tests of either or both of these hypotheses, have produced varied results.

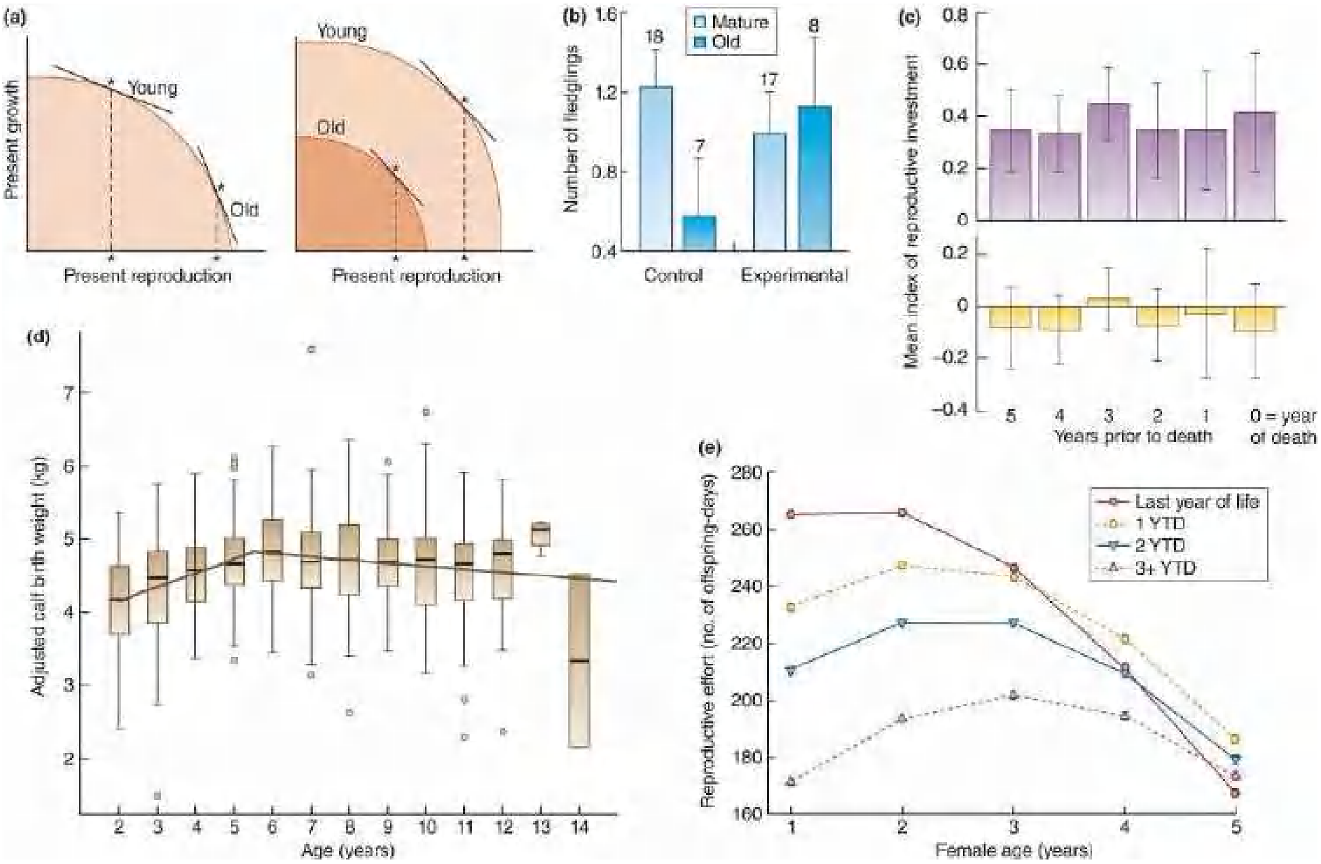


Figure 7.12 Proposed and observed patterns in reproductive allocation as individuals age. (a) Alternative views of how reproductive allocation may change as organisms age. Left: the terminal investment hypothesis. Towards the end of life, the cost of reproduction is small (near-vertical fitness contours) and a large investment in reproduction is favoured. Right: senescence. Towards the end of life, only a small options set is available and so only a small investment in reproduction is possible. (b) Senescence and terminal investment in male blue-footed boobies, *Sula nebouxii*, in Mexico. The number of fledglings (means ± SEs) reared was normally greater for young than for mature males, suggesting senescence. But when birds were given an infection that reduced their life expectancy, fledgling success went up for old males, suggesting terminal investment (treatment, $F_{1,44} = 2.95$, $P = 0.09$; age, $F_{1,44} = 0.25$, $P = 0.62$; treatment × age, $F_{1,44} = 5.35$, $P = 0.025$). (c) No evidence of senescence or terminal investment in Californian oaks. Above, the mean index of reproductive investment (± 95% CI) in trees in their last six years prior to death (five years before to year of death (0)); and below, the difference (± 95% CI) between this index and the mean for these trees for the years between 10 and six years before death. The index was the acorn crop divided by the sum of acorn crop and annual growth, both standardised so that different species could be combined. (d) Senescence in reindeer, *Rangifer tarandus*. Reproductive effort (calf birth weight adjusted for calf sex, etc.) declined from age 5.5 years to the end of life. Boxes are 25–75% quartiles. The line shows the best fit (two-stage) model. (e) Senescence and terminal investment in song sparrows, *Melospiza melodia*. Reproductive effort (number of offspring × number of days spent caring for them) declined in older animals, but in younger animals was greater in birds that were closer to death (YTD, years to death). Lines (± SEs) are the output of a statistical model fitted to the data.

Source: (b) After Velando *et al.* (2006). (c) After Koenig *et al.* (2017). (d) After Weladji *et al.* (2010). (e) After Tarwater & Arcese (2017).

Figure 7.12b shows a good example in which the terminal investment hypothesis is apparently supported. Older individuals of a long-lived bird in Mexico, the blue-footed booby, *Sula nebouxii*, normally made a smaller reproductive allocation (suggesting a senescence effect), but when they were given an experimental infection, likely to shorten their life, the older birds, seemingly staring death in the face, significantly increased their reproductive allocation (made a large terminal investment), whereas younger birds, with a greater residual reproductive value to protect, actually reduced their allocation marginally. On the other hand, in an observational study of eight species of Californian oaks, *Quercus* spp., there was no evidence of any increased reproductive investment (as a proportion of available resource) even as they underwent physical decline in their final years of life (Figure 7.12c). And in a study of reproductive output in female reindeer, *Rangifera tarandus*, in Finland, several aspects of reproduction, including calf birth mass (Figure 7.12d), suggested a senescence effect rather than an increased terminal investment.

But on the other hand again, in song sparrows, *Melospiza melodia*, from Mandarte Island, Canada, there was an apparent senescence effect – females generally made a reduced reproductive allocation as they aged – but also, younger females made a greater allocation the closer they were to death (Figure 7.12e), and this effect was itself greater the younger the females were. Taking this pattern at face value, it suggests support for the terminal investment hypothesis, and certainly there was a terminal *allocation* to reproduction. But as with any observational study, it is difficult to separate cause and effect. It may have been that increased investment in younger females shortened their lives, rather than their having adjusted investment in anticipation of death. Such ambiguity is difficult to avoid. Senescence and terminal investment effects are not mutually exclusive. It is always going to be difficult to tease them apart.

age at maturity

One key aspect of an organism's lifetime profile of reproductive allocation is when it first initiates reproduction: its age of maturity. Since relatively high CR habitats should favour low reproductive investments, maturity should be relatively delayed in such habitats, but should occur at a relatively large size following a more extended period of growth made possible by reproductive restraint.

These ideas are supported by a classic study of guppies, *Poecilia reticulata*, a small fish species in Trinidad that also provides support for the patterns of overall reproductive allocation discussed above. The guppies live in small streams that can be divided into two contrasting types. In one, their main predator is a cichlid fish, *Crenicichla alta*, which exerts high predation rates that are directed at all size-classes of guppies. This leads to a high predation risk overall, low densities, and good availability of resources for the guppies. The cichlid sites are therefore low CR for the guppies, since the reduced growth that would accompany reproduction is harmful neither to competitive ability (since competition itself is unimportant) nor to predation risk (since the cichlids attack all size-classes). In the other type of stream, the main predator is a killifish, *Rivulus hartii*, which prefers small, juvenile guppies and exerts lower predation pressure, giving rise to more crowded populations. The killifish sites are therefore high CR for the guppies, since the reduced growth that would accompany reproduction harms both competitive ability, which is important, and predation risk, which is highest in smaller fish. As predicted, the guppies from low CR, cichlid sites mature earlier and at a smaller size and make a larger reproductive allocation (Reznick, 1982).

These observations were also supported by a field experiment. When 200 guppies were introduced from a cichlid to a killifish site, and lived there for 11 years (30 – 60 generations), not only did the phenotypes in the field come to resemble those of other high CR (killifish) sites, but these differences were also discernible under laboratory conditions, indicating that they were heritable and had evolved (Reznick *et al*., 1990).

Interestingly, the killifish themselves are subject to contrasting patterns of predation and crowding at different sites in Trinidad, and when reared in a common environment, those from high predation–low crowding sites again matured earlier, at a smaller size, and made a greater reproductive allocation (Figure 7.13), though as we have seen before, this pattern was itself influenced by the immediate environment as well as evolutionary history. The pattern was far more marked when the fish developed at high food levels than at low – an effect the authors of the study ascribe to the killifish from high predation–low crowding sites being better adapted to the high food levels they would have experienced there (Walsh & Reznick, 2008).

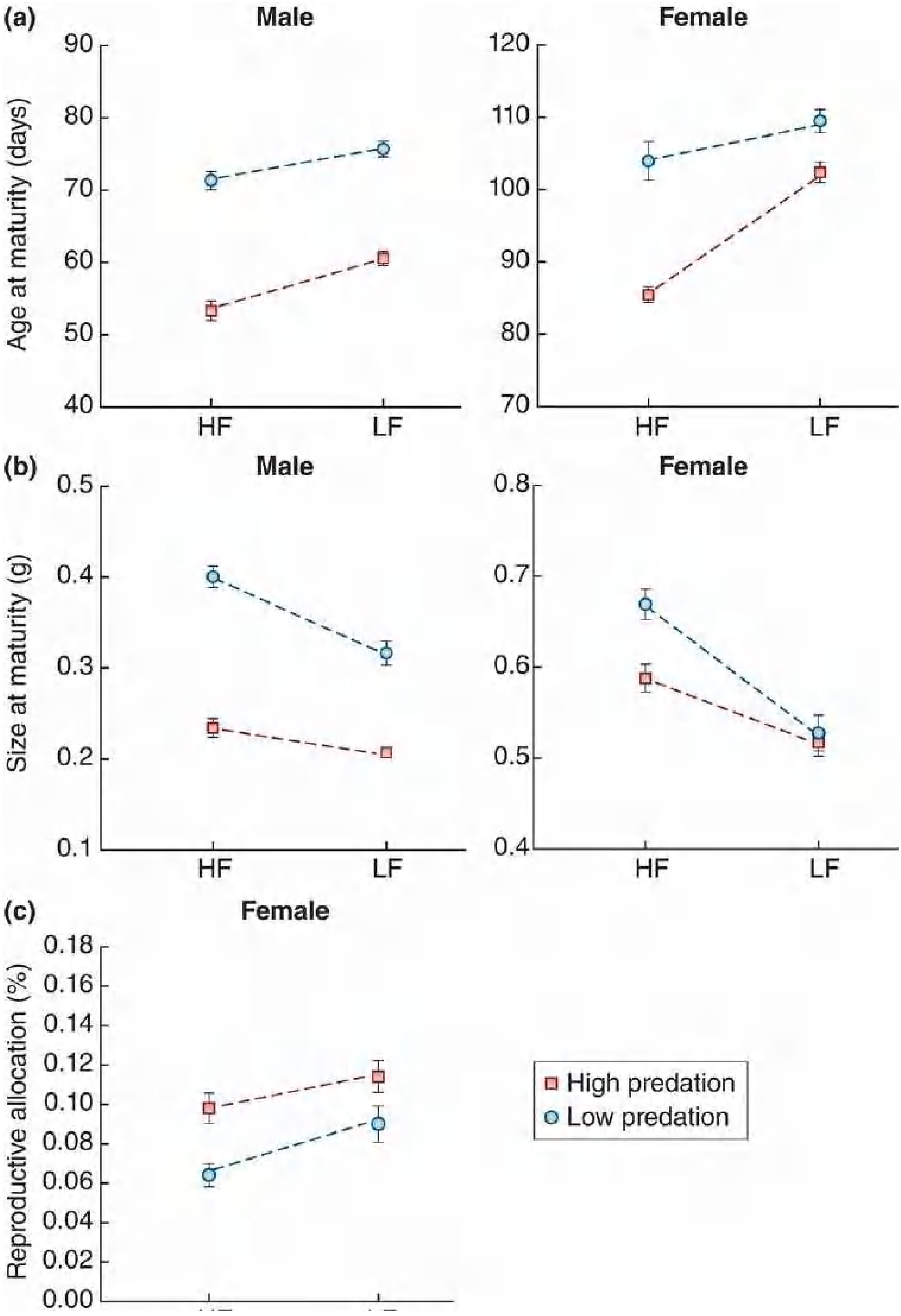

(a)
Male
Female
Age at maturity (days)
90
80
70
60
50
40
120
110
100
90
80
70
HF
LF
HF
LF
(b)
Male
Female
Size at maturity (g)
0.5
0.4
0.3
0.2
0.1
0.8
0.7
0.6
0.5
0.4
HF
LF
HF
LF
(c)
Female
Reproductive allocation (%)
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
High predation
Low predation

Figure 7.13 Effects of predation risk on patterns of reproduction in killifish. When Trinidadian killifish, *Rivulus hartii*, from low predation sites (where *Rivulus* was the only piscivorous fish) and high predation sites (containing other highly piscivorous fish) were reared in a common environment, they differed in (a) age and (b) size at maturity in males and females, and (c) reproductive allocation in females, at high and low food levels (HF and LF). Fish in low predation sites (more crowded, higher cost of reproduction) matured later and larger and made a smaller reproductive allocation ($P < 0.05$ in all cases), especially at high food levels (interaction term also significant). Bars are SEs.

Source: After Walsh & Reznick (2008).

semelparity

Semelparity is arguably the most extreme lifetime profile of reproductive investment: a switch in reproductive allocation from zero to maximum, with a cost of reproduction so large that no further reproduction is possible. In a classic analysis of models of population growth, Cole (1954) noted that a semelparous individual able to produce just one more offspring than its iteroparous equivalent would generate the same population growth rate, and asked why, in that case, semelparity was not more common than it is. Further analysis resolving 'Cole's paradox' (Young, 1981) emphasised especially the importance of juvenile survival rates (before reproduction) and adult survival rates following reproduction, which Cole had assumed were the same. The analysis showed that semelparity would increasingly be favoured by higher rates of juvenile survivorship and lower rates of adult survivorship. For example, if juvenile rates were very low, a semelparous individual would have to produce very many more offspring to match its iteroparous equivalent.

We can see essentially the same prediction within the framework of options sets and fitness contours if we relax our assumption that all options sets are convex-outwards. It is then apparent that semelparity is especially likely to evolve in organisms with options sets that are *concave*-outwards. In these, even a low level of present reproduction leads to a considerable drop in, say, future survival, but further increases in reproduction have increasingly little influence on survival (see Figure 7.10d). In other words, adult (postreproductive) survival is low, irrespective of the reproductive allocation. This is the likely explanation for the suicidal semelparity displayed by many species of salmon. Reproduction for them demands a dangerous and effortful upstream migration from the sea to their spawning grounds, but the risks and extra costs associated with the 'act' of making a reproductive allocation are largely independent of its magnitude.

If we revert to convex-outwards options sets, semelparity is most likely to evolve where the highest reproductive allocations are favoured, that is, in low CR habitats (Figure 7.11b). There is support for this in experiments on the plant, *Lobelia inflata*, in Canada, which is semelparous in the sense that plants always senesce at the end of a single period of flowering. However, it is possible to vary the initiation of flowering from mid-June to mid-September by manipulating the times at which seeds are germinated and seedlings planted out, following which flowering finishes predictably around mid-October, with the first frosts. Plants that start flowering later have a shorter period of potential reproduction ahead of them and thus a lower cost of reproduction. As predicted, they were 'more' semelparous: they started to reproduce sooner and at a smaller size, but also squeezed their reproduction into a more intense burst, even though their total reproductive output was similar to plants that started earlier (Figure 7.14). There appeared, therefore, to be 'degrees' of semelparity, or, put another way, a continuum from semelparity to iteroparity. As so often, nature is more nuanced than our simple distinctions would imply.

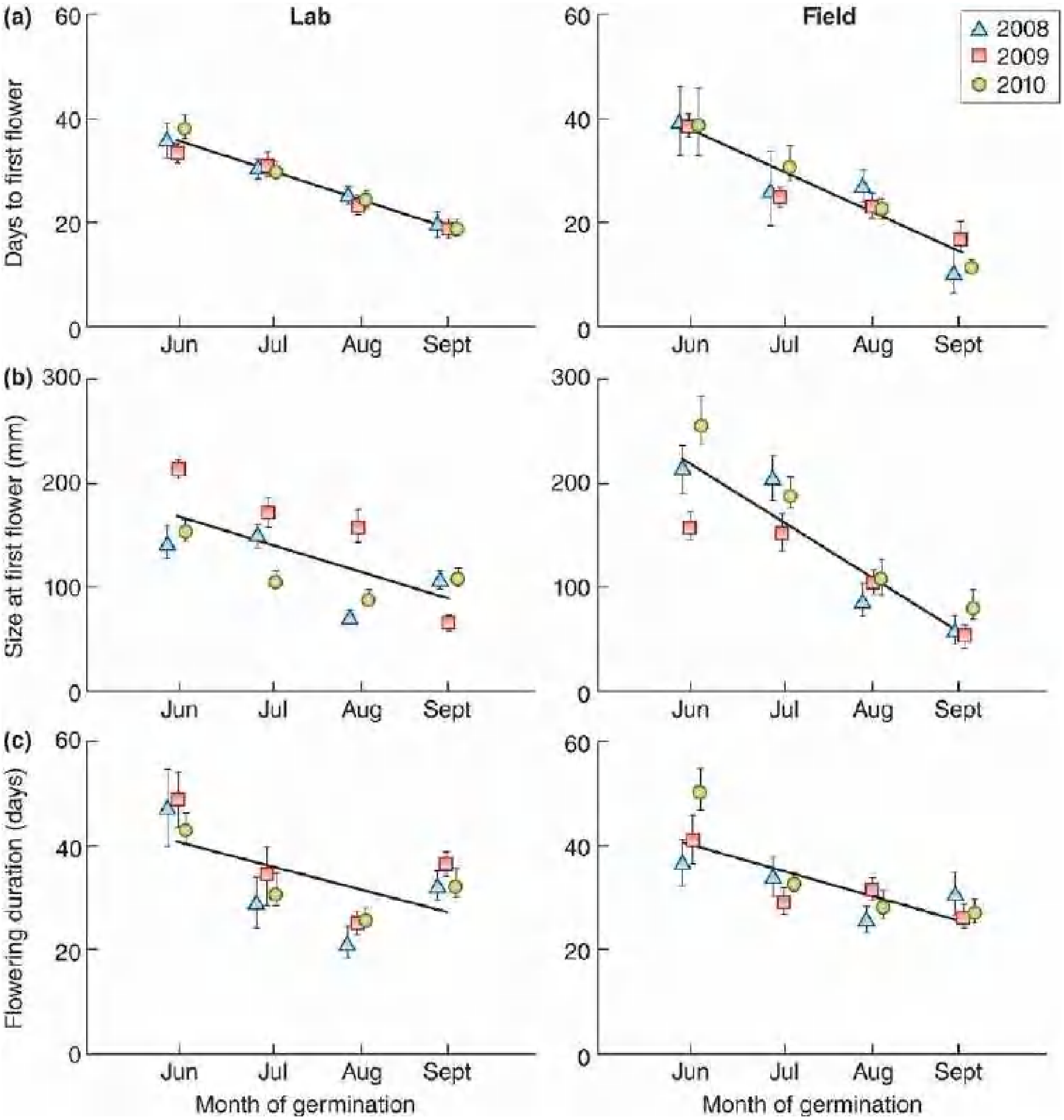


Figure 7.14 The 'degree' of semelparity in the plant *Lobelia inflata* depends on whether it germinates earlier or later (June to September). Observations are shown for both laboratory and field situations. Plants that started later were 'more' semelparous in that they (a) flowered earlier ($P < 0.05$), (b) flowered when smaller ($P < 0.05$) and (c) had a shorter reproductive period ($P < 0.05$) even though the overall reproductive output remained similar. Data are shown from three years; bars are 95% CIs.

Source: After Hughes & Simons (2014).

7.5 The size and number of offspring

a classification of habitats for newly born offspring

When we come to classify habitats from the point of view of newly born offspring, a similar approach can be taken to that in Figure 7.10. We start with an options set encompassing all possible combinations of offspring size and the numbers of them produced in a particular reproductive episode (the litter or clutch size), and link these, again, to fitness contours. In 'offspring size-insensitive' habitats (indiscriminate mortality, or superabundant resources, or sources of mortality to which larger individuals are more prone), the reproductive value of individual offspring is little affected by their size and so fitness contours run near-parallel to the offspring-size axis. We therefore expect a given reproductive allocation to be divided into a larger number of smaller offspring (Figure 7.15). By contrast, in 'offspring size-sensitive' habitats, the reproductive value of individual offspring rises significantly with size, perhaps because of competition early in life, or important sources of mortality to which small offspring are especially vulnerable. Fitness contours therefore run diagonally, not near-vertically, and we expect a given reproductive allocation to be divided into a smaller number of larger offspring (Figure 7.15).

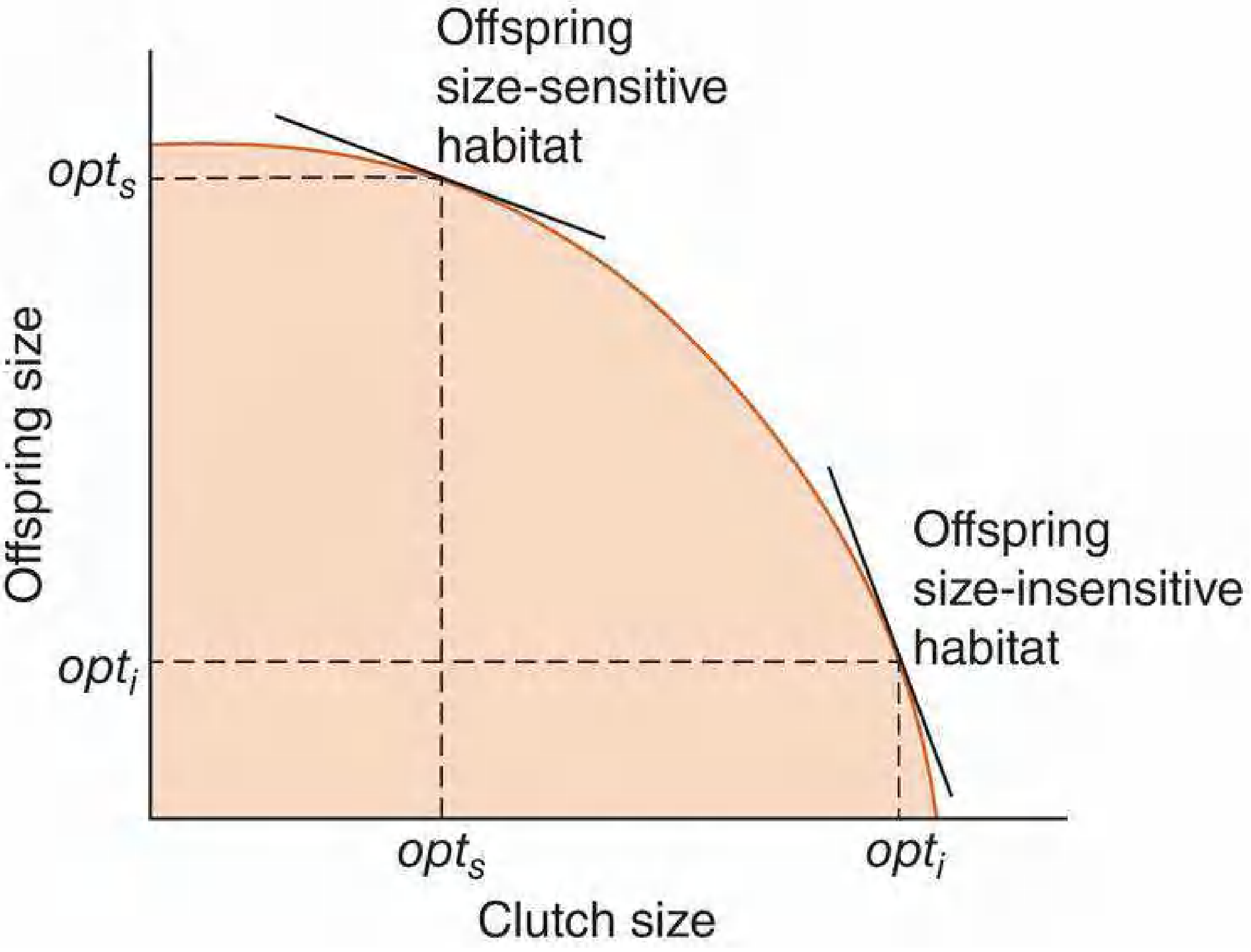


Figure 7.15 The optimisation of offspring size and number in a clutch or litter. Fitness contours run more steeply in offspring size-insensitive than offspring size-sensitive habitats, intersecting the options set at a larger clutch of smaller offspring (opt_i compared with opt_s).

Evidence of this is again provided by the Trinidadian guppies we discussed in the previous section. Offspring size is larger in the killifish sites, in which, as we have seen, predation is less intense and concentrated on smaller individuals (exerting pressure to move quickly beyond the most vulnerable stages) and competition is more intense (Reznick, 1982). These correlations provide a plausible explanation for the size differences, but in the case of competition, at least, there is also more direct, experimental evidence. Groups of small and large newborn guppies (around 7.2 and 7.7 mm in length, respectively) were introduced to artificial tanks already containing a population of guppies designed to mimic natural size distributions. This was done using newborns from both low predation–high competition sites and high predation–low competition sites and carried out at both high and low densities (56 and 14 guppies per tank – again mimicking natural variations). Results in the high-density tanks were especially informative. Certainly, there were indications of competition amongst the newborns there: the faster the growth of individuals in the resident population, the slower the growth of the newborns (Figure 7.16a). And it was also clearly the case that the smallest newborns were affected most by this, especially if they were from the low predation–high competition site (Figure 7.16b, c). Producing larger newborns (and hence fewer of them) is clearly adaptive in the more highly competitive environment.

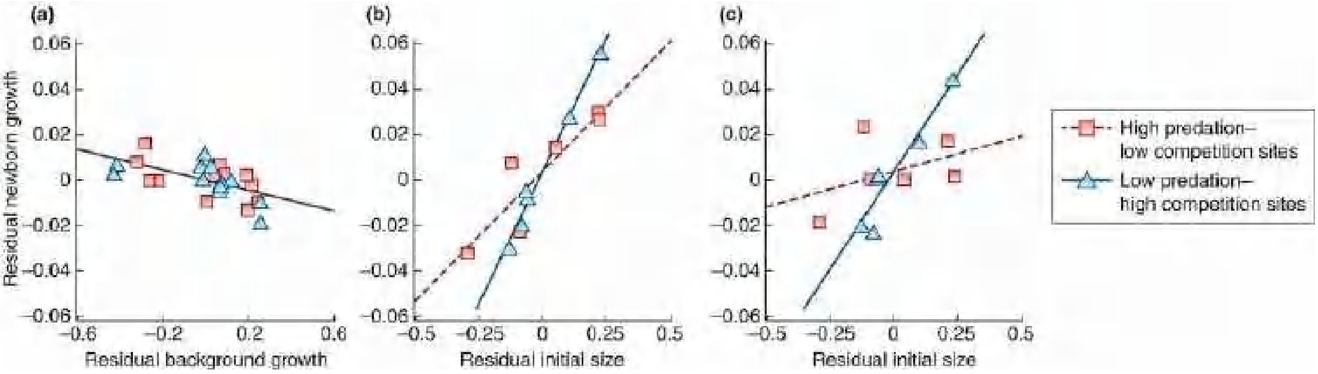


Figure 7.16 Guppies produce fewer, larger offspring in a more highly competitive environment. Newborn guppies, *Poecilia reticulata*, from low predation–high competition sites and high predation–low competition sites were added to populations at high density, constructed with a natural size distribution. Data shown are 'residuals' from statistical models in order to take account of other factors acting on the variables in each case. (a) Irrespective of the site of origin, newborns grew more slowly in populations where the residents were growing faster ($F_{1,5} = 7.39$, $P < 0.05$). (b) Amongst smaller newborns, those that were initially largest grew fastest, especially when from the low predation site. (c) Similar results for the larger newborns (size class $F_{1,5} = 22.56$, $P < 0.01$; initial size $F_{1,5} = 117.39$, $P < 0.001$; initial size × population $F_{1,5} = 41.05$, $P < 0.01$).

Source: After Bashey (2008).

7.5.1 The number of offspring: clutch size

the Lack clutch size

The underlying idea of an optimum clutch size goes back at least as far as Lack (1947), who emphasised that natural selection will favour not the largest clutch size but a compromise clutch size, which, by balancing the number produced against their subsequent survival, leads to the maximum number surviving to maturity (Figure 7.17a). Lack was especially interested in birds, and was thinking mostly of the difficulties parents have in rearing large numbers of nestlings when available resources are spread too thinly. This optimum clutch size has therefore come to be known as the 'Lack clutch size'. As we saw in Figure 7.15, what that optimum turns out to be depends on the habitat (the angle of the fitness contours), the level of resources available (the size of the options set), and so on.

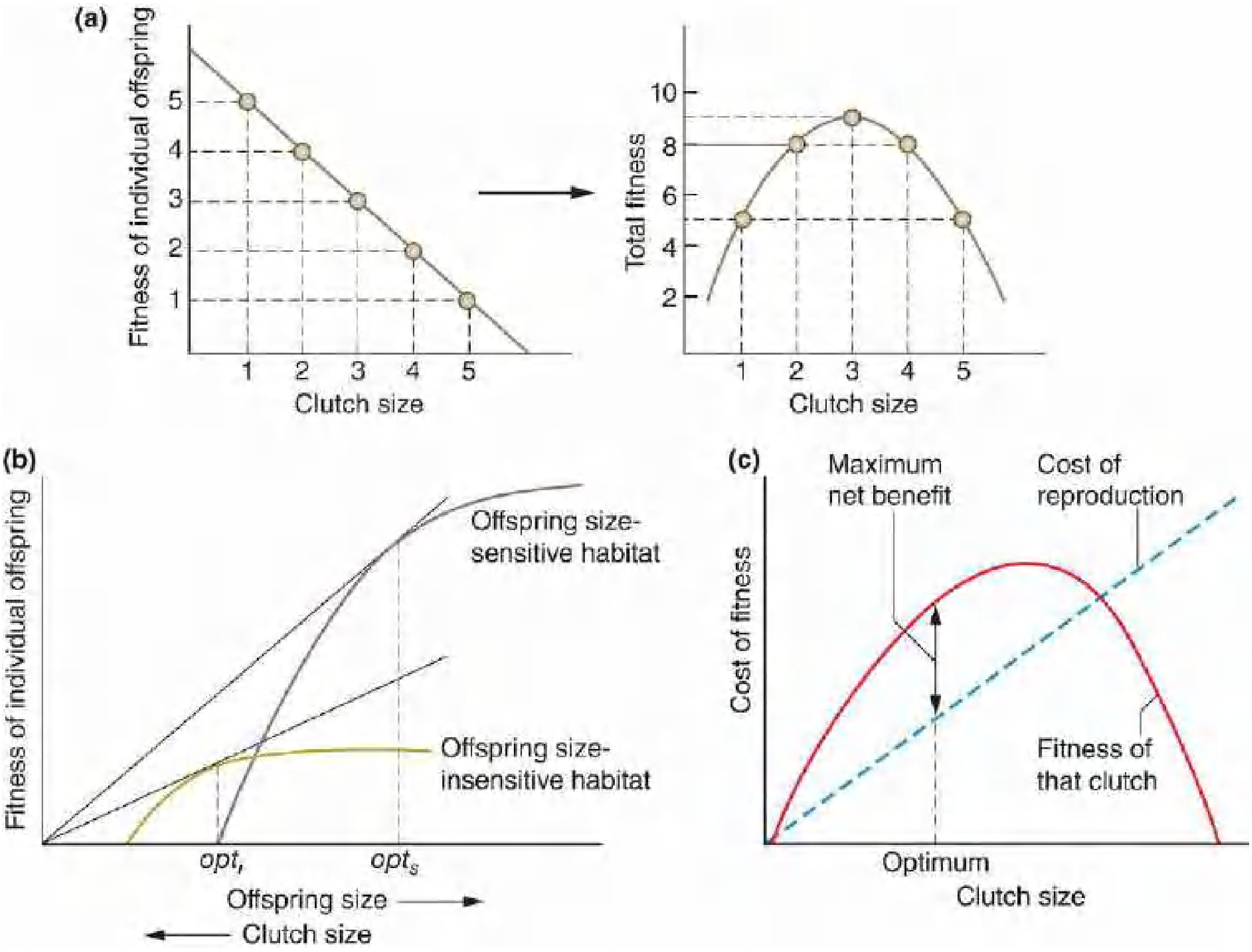


Figure 7.17 The optimisation of offspring size and number in a clutch or litter. (a) The ‘Lack clutch size’. If the fitness of each individual offspring decreases as total clutch size increases, as shown on the left, then the total fitness of a clutch (right; the product of number and individual fitness) must be maximised at some intermediate (‘Lack’) clutch size. (b) The Smith and Fretwell (1974) approach to the optimisation of offspring size and number. Offspring fitness increases with offspring size, but this increase follows a saturating curve. The curve saturates faster in more offspring size-insensitive habitats, but offspring size increases at the expense of clutch size, as noted on the horizontal axis. The optimum clutch size is where total fitness, the product of offspring number and their individual fitness, is greatest. Straight lines from the origin have the same total fitness. Steeper lines, within a given habitat, have greater total fitness, and the steepest line, with the greatest fitness, just touches the saturating curve. Hence, smaller clutches of larger offspring are favoured in more offspring size-sensitive habitats (opt_s compared with opt_i). (c) However, if there is also a cost of reproduction, then, following from (a), the ‘optimum’ clutch size is that where the net fitness is greatest, i.e. here, where the distance between the cost line and the ‘benefit’ (total clutch) curve is greatest.

Source: After Charnov & Krebs (1974).

A complementary approach to seeking an optimum clutch size was initiated by Smith and Fretwell (1974), who focused instead on the size of an individual offspring and its own fitness, arguing that there were diminishing returns in fitness as size increased (a ‘saturating’ curve; Figure 7.17b). Thus, since natural selection would favour parents with the maximum *product* of offspring number and their individual fitness, an optimum would again be favoured, dependent on resource availability and so on – in this case affecting the shape and location of the offspring fitness–size relationship (Figure 7.15b; see Dani & Kodandaramaiah, 2017). As discussed, for example, by Gaillard *et al.* (2014), there is no contradiction between the two approaches, though

Lack's has been more associated with the idea of a single optimum for a population, and Smith and Fretwell's with attempts to account for individual variation.

A number of attempts have been made to test the validity of Lack's proposal through experimental manipulation (especially with birds and to a lesser extent with insects) – adding or removing eggs from natural clutches, determining which clutch size is ultimately the most productive, and comparing this with the normal clutch size. Many of these have suggested that Lack's proposal is wrong, at least when taken at face value. A meta-analysis suggested, for example, that enlarged broods produced 0.55 more eggs than unmanipulated broods, on average (Van der Werf, 1992). The clutch size most commonly observed 'naturally' seems not always to be the most productive. However, we should be careful how we interpret such studies.

beyond the Lack clutch size

Firstly, many of the studies are likely to have made an inadequate assessment of the fitness of individual offspring. It is not enough to add two eggs to a bird's normal clutch of four and note that six apparently healthy birds hatch, develop and fledge from the nest. How well do they survive the following winter? How many chicks do they have themselves? In a long-term study of great tits (*Parus major*) near Oxford, UK, for example, 'addition' nests were more immediately productive than control nests (10.96 fledglings, on average, compared with 8.68), which were more productive than removal nests (5.68), but recruitment (i.e. survival of offspring to become breeding adults themselves) was highest from the natural, unmanipulated clutches (Figure 7.18).

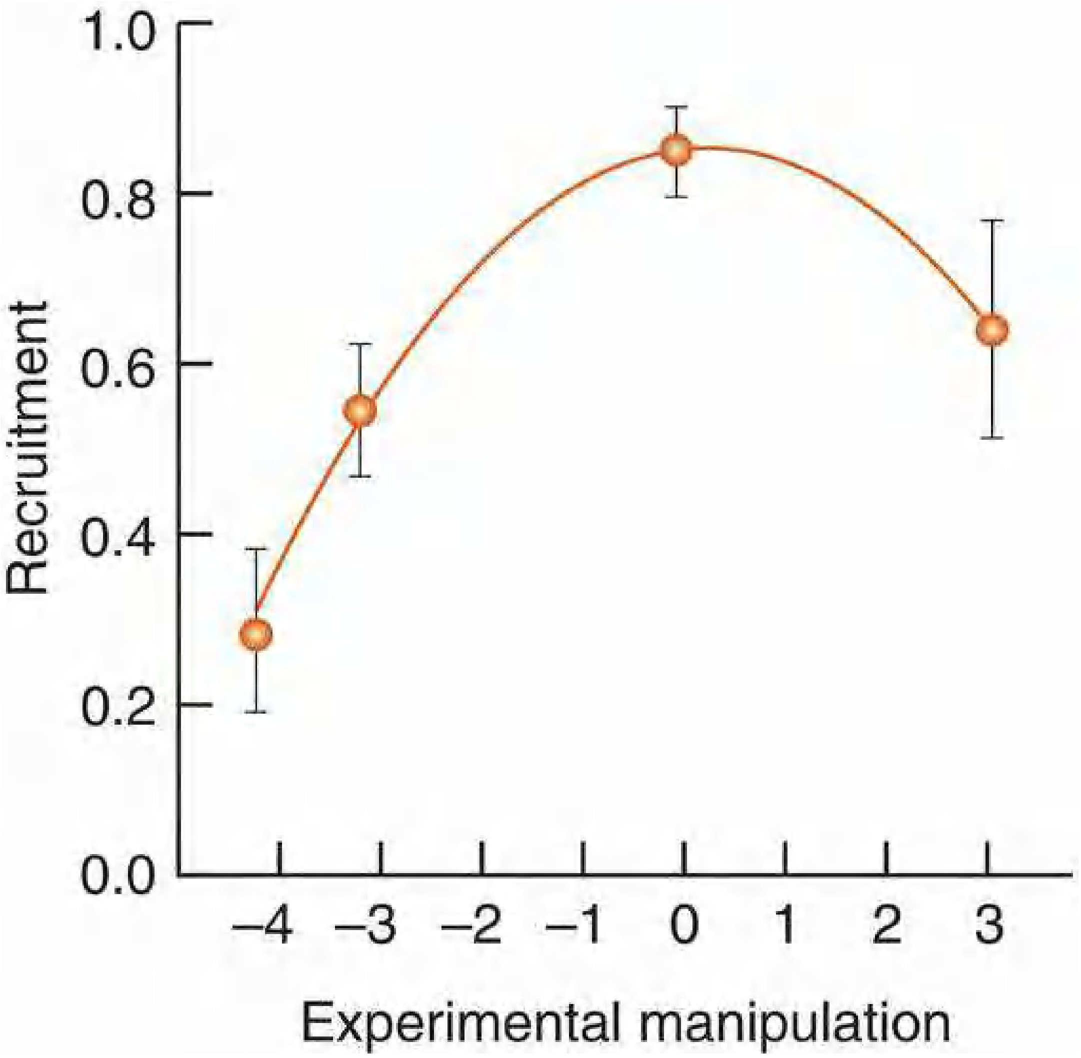
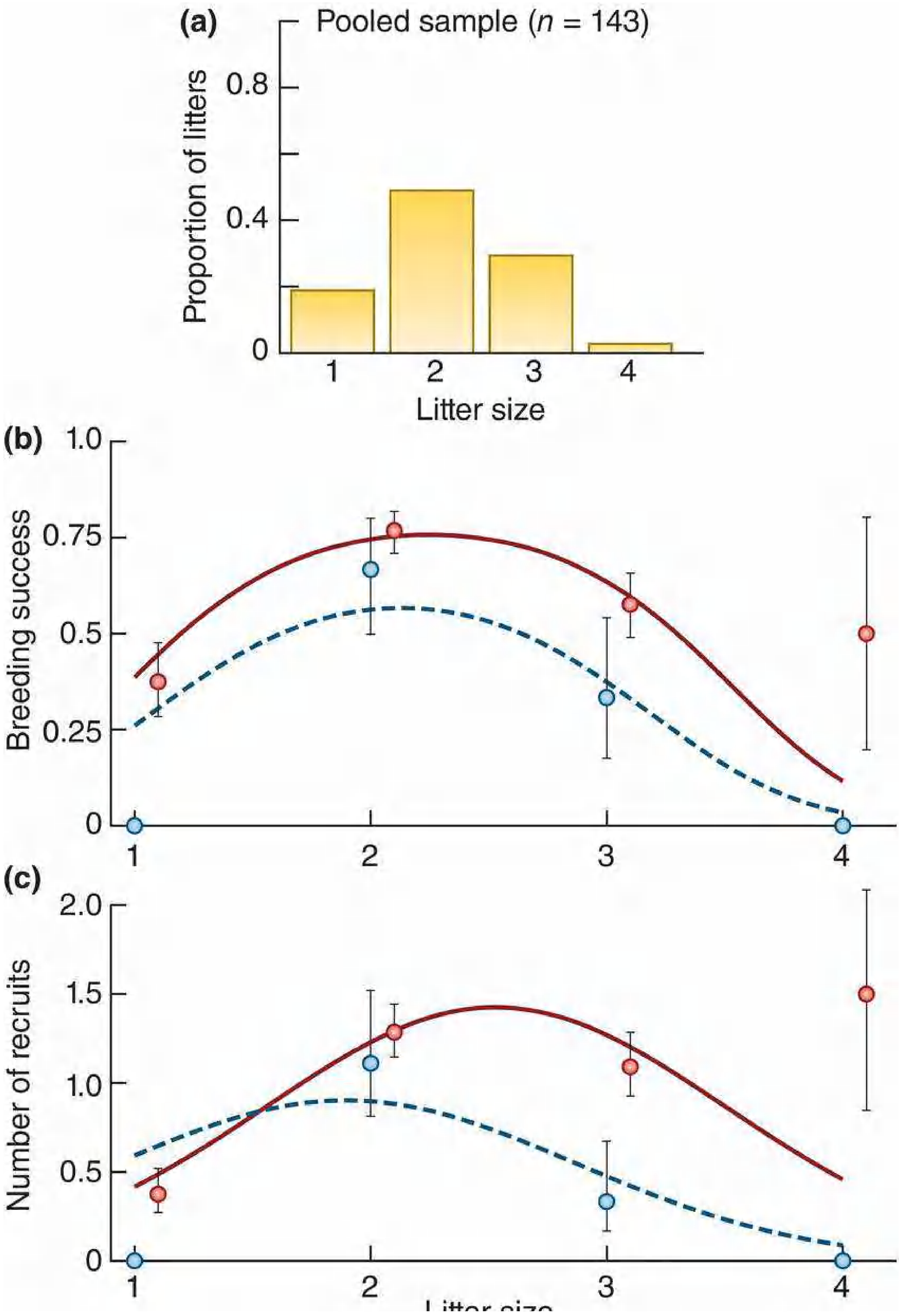


Figure 7.18 Recruitment from great tit, *Parus major*, nests is highest from natural, unmanipulated clutches. The mean observed number of young recruited per nest (R) (± SE) relative to experimental manipulations, M (additions to, or removals from, clutches) in great tits. The curve is the polynomial $R \sim M + M^2$.

Source: After Pettifor *et al.* (2001).

Secondly, and perhaps the most important omission from both Lack's and Smith and Fretwell's approaches, is their failure to consider any subsequent cost of reproduction. Natural selection will favour a lifetime pattern of reproduction that gives rise to the greatest fitness overall. A large and apparently productive clutch may extract too high a price in terms of RRV. The favoured clutch size will then be less than what appears to be the most productive in the short term (Figure 7.17c). Few studies have been sufficiently detailed to allow the cost of reproduction to be taken into account in assessing an optimal clutch size. In one, though, female bank voles (*Myodes glareolus*) were treated with gonadotropin hormones, inducing them to increase their reproductive allocation to a larger litter (Oksanen *et al.*, 2002). Treated females were considerably more productive at the time the litters were born, and a very small but nonetheless significant increase was maintained in the number of offspring surviving to the following winter. However, the treated females also paid a significant cost for their increased reproductive efforts: higher mortality during nursing, decreased body mass gain and a decreased probability of producing a subsequent litter. It may have seemed possible to 'improve' upon the natural optimum in the short term. But longer-term costs seem to have at least cancelled out these apparent improvements.

There are, nonetheless, studies that provide support for the Lack clutch size in the most direct sense of demonstrating the existence of an optimal, adaptive clutch size apparently independent of the state or circumstances of individual females. Eurasian lynx, *Lynx lynx*, in Sweden and Norway were studied over an extended period (7–13 years) from contrasting populations. Crucially, females were fitted with collars that allowed them to be identified individually and followed. The numbers of kittens they gave birth to (May and early June) were then counted, and the numbers of these recruited into the population noted the following November to January from tracks observed in the snow as they followed their mother before they became independent (when their dispersal might be misinterpreted as mortality). Although there was variation in litter size (Figure 7.19a), twinning (a litter size of two) was the most common, as indeed it is in captivity. It was also the most successful litter size, expressed either as a probability of recruit survival or the number of recruits reared from a litter (Figure 7.19b, c), and this was true irrespective of female body mass, of whether females were breeding for the first time or not, and of site and year. For lynx, at least, there is good evidence that the litter size most commonly observed, while not the physiological maximum, is generally the most productive.


(a)
Pooled sample (n = 143)
Proportion of litters
0.8
0.4
0
1
2
3
4
Litter size
(b)
Breeding success
1.0
0.75
0.5
0.25
0
1
2
3
4
(c)
Number of recruits
2.0
1.5
1.0
0.5
0
1
2
3
4

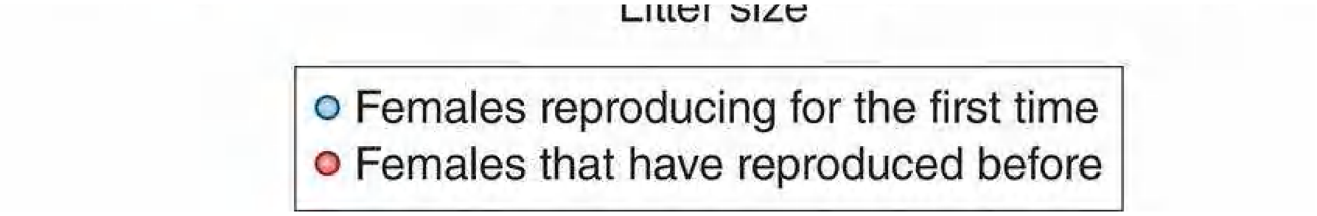


Figure 7.19 Evidence for an intermediate, Lack clutch (litter) size in lynx, *Lynx lynx*. (a) Frequency distribution of litter sizes of lynx in Sweden and Norway for four populations combined. (b) The relationship between litter size and breeding success for two-year-old females (probability of producing at least one recruit). Females reproducing for the first time are shown blue; those that had reproduced before are in red. Curves are for quadratic models that were significantly better fits to data than linear models. (c) Similar results but for the number of recruits. Bars are SEs.

Source: After Gaillard *et al*. (2014).

APPLICATION 7.2 Kiwis and Operation Nest Egg

Where an environment remains constant, the evolution of egg size and number, like evolution generally, can bring a species close to an optimal match-up (see Chapter 1). But it can become a trap when an environment is subject to sudden change. The native fauna of New Zealand, including the country's national symbol, the flightless kiwis (*Apteryx* spp.), evolved in an environment that provided little by way of threat from predators in the virtual absence of mammals. But the arrival of humans around 750 years ago, and of Europeans around 350 years later, led to the introduction of mammalian predators, most notably stoats (*Mustela erminea*), that now take a massive toll of the birds' eggs and chicks. There are predator-free offshore islands where the kiwis are safe, but in mainland sites their long-term persistence is threatened. Of particular relevance here, there is no realistic egg or chick size available to the kiwis that would take them out of the range that is vulnerable to stoat (and other) predation. In fact, the birds have evolved to produce a clutch of just one enormous egg – around six times the size of a hen's egg despite the adults being a similar size – suggestive of an offspring size-sensitive habitat prior to the recent introduction of predators: probably a low predation–high competition environment. We cannot be certain how egg size and number would have evolved had the predators been present for millennia, nor how they might evolve in future (although there is considered to be a significant probability of their extinction), but for now the birds are chained by their evolutionary past to a reproductive strategy that threatens their long-term survival.

There have, unsurprisingly, been campaigns to protect the birds by eliminating the predators. But an interesting alternative conservation strategy, initiated in the 1990s, is Operation Nest Egg (Colbourne *et al.*, 2005). The principle is simple. Kiwi eggs are taken from the wild and reared in 'crèches' – either in captivity or on predator-free islands – until they are old enough (usually around six months) and large enough to no longer be vulnerable to stoat predation (around 1200 g) – at which point they are released back into mainland sites. For an initial period when Operation Nest Egg was being established, there were five sites on the North Island for the brown kiwi, *A. mantelli*, and one on the South Island for the critically endangered rowi, *A. rowi*. For the brown kiwis, survival to six months was 81%, which was far greater than the 11% observed in unmanaged populations and greater, too, than the *c.*20% estimated to be necessary to sustain the kiwis there. Similarly, 21% of the offspring of eggs in the programme survived to become breeding adults three to four years later, exceeding the estimated *c.*6% required for sustained persistence. For the rowi, an 81% survival to six months compares with only 16% in a natural population where stoats were being intensively trapped. As a result, the population increased from around 160 to 200 over six breeding seasons. The programme continues to grow and expand and to be successful (https://www.kiwisforkiwi.org/).

7.6 Classifying life history strategies

7.6.1 *r*- and *K*-selection

Given the vast array of life histories that species exhibit – at some level, every life history is unique – there has been an understandable attraction in making sense of this diversity by classifying them into a limited number of 'types', and then associating these with the types of environmental forces that have led to their evolution. The most influential of these classifications has been the concept of *r*- and *K*-selection, originally propounded by MacArthur and Wilson (1967 ; MacArthur, 1962) and elaborated by Pianka (1970) (see also Boyce, 1984). The letter *r* refers to the intrinsic rate of natural increase (see Chapter 4) and indicates that *r*-selected individuals have been favoured for their ability to reproduce rapidly (i.e. have a high *r* value). The letter *K* (see Chapter 6) refers to the size ('carrying capacity') of a crowded population, limited by competition. Thus, *K*-selected individuals have been favoured for their ability to make a large proportional contribution to a population that remains near that carrying capacity – in other words, to be good competitors. The concept is therefore based on there being two contrasting types of habitat: *r*-selecting and *K*-selecting. It originally emerged (MacArthur & Wilson, 1967) from the contrast between species that were good at rapidly colonising relatively 'empty' islands (*r* species), and species that were good at maintaining themselves on islands once many colonisers had reached there (*K* species). Subsequently, the concept was applied much more generally. Like all generalisations, this dichotomy is an oversimplification – but one that has been immensely productive. It is important to remember, moreover, that the scheme deals in comparisons, not absolutes. It seeks to understand the differences between species (or populations of a species) by contrasting their characteristics (and the habitats in which they live) with one another, not by claiming, for example, that a species is '*r*-selected' when considered in isolation, based on the absolute size of its eggs.

K-selection

A *K*-selected population, then, lives in a habitat with little environmental fluctuation. As a consequence, a crowded population of fairly constant size is established. There is intense competition amongst the adults, and the results of this competition largely determine the adults' rates of survival and fecundity. The young also have to compete for survival in this crowded environment, and there are few opportunities for the young to become established as breeding adults themselves. In short, the population lives in a habitat that, because of intense competition, in the terms we have used above, is both high CR and offspring size-sensitive.

The predicted characteristics of these *K*-selected individuals, again as we have seen above, are therefore larger size, deferred reproduction, iteroparity (i.e. more extended reproduction), a lower reproductive allocation and larger (and thus fewer) offspring. The individuals will often invest in attributes that increase survival, as opposed to reproduction. But in practice, because of the intense competition, many of them will nonetheless have very short lives.

r-selection

By contrast, an *r*-selected population lives in a habitat that is either unpredictable in time or short lived. Intermittently, the population experiences benign periods of rapid population growth, free from competition (either when the environment fluctuates into a favourable period, or when a site has been newly colonised). But these benign periods are interspersed with malevolent periods of unavoidable mortality (either in an unpredictable, unfavourable phase, or when an ephemeral site has been fully exploited or disappears). The mortality rates of both adults and juveniles are therefore highly variable and unpredictable, and they are frequently independent of population density and of the size and condition of the individuals concerned. In short, the habitat is both low CR and offspring size-insensitive.

The predicted characteristics of *r*-selected individuals are therefore smaller size, earlier maturity, possibly semelparity, a larger reproductive allocation and more (and thus smaller) offspring. The individuals will invest little in survivorship, but their actual survival will vary considerably depending on the (unpredictable) environment in which they find themselves.

evidence for the *r*/*K* concept

The *r*/*K* concept has been useful as an organising principle in helping ecologists think about life histories, and there have certainly been studies that have provided broad support for the scheme. For instance, in a study of the common dandelion, *Taraxacum officinale*, plants were either left for five years in crowded, multispecies communities (*K*-selecting), or were 'weeded out' twice per year, early in their adult lives, such that they had to regrow again from root stumps or re-establish from newly deposited seed (*r*-selecting). They were then grown side by side, in a common garden, either from seeds or as cuttings from established plants (Figure 7.20). Plants from the *r*-selecting environment invested more in reproduction, as opposed to growth and survival: they were smaller (Figure 7.20a) but nonetheless produced more flower heads (Figure 7.20b) having begun to reproduce earlier (Figure 7.20c). As a result, they produced more seeds (Figure 7.20d), but those seeds were smaller (Figure 7.20e).

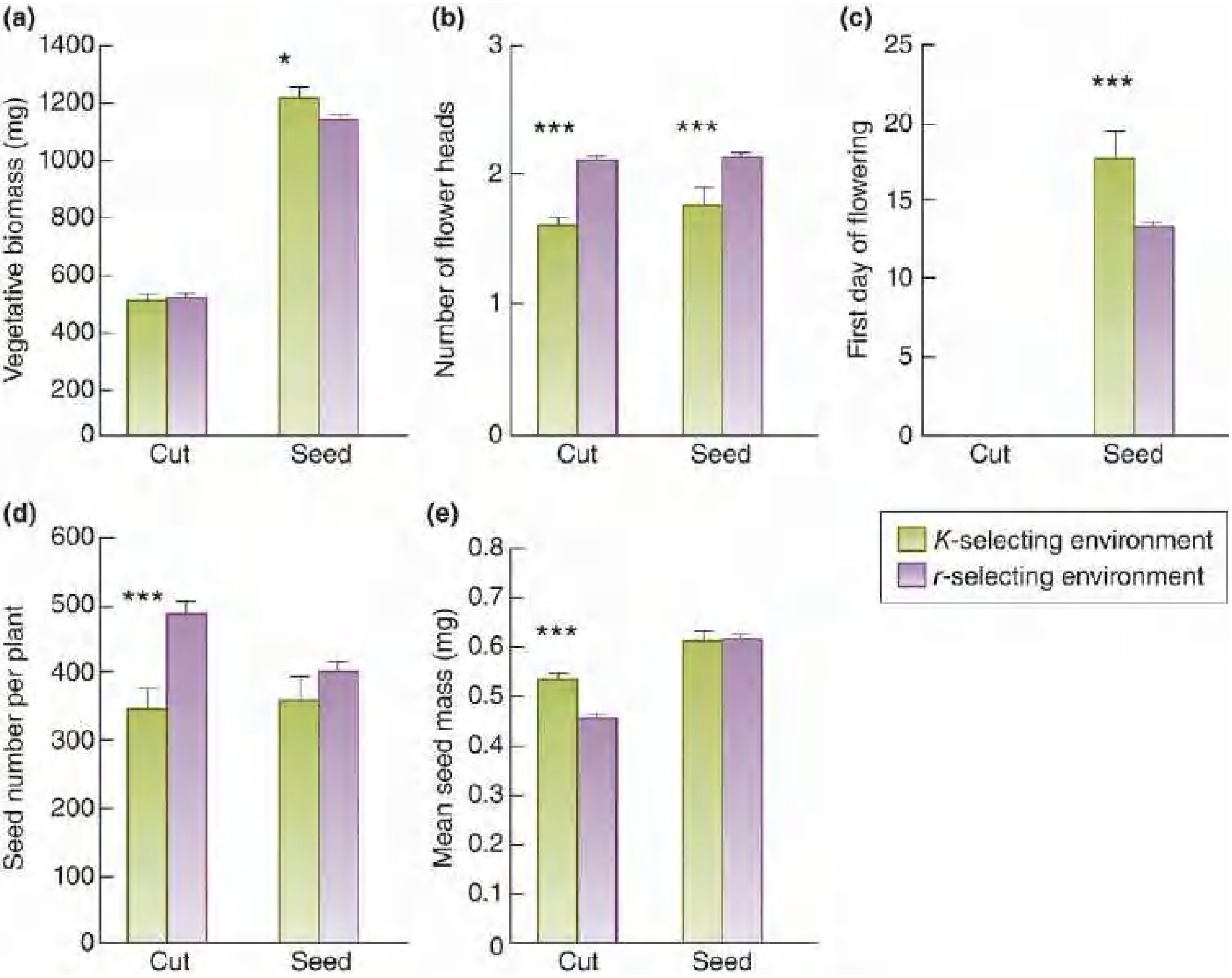


Figure 7.20 Evidence for *r*- and *K*-selection in dandelions. Results of an experiment in which dandelions, *Taraxacum officinale*, were subjected for five years to either a *K*-selecting environment (grown crowded with several other species) or an *r*-selecting environment (repeatedly weeded out and hence subjected to persistent density-independent mortality). They were then grown in a common garden either as cuttings from established plants ('cut') or from seed. Bars are ± 1 SE; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. (a) Vegetative biomass, (b) number of flower heads per plant, (c) period to first flowering, (d) seed number per plant, (e) mean seed mass.

Source: After Lipowsky *et al.* (2012).

the scheme explains much – but leaves as much unexplained

But support for the *r*/*K* scheme has been far from universal. Generally, reviews of studies pertinent to the concept, going back to Stearns (1977), have found roughly equal numbers of studies conforming to or in some way contradicting the scheme. We might regard this as a damning criticism of the *r*/*K* concept, since it shows that its explanatory powers are limited. On the other hand, a 50% success rate is hardly surprising given the number of additional factors that may further our understanding of life history patterns. For example, growth and large size, rather than reproduction, may be favoured because of size-specific predation (not intense competition, as in *K*-selecting environments); similarly eggs may be larger because their mothers are larger – there is a widespread correlation between the two (Cameron *et al*., 2016) – and not because of *K*-selection; and so on. It is therefore equally possible to regard it as very satisfactory that a relatively simple concept can help make sense of a large proportion of the multiplicity of life histories. Nobody, though, can regard the *r*/*K* scheme as the whole story.

7.6.2 A fast–slow continuum

An alternative, popular approach has been to consider organisms to be distributed along a 'fast–slow' continuum (Stearns, 1983), with high allocation to reproduction (and hence fast population growth) at one end and high allocation to survival at the other. This is clearly related to the *r*/*K* concept – *r* species can be thought of as fast species and *K* species as slow species – but differs in two important respects. Firstly, it is less 'inclusive' than the *r*/*K* scheme: species can be placed on a fast–slow continuum on the basis of only a subset of the traits included in the *r*/*K* scheme. Secondly, it seeks primarily to classify life histories, and then, in some cases, relate points on the continuum to different habitats, taxonomic groups and so on – whereas the *r*/*K* scheme incorporates contrasting habitats as an integral element from the outset. It is therefore less ambitious in its scope than the *r*/*K* scheme, and hence explains less in those cases where it is successful, but it is also more flexible and less likely to fail, no doubt accounting for its popularity.

a global study of plants

A good, and in many respects typical example of the use of the fast–slow continuum is shown in Figure 7.21, which describes the analysis of 625 plant species from around the world (Salguero-Gomez, 2017). A total of nine life history traits were examined for each species, including age at sexual maturity, degree of iteroparity, and so on, but rather than deciding *a priori* which of these would be important in distinguishing the species from one another, a principal component analysis was carried out. Very briefly, this statistical technique takes a set of explanatory variables that may themselves be correlated with one another and converts them to a smaller set of uncorrelated 'principle components' to which the original variables contribute, and to which their level of contribution can be measured (their 'loading'). The first principle component (PCA1) explains the greatest proportion of the original variation in traits, PCA2 explains the next greatest proportion, and so on. In this case, PCA1, which explains 35% of the variation, does indeed have loadings that place 'slow' traits (for example, long generation times, delayed maturity, low reproductive output) at one end of the axis and corresponding 'fast' traits at the other end. PCA2, however, explained a further 25% of the variation. Salguero-Gomez refers to this as a 'reproductive strategy' axis. It has species with extended reproductive periods during which they reproduce repeatedly to generate a large reproductive output at one end, and semelparous/low output species at the other. No further principal components had significant explanatory power.

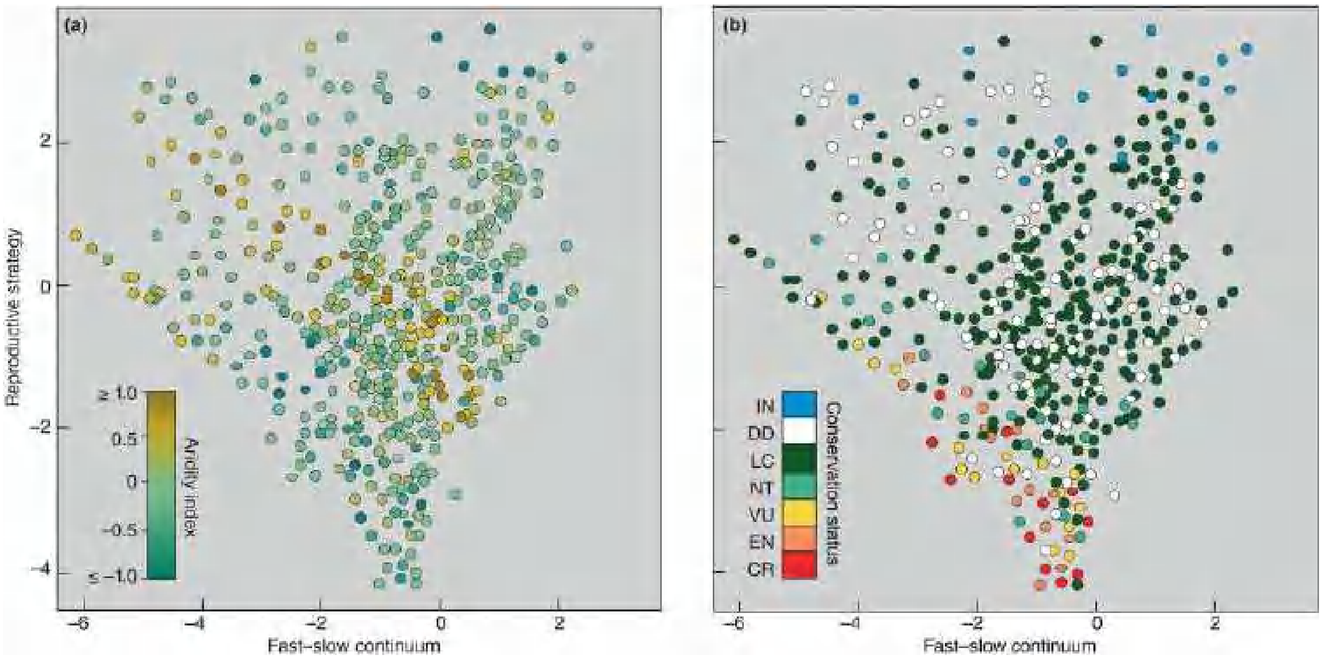


Figure 7.21 A fast–slow continuum explains life history variation in plants, but is not the whole story. (a) Results of a principle component analysis of 625 plant species, in which individual species are located relative to the first two principle components, respectively describing a fast–slow continuum and reproductive strategy, as explained in the main text. Species are colour-coded according to the score of their typical habitat on an aridity index – scores greater than zero being 'dry'. (b) A similar plot but in this case with the species colour-coded according to their conservation status: IN, invasive; DD, data deficient; LC, least concern; NT, near threatened; VU, vulnerable; EN, endangered; CR, critically endangered.

Source: After Salguero-Gomez (2017).

The study therefore supports the utility of the fast–slow continuum as an organising concept – this, in essence, was PCA1. But it also reaches an unsurprisingly common conclusion: that just like the *r*/*K* concept, the fast–slow continuum is not the whole story. Furthermore, it illustrates how the classification can be related subsequently to habitat or environmental correlates, without having incorporated these *a priori*. Figure 7.21a, for example, shows how the two-component classification relates to the aridity of the plants' habitats. There is an overall tendency (though by no means a general rule) for species at the fast end of the continuum to be found in more arid environments. This makes sense in that desert species, for example, must be able to exploit rapidly the brief periods when their growth and reproduction would be favoured. But the relationship of reproductive strategy to aridity is more complex with an apparent interaction between this and the fast–slow axis, such that plants of arid environments are concentrated along an upper-left to lower-right diagonal in Figure 7.21a. This suggests that desert-like environments may favour both faster, more iteroparous life histories and also slower, more semelparous ones. More generally, we can see here how the exercise itself suggests associations between life history and habitat that amount to hypotheses that subsequent studies can explore.

fast–slow continua for mammals

Another instructive study based on the fast–slow continuum looked at the life histories of mammals (Bielby *et al.*, 2007). Like the plant study just examined, it compiled data on a range of life history traits (seven) and then applied a statistical technique, in this case factor analysis, to generate a smaller number of 'factors' with significant explanatory power in accounting for the overall variation. Again, each factor combined a subset of the seven traits with appropriate loadings. The analysis was applied to all mammals and separately to various subgroups for which the dataset was rich, making nine analyses in all. As in the plant study, two factors stood out, explaining 53–78% of the total variance between them, such that in seven of the nine cases, adding a third factor did not significantly improve explanatory power. The analyses varied in which of the factors had most and which the second-most explanatory power. Figure 7.22 shows results for five of the subgroups.

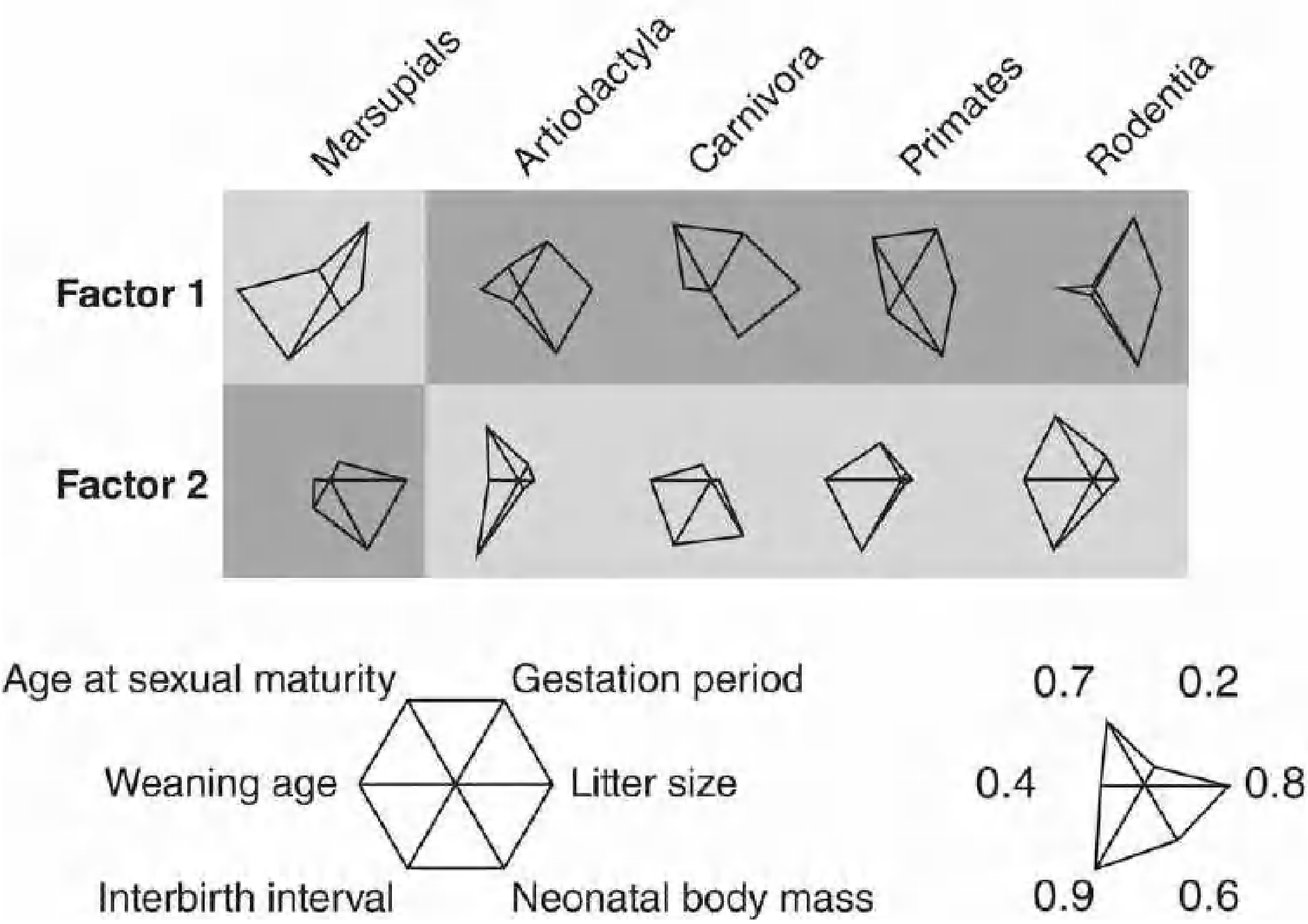


Figure 7.22 A fast–slow continuum explains life history variation in mammals, but is not the whole story. Results of a factor analysis of the life histories of five groups of mammals, as indicated at the top of the figure. Details are shown for the two main factors, incorporating six life history traits. The loadings of the six traits are visualised by star plots explained at the base of the figure. Each trait's loading (from 0 to 1) is represented by the length of the line from the centre to the tip of its point. An idea of scale is given by the illustrative example to the right. For each mammal group, the most important factor is placed above and the second most important below. The fast–slow factor is light shaded; the offspring size and number factor is dark shaded.

Source: After Bielby *et al.* (2007).

This time, however, the authors described both factors as reflecting a fast–slow continuum, and one of the two certainly did, with heavy loadings from age at weaning, age at maturity and interbirth interval. The other factor had heavy loadings from neonatal body mass, gestation length and (with an opposite sign) litter size. It therefore represented the trade-off between offspring number and size. These offspring traits are by no means always included in fast–slow comparisons, though it is easy to see why the production of lots of small offspring might be considered fast rather than slow. Of course, doing so brings this two-factor classification even closer to the *r*/*K* concept, and this mammal study therefore illustrates further the connection between *r* and *K* and fast and slow, as well as emphasising that any organisation of life histories along a fast–slow continuum – or indeed any other single axis – leaves residual variation to be explained. It also shows again how a life history classification can be related subsequently to other factors – in this case mammal phylogeny – generating hypotheses for further testing. It is clear from Figure 7.22, for example, that age at sexual maturity played a large part in placing species on the fast–slow continuum for rodents and artiodactyls but only a negligible role for carnivores and marsupials. The explanation remains to be determined.

APPLICATION 7.3 The fast–slow continuum, invasion and conservation

Salguero-Gomez's plant study also illustrates how relating the fast–slow continuum to 'outside' factors can have applied significance. Thus, Figure 7.21b shows his two-dimensional classification related to that used by the International Union for the Conservation of Nature for conservation and invasive status: endangered, critically endangered, invasive species and so on (see Section 15.4.2). The ideal in conservation is to be prepared for problems before they occur and perhaps even act to avoid or mitigate them. In this case, both axes were strongly correlated with conservation status. Species were more likely to be invasive and less likely to be endangered at the slow end of the spectrum and if they were iteroparous with an overall high reproductive output. This suggests that if we wish, for example, to predict and prepare for a species' invasion, we should be worried most about slow species that are able to 'hold their ground' rather than fast species that start reproducing quickly, especially if these reproduce only once. These fast species, however, seem most prone to being endangered.

Intriguingly, the picture may be different when we turn from plants to animals. Cortes (2002), for example, has explored the relationship between body size, age at maturity, generation time and the finite rate of population increase λ (referred to in Chapter 4 as *R*), by generating age-structured life tables for 41 populations of 38 species of sharks that have been studied around the world. A three-dimensional plot of λ against generation time and age at maturity reveals a fast–slow continuum running from upper-right-front to lower-left-back, with species characterised by early age at maturity, short generation times and generally high λ at the fast end of the spectrum (Figure 7.23a) and also having a smaller body size (Figure 7.23b). Cortes (2002) then used elasticity and sensitivity analyses (again, see Chapter 4) to assess the various species' ability to respond to changes in survival (due, for example, to human disturbance such as pollution or harvesting). 'Fast' sharks, such as *Sphyrna tiburo*, could compensate for a 10% decrease in adult or juvenile survival by increasing the birth rate. But for large, slow-growing, long-lived species, such as *Carcharhinus leucas*, even moderate reductions to adult or, especially, juvenile survival require a level of compensation in the form of fecundity or postbirth survival that such species cannot provide. It seems therefore to be the larger, slower species that are more vulnerable to extinction in this case. Work by Dulvy and Reynolds (2002) on skates and rays (Rajidae), closely related to sharks, provide a graphic illustration of Cortes' warning. Of the world's 230 species, only four are known to have undergone local extinctions and significant range reduction, and these are among the largest of their group (Figure 7.23c). Dulvy and Reynolds therefore propose that seven further species, each as large or larger than the locally extinct taxa, should be prioritised for careful monitoring. A similar pattern, too, is shown among Australian terrestrial marsupials (Figure 7.23d).

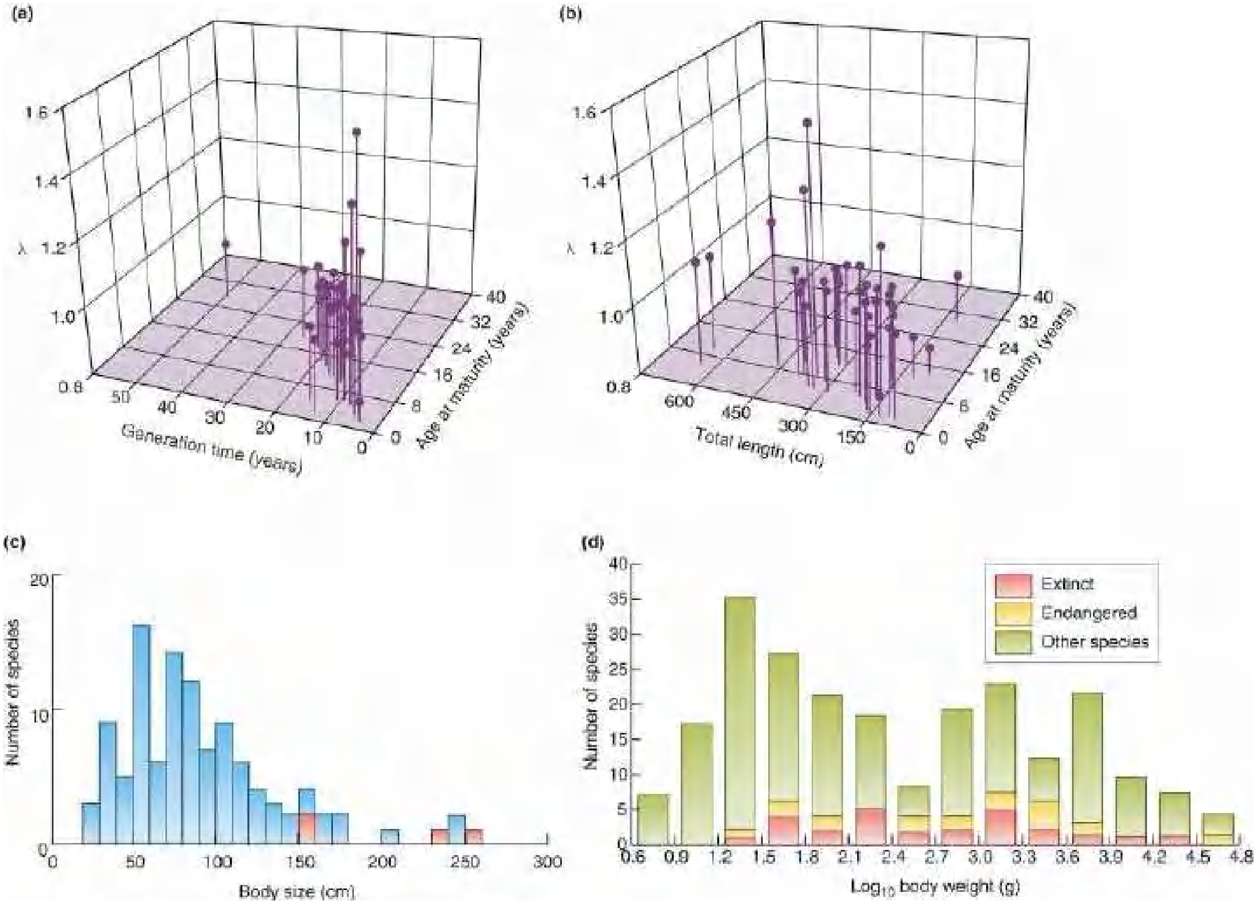
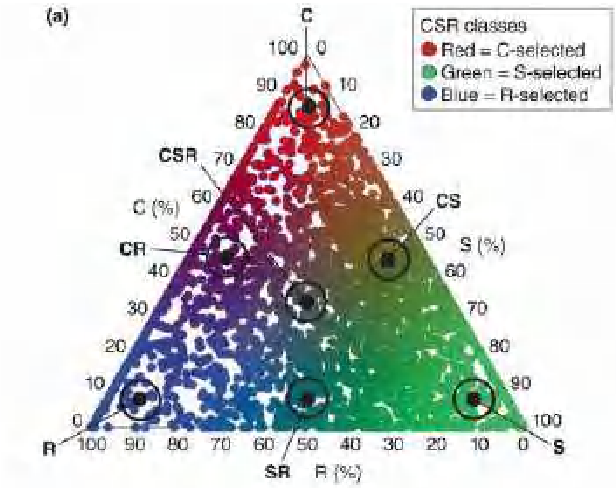


Figure 7.23 The fast–slow continuum can help guide conservation priorities in sharks and marsupials. Mean population growth rates, λ, of 41 populations from 38 species of shark in relation to: (a) age at maturity and generation time and (b) age at maturity and total body length. (c) Frequency distribution of skate body size with the four species that have gone locally extinct coloured red. (d) Body size frequency distribution of the Australian terrestrial marsupial fauna including 25 species that have gone extinct in the last 200 years (red). Sixteen species currently considered endangered are shown in yellow.

Source: (b) After Cortes (2002). (c) After Dulvy & Reynolds (2002). (d) After Cardillo & Bromham (2001).

7.6.3 Grime’s CSR triangle

Grime (1974; see also Grime & Pierce, 2012) produced a classification of plant life histories and habitats that bears at least some similarities to the *r*/*K* scheme and hence also to the fast–slow continuum. We can reasonably suppose that this consistency between schemes is evidence of an underlying truth. The differences between them are more about how they are to be applied and to which groups of organisms, though Grime’s, like the *r*/*K* scheme, incorporates habitat from the outset. Habitats are seen as varying in two significant respects – in their level of disturbance (brought about by grazing, disease, trampling or adverse weather) and in the extent to which they experience ‘stress’ (shortages of light, water or nutrients that limit photosynthesis). Grime then argued that a stress-tolerant strategy (S) is appropriate when stress is severe and disturbance uncommon; that a so-called ruderal strategy (R) is appropriate when disturbance levels are high but conditions are benign and resources abundant; and that a competitive strategy (C) is appropriate when disturbance is rare, resources are abundant and crowded populations develop. R species are typically annuals or short-lived herbaceous perennials with a capacity for rapid seedling establishment and growth, and a tendency to direct a high proportion of photosynthate into seeds. They are therefore similar to *r* or fast species. C species typically have high relative growth rates (high growth rates for their size), the ability to spread by vegetative means, and a tall stature. They therefore have some similarity to K or slow species (more growth than reproduction) but the focus is more on traits that are especially relevant to plants (and other modular organisms). Finally, S species tend to be small in stature with low relative growth rates. Such species are therefore also ‘slow’, but they are investing in tolerating harsh conditions rather than holding their ground in a competitive environment. Hence, they have no real equivalent in the r/K scheme. Taken together, this allows species to be placed in a triangle with extreme R, C and S species at the corners (Figure 7.24), though Grime’s classification also allows for the possibility of intermediate species and habitats – CR, CS, SR and even CSR.



(b) Tropical and subtropical moist broadleaf forests

(c) Mediterranean forests, woodlands and scrub

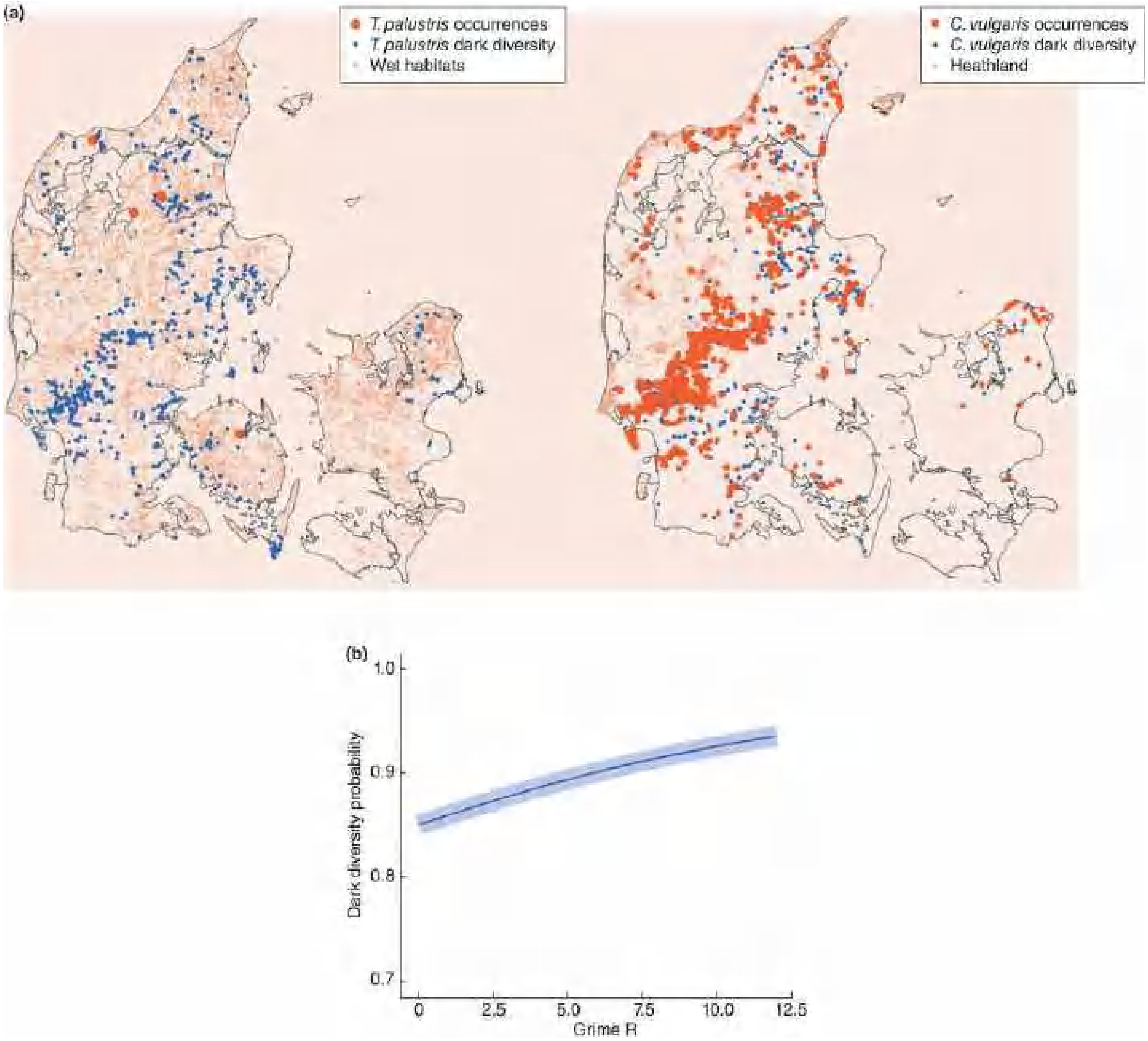
[**Figure 7.24**](#) **Grime's CSR triangle as an organising principle for plant life histories.** (a) The distribution of 3068 terrestrial higher plant species from around the world according to their position in Grime's CSR triangle, as explained in the main text. (b, c) The distribution of all plant species and the median of these (large black dot in triangles on right) for, respectively, tropical and subtropical moist broadleaf forests (b) and Mediterranean forests, woodlands and scrubs (c). The global locations of these biomes are shown on the maps to the left.

Source: After Pierce *et al.* (2017).

Figure 7.24a illustrates, in particular, the CSR classification of 3068 higher land plants from throughout the world, based on just three plants traits that were available for all plants, which Pierce *et al.* (2017) showed were very nearly as informative as a much fuller set of 14 traits that were available for a subset of plants, allowing the comparison to be made. These traits were leaf area (LA), specific leaf area (SLA; leaf area per unit mass) and leaf dry matter content (LDMC; leaf dry mass as a percentage of water-saturated fresh mass). In a principal component analysis, high SLA and high LDMC were at opposite ends of PCA1, while PCA2 was dominated by variation in LA. This then facilitated a CSR classification based on C species having high LA (and tending always to be intermediate on PCA1), S species having a low LA, a low SLA and high LDMC (so-called ‘conservative’ leaf economics, investing in long-lived rather than rapidly photosynthesising leaves), and R species having a low LA but a high SLA and low LDMC (‘acquisitive’ leaf economics with rapid photosynthesis). The classification can then be used, for example, to characterise and then contrast different biomes from across the globe. We can see in Figures 7.24b and 7.24c that plants in tropical and subtropical moist broadleaf forests are concentrated in the CS region of the triangle with very little contribution from R species, whereas plants in Mediterranean biomes are much more evenly spread with, if anything, an absence of CS species. Further utility of the CSR triangle is shown in Application 7.4.

APPLICATION 7.4 CSR and dark diversity

‘Dark diversity’ refers to an imaginary community of organisms that are present in the regional pool of species (and hence ‘available’) but absent from a given location even though conditions there would favour their presence. Moeslund *et al*. (2017) carried out an assessment of dark diversity for 564 plant species throughout Denmark at a fine spatial resolution that could aid local decision making (areas of 78 m^2). An argument can be made that our conservation efforts are best directed at those species that contribute most to dark diversity, because it is for these species that the gap between their actual and potential distribution is greatest. We can see in Figure 7.25a, for example, the contrast in Denmark between *Tephroseris palustris* and *Calluna vulgaris* (heather). *Calluna* makes little contribution. It is already present in two-thirds of the sites where it ‘could’ be. *Tephroseris*, on the other hand, is present in fewer than 1% of potential sites, and offers, arguably, much greater opportunities for management interventions to increase diversity. For this, though, it is necessary to understand what features are most important in determining contributions to dark diversity, and in this case, typical dark diversity species tended to live at low light and nutrient levels and be poor dispersers, but in the present context were also ruderals (R) and stress intolerant (not S) (Figure 7.25b). This would suggest that one important way of promoting increased diversity, enabling species to live throughout more of their available range, would be to ensure a sufficiency of bare soil in which ruderal species can germinate successfully.



[Figure 7.25](#) Grime's CSR triangle and dark diversity as a guide to conserving plants in Denmark. (a) Contrasting maps illustrating a species with a high dark diversity, above (*Tephroseris palustris*, 99.6%) and a species with a low dark diversity probability below (*Calluna vulgaris*, 37.3%). The distributions of the plants' typical habitat types are also shown. Species are not necessarily predicted to exist wherever their habitat type is present but may nonetheless be absent (contributing to 'dark diversity' at a range of sites). (b) The contribution of Grime's R (ruderality) to a plant species' probability of dark diversity as output of a statistical model also incorporating other predictors (nutrient availability, etc.). The shading represents uncertainty around the predicted relationship (blue line) as estimated from the variation in output among 1000 simulations of the model.

Source: After Moeslund *et al.* ([2017](#)).

7.7 Phylogenetic and allometric constraints

organisms are prisoners of their evolutionary past

The life histories that natural selection favours (and we observe) are not selected from an unlimited supply but are constrained by the phylogenetic or taxonomic position that organisms occupy. For example, in the entire bird order Procellariiformes (albatrosses, petrels, fulmars) the clutch size is one, and the birds are 'prepared' for this morphologically by having only a single brood patch with which they can incubate this one egg (Ashmole, 1971). A bird might produce a larger clutch, but this is bound to be a waste unless it exhibits concurrent changes in all the processes in the development of the brood patch. Albatrosses are therefore prisoners of their evolutionary past, as are all organisms, and as such are confined to a limited range of habitats.

It follows from these 'phylogenetic' constraints that caution must be exercised when life histories are compared. The albatrosses, as a group, may be compared with other types of birds in an attempt to discern a link between the typical albatross life history and the typical albatross habitat. The life histories and habitats of two albatross species might reasonably be compared. But if an albatross species is compared with a distantly related bird species, then care must be taken to distinguish between differences attributable to habitat (if any) and those attributable to phylogenetic constraints.

7.7.1 Effects of size and allometry

One element of phylogenetic constraint is that of size. Figure 7.26a shows the relationship between time to maturity and size (weight) in a wide range of organisms from viruses to whales. We can see that particular groups of organisms are confined to particular size ranges. For instance, unicellular organisms cannot exceed a certain size because of their reliance on simple diffusion for the transfer of oxygen from their cell surface to their internal organelles. Insects cannot exceed a certain size because of their reliance on unventilated tracheae for the transfer of gases to and from their interiors. Mammals, being endothermic, must exceed a certain size, because at smaller sizes the relatively large body surface would dissipate heat faster than the animal could produce it, and so on. We can also see that time to maturity and size are strongly correlated. In fact, as Figure 7.26b–d illustrates, size is strongly correlated with many (probably most) life history components. Thus, since the sizes of organisms are constrained by their phylogenetic position, these other life history components will be constrained too.

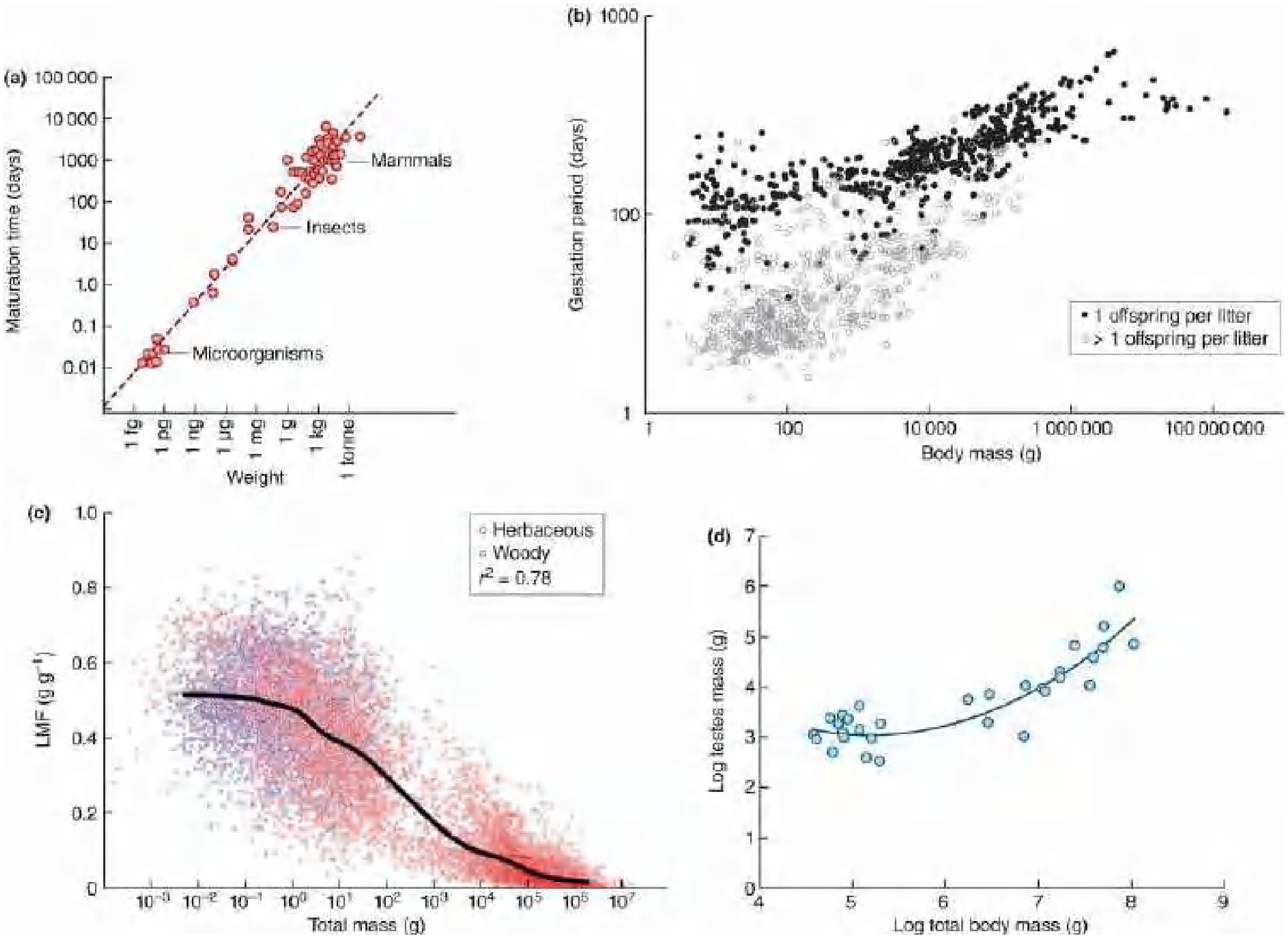


Figure 7.26 **Allometric life history relationships**, all plotted on log scales. (a) Maturation time as a function of body weight for a broad range of animals and microorganisms. (b) Gestation time in eutherian mammals as a function of body mass. Species with one offspring per litter and more than one offspring are shown separately. Slopes are, for one offspring, 0.10 (0.10–0.11 95% CI) with ordinary regression and 0.08 (0.07–0.10 95% CI) when taking account of phylogenetic structure within the dataset, and for more than one offspring, 0.17 (0.15–0.18 95% CI) with ordinary regression and 0.08 (0.06–0.10 95% CI) when taking account of phylogenetic structure within the dataset. $P < 0.001$in all cases. (c) The fraction of total mass allocated to leaves (LMF) as a function of total mass in a dataset of 8170 woody and 2960 herbaceous plant species. The bold line is a curve fitted through the means of 50 consecutive size classes. The slope (the allometric relationship) clearly changes as total mass changes. (d) Testes mass as a function of somatic body mass in 31 species of cetaceans. Again, the relationship clearly changes with somatic mass.

Source: (a) After Blueweiss *et al*. (1978). (b) After Clauss *et al*. (2014). (c) After Poorter *et al*. (2015). (d) After Macleod (2010).

the importance of allometries

Note in particular that the examples in Figure 7.26 are all examples of *allometric* relationships that were first discussed in Section 3.9 as part of a metabolic theory of ecology – that is, relationships in which a physical or physiological property of an organism changes *relative* to the size of the organism, rather than changing in direct proportion to the changing size. What, though, is the significance of allometry in the study of life histories? The usual approach to the ecological study of life histories has been to compare the life histories of two or more populations (or species or groups), and to seek to understand the differences between them by reference to their environments. It is clear by now, however, that taxa can also differ because they lie at different points on the same allometric relationship, or because they are subject to different phylogenetic constraints generally. It is therefore important to disentangle ‘ecological’ differences from allometric and phylogenetic differences (see Harvey & Pagel, 1991 ; Freckleton et al., 2002). This is not because the former are ‘adaptive’ whereas the latter are not. Indeed, we saw in Chapter 3 that in allometric relationships, a fit between organism and environment is maintained as size changes. Rather, it is a question of the evolutionary responses of a species to its habitat being limited by constraints that have themselves evolved.

comparing salamanders: dangerous if allometries are ignored

These ideas are illustrated in Figure 7.27a, which shows the allometric relationship between clutch volume and body volume for salamanders generally. Figure 7.27b then shows the same relationships in outline; but superimposed upon it are the allometric relationships within populations for two salamander species, *Ambystoma tigrinum* and *A. opacum* (Kaplan & Salthe, 1979). If the species’ means are simply compared, without reference to the general salamander allometry, then the species are seen to have the same ratio of clutch volume : body volume (0.136). This seems to suggest that the species’ life histories ‘do not differ’, and that there is therefore ‘nothing to explain’ – but any such suggestion would be wrong. *A. opacum* conforms closely to the general salamander relationship. *A. tigrinum*, on the other hand, has a clutch volume which is almost twice as large as would be expected from that relationship. Within the allometric constraints of being a salamander, *A. tigrinum* is making a much greater reproductive allocation than *A. opacum*; and it would be reasonable for an ecologist to look at their respective habitats and seek to understand why this might be so.

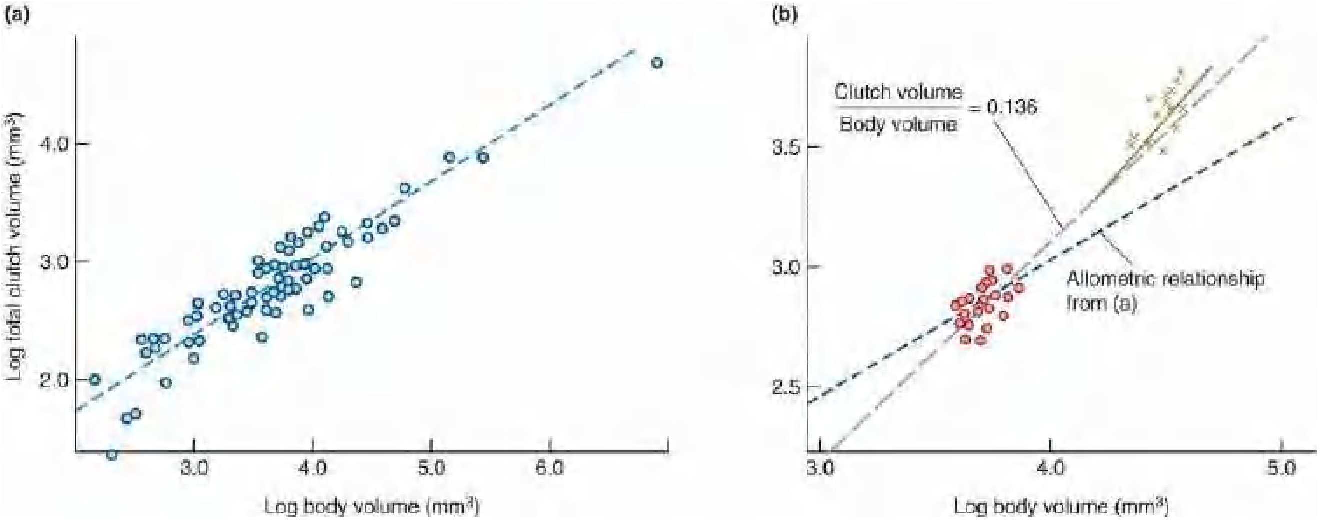


Figure 7.27 Allometric relationships between total clutch volume and body volume in female salamanders. (a) The overall relationship for 74 salamander species, using one mean value per species ($P < 0.01$). (b) The relationships within a population of *Ambystoma tigrinum* (green crosses) ($P < 0.01$), and within a population of *A. opacum* (red circles) ($P < 0.05$). The allometric relationship from (a) is shown as a blue dashed line: *A. opacum* conforms closely to it; *A. tigrinum* does not. However, they both lie on an isometric line along which clutch volume is 13.6% of body volume (purple dashed line).

Source: After Kaplan & Salthe (1979).

In other words, it is reasonable to compare taxa from an ‘ecological’ point of view as long as the allometric relationship linking them at a higher taxonomic level is known (Clutton-Brock & Harvey, 1979). It will then be their respective deviations from the relationship that form the basis for the comparison. Problems arise, though, when allometric relationships are unknown (or ignored). Without the general salamander allometry in Figure 7.27a, the two species would have seemed similar when in fact they are different. Conversely, two other species might have seemed different when in fact they were simply conforming to the same allometric relationship. Comparisons that ignore allometries are clearly perilous, but regrettably, ecologists are frequently oblivious to allometries. Typically, life histories have been compared, and attempts have been made to explain the differences between them, in terms of habitat differences. As previous sections have shown, these attempts have often been successful; but they have also often been unsuccessful, and unacknowledged allometries undoubtedly go some way towards explaining this.

7.7.2 Effects of phylogeny

An impression of the strength of the influence of phylogeny can be gained from an early study of the variation in seven life history traits amongst a large number of mammal species (Read & Harvey, 1989). A nested analysis of variance led to the determination of the percentage of the total variance attributable to: (i) differences between species within genera; (ii) differences between genera within families; and so on. Species vary very little within genera (3–11% of the variance) and genera vary little within families (6–14%). Far and away the largest part of the variance, for all the traits, is accounted for by differences between orders within the mammalian class as a whole (55–71%). This emphasises that in simply comparing two species from different orders, we are in essence comparing those orders (which probably diverged many millions of years ago) rather than the species themselves. (It does not mean, however, that comparing species in the same genus, say, is pointless. Even when two species are very similar in their life histories and habitats, if one makes a greater reproductive allocation and also lives in a habitat that is lower CR, then this allows us to build a pattern linking the two.)

the independent contrasts approach

Studies such as these have led to the very widespread adoption of ‘phylogenetic comparative’ statistical methods (PCMs) that only compare life histories after the effects of shared ancestry have been taken into account, especially the ‘independent contrasts’ approach (Harvey & Pagel, 1991 ; O’Meara, 2012 .) Very briefly, the method requires the construction of a phylogeny linking the species so that the evolutionary distance between two species and their shared ancestor can be estimated, and this connection is then taken into account rather than assuming, say, that two species with a recent common ancestor are statistically independent units of observation. We see an example where this can make an important difference if we return to the allometric relationship between gestation time and size in eutherian mammals shown in Figure 7.26b (but this time combining all species, whether they produce one or more than one offspring per litter). Figure 7.28a shows the slope (and 95% confidence interval) when ordinary least squares analysis (OLS) was compared with a phylogenetic generalised least squares analysis (PGLS) for all species (as in Figure 7.26b), but also for 100 random subsets that selected only one species from each of 105 families, or selected one species from each of 20 orders. When all species were combined, the OLS and PGLS slopes differed, but when only one species was included per family or per order they did not, and their shared slope values converged on the OLS value for all species. This suggests that when all species were included, the PGLS slope was a truer reflection of interspecific differences, whereas the OLS slope also incorporated relationships between whole families and orders, and hence, to a significant degree, was reflective of phylogenetic history.

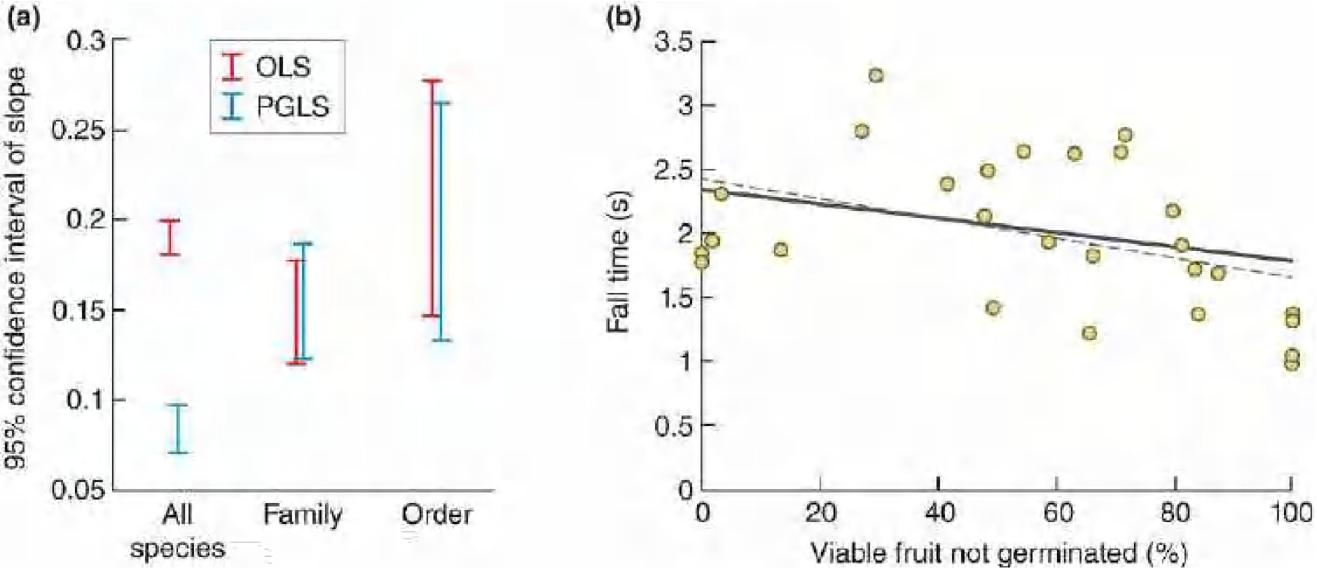


Figure 7.28 Phylogenetic comparative methods, taking account of shared ancestry, help make sense of patterns of life history variation. (a) Analysis of the slope of allometric relationship (95% CI) between gestation period and body mass in mammals (see Figure 7.25b), using either ordinary least squares (OLS) or taking phylogeny into account by using phylogenetic generalised least squares (PGLS), and applying this either to all species or to subsets that allowed only one species per family or per order. (b) The allometric relationship between dispersibility (time taken for a fruit to fall through a 2.5 m tube) and the tendency to remain dormant (% of viable fruit not germinating) in South African daisies. The regression accounting for phylogeny (solid line; $P = 0.014$) was statistically superior to one that did not (dashed line; $P = 0.025$), but the relationships were functionally very similar.

Source: (a) After Clauss *et al.* (2014). (b) After de Waal *et al.* (2016).

In other cases, the strength of the statistical relationship is improved by taking phylogeny into account but the functional form of the relationship is little altered. Figure 7.28b shows an example of this for the relationship between dispersibility and a tendency to remain dormant in South African plants in the family Asteraceae (daisies). In still other cases, using PCMs makes little difference either functionally or statistically. We should also remember that PCMs, like all statistical methods, are only as solid as the validity of the assumptions on which they are based – here, in particular, that the phylogeny is correct – and that an analysis incorporating phylogeny should therefore not be assumed, uncritically, to be superior (Cooper *et al.*, 2016). Nonetheless, such methods have increasingly, and quite rightly, become the norm.

On the other hand, the strong tendency for life history features to be shared at the higher taxonomic levels does not mean that attempts to relate life histories to lifestyles and habitats should be abandoned there, since lifestyles and habitats are also constrained by an organism's size and phylogenetic position. There may still, therefore, at these higher levels, be patterns linking habitats and life histories rooted in natural selection. For example, in the earliest stages of the development of the r/K scheme, insects (small size, many offspring, high reproductive allocation, frequent semelparity) were described as relatively r-selected, compared with mammals (large size, few offspring, etc. – relatively K-selected) (Pianka, 1970). Such differences could be dismissed as being 'no more' than the product of an ancient evolutionary divergence, but as we have stressed, an organism's habitat reflects its own responses to its environment. Hence a mammal and an insect living side by side are also almost certain to experience very different habitats. The larger, homeothermic, behaviourally sophisticated, longer-lived mammal is likely to maintain a relatively constant population size, subject to frequent competition, and be relatively immune from environmental catastrophes and uncertainties. The smaller, poikilothermic, behaviourally unsophisticated, shorter-lived insect, by contrast, is likely to live a relatively opportunistic life, with a high probability of unavoidable death. Insects and mammals are prisoners of their evolutionary past in their range of habitats just as they are in their range of life histories – and the r/K scheme provides a reasonable (although certainly not perfect) summary of the patterns linking the two.

It is undoubtedly true that life history ecology cannot proceed oblivious to phylogenetic and allometric constraints. Yet it would be unhelpful to see phylogeny as an alternative explanation to habitat in seeking to understand life histories. Phylogeny sets limits to an organism's habitat just as it does to its life history. The essentially ecological task of relating life histories to habitats remains a fundamental challenge.

allometry, phylogeny and the fast–slow continuum in mammals

A study bringing many of these ideas together was carried out by Sibly and Brown (2007), who examined a dataset of 637 placental mammal species and analysed the relationship between body mass and reproductive output per year (combining offspring size, litter size and numbers of litters per year). Overall, there was a clear allometric relationship. For the dataset as a whole, reproductive output declined with increasing body mass with a slope of −0.28 – a value between the two theoretical values of b −1 of −0.25 and −0.33 that we discussed as part of the metabolic theory of ecology in Section 3.9. However, there was considerable variation around this overall relationship – the fitted line accounted for only 59% of the overall variation – and it is clear from Figure 7.29a that phylogeny accounts for much of the remainder. In fact, when separate lines were fitted to the different mammalian orders, differences in their intercepts – their vertical location on the plot – were most important, rather than their slopes. A statistical model allowing intercepts to differ accounted for 72% of the variation, whereas allowing the slopes also to vary only shifted this up to 74%.

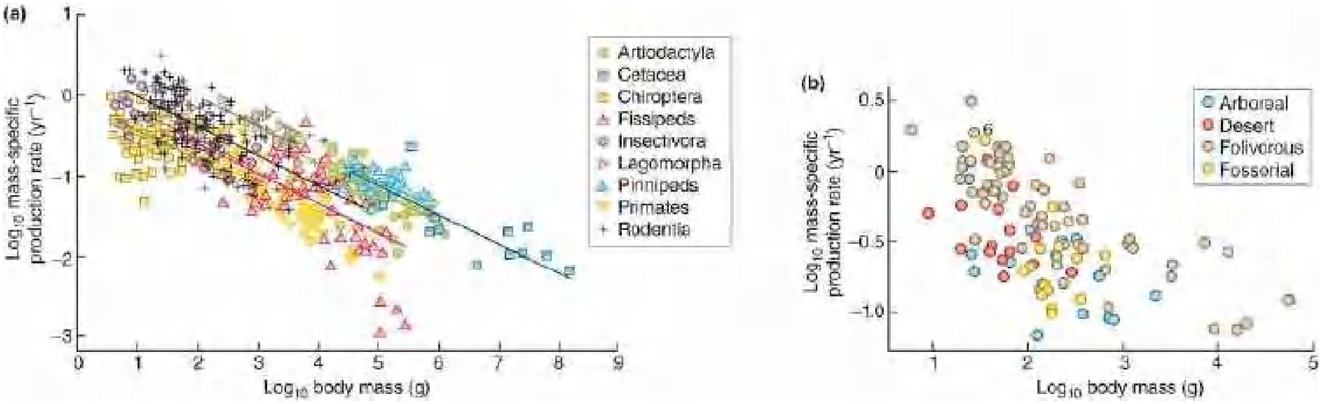


Figure 7.29 Allometry, phylogeny and the fast–slow continuum combine to make sense of life history variation in mammals. (a) A plot, on log scales, showing the allometric relationship in various groups of placental mammals between mass-specific production rate per year (reproductive output per unit body mass, combining offspring size and number and number of litters) and body mass. Relationships shared the best-fit slope for the whole dataset, −0.28, but varied in their intercepts. For clarity, an original dataset of 637 species has been condensed to include only those groups with more than 10 species in the dataset, as indicated. (b) A similar plot for rodent species only, divided into those that are arboreal (tree-dwelling), desert, folivorous (leaf-eating) and fossorial (living underground).

Source: After Sibly & Brown (2007).

In particular, taxa with higher intercepts included leaf-eating orders like lagomorphs (rabbits, hares and pikas) and artiodactyls (sheep, cattle, pigs, deer, etc.) and also marine mammals such as the cetacea (whales, porpoises and dolphins) and pinnipeds (seals and sea lions) (Figure 7.29a). Sibly and Brown argue that these higher rates of production for a given body size are possible because the leaf-eaters have evolved the means to ingest and digest an abundant and generally reliable source of nutrition in green vegetation, and similarly, the whales and seals, by invading the marine environment, have access to a rich supply of invertebrates and fish. By contrast, taxa with lower intercepts include bats and primates. In this case, Sibly and Brown argue that the animals have evolved to occupy aerial or arboreal habitats where predation and hence mortality rates are low, and that they therefore divert more resources to promoting their survival in a crowded environment and less to the production of offspring.

Such arguments imply that the differences between the groups have as much to do with differences in lifestyle as with taxonomic difference *per se*, and this is supported by plots like that in Figure 7.29b, which separate rodents on the basis of their lifestyle, rather than any taxonomic grouping. We see explicitly in the figure that species living in trees, or underground, or in deserts, with relatively little predation, have lower production rates than their folivorous relatives. This led Sibly and Brown to suggest that the vertical axis in Figure 7.29a itself represents the fast–slow continuum, linking back to earlier sections of this chapter. And this in turn reinforces the idea that the importance of phylogenetic linkage, and the shared morphology that this inevitably implies, rests as much in shared lifestyle as in common life history traits. We can think of organisms as having traits appropriate to the habitats they occupy. Or we can think of organisms occupying habitats appropriate to the traits they inherit.

Chapter 8
Interspecific Competition

8.1 Introduction

Having been introduced to *intra*specific competition in Chapter 5, we can readily understand that *inter*specific competition occurs when individuals of one species suffer a reduction in fecundity, survivorship or growth as a result of either exploitation of resources or interference by individuals from *another* species. These individual effects are likely to influence the population dynamics and distributions of the competing species, which go to determine the compositions of the communities of which they are part. But the individual effects will also influence a species' evolution, which, in turn, can influence the species' distributions and dynamics.

two separate questions – the possible and actual consequences of competition

This chapter, then, is about both the ecological and the evolutionary effects of interspecific competition on individuals and on populations. In Chapter 16, we go on to examine its role (along with those of predation and parasitism) in shaping the structure of ecological communities. But the present chapter also addresses a more general issue in ecology and indeed in science – that there is a difference between what a process can do and what it *does* do: a difference between what, in this case, interspecific competition is capable of doing and what it actually does in practice. These are two separate questions, and we must be careful to keep them separate.

The way these different questions can be asked and answered will be different, too. To find out what interspecific competition is capable of doing is relatively easy. Species can be forced to compete in experiments, or mathematical models of competition can be explored, or they can be examined in nature in pairs or groups chosen precisely because they seem most likely to compete. But it is much more difficult to discover how important interspecific competition actually is. It will be necessary to ask how realistic our experiments are, and whether the pairs and groups of species we choose to study are really typical of pairs and groups more generally.

The chronology of ecologists' investigation of interspecific competition is perhaps surprising. It began (least surprisingly) with descriptive observations of natural systems, but in the 1920s and 1930s was hugely influenced by the mathematical models of Alfred Lotka, working in the USA, and Vito Volterra, working in Italy (see Section 8.4), and in the 1930s by the laboratory work on protists in test tubes of the Russian ecologist Georgyi Gause, who may be said to have given the Lotka–Volterra models 'experimental teeth' (see Kingsland (1985) for a history of the field). Since then, there have been other models, other controlled experiments in artificial environments, and certainly many more quantitative and even experimental (manipulative) studies in the field. However, rather than trying to recreate here the way in which the topic has developed, we begin with a series of real examples from both natural and artificial environments (Section 8.2) and see what general principles emerge from these (Section 8.3). Then, in the following sections, we examine the Lotka–Volterra and other mathematical models of interspecific competition, before exploring what lessons can be learned from these, and how these apply – and sometimes do not apply – in the real world.

8.2 Some examples of interspecific competition

8.2.1 Competition among phytoplankton species for phosphorus

more efficient exploiters exclude less efficient ones

Competition was investigated in the laboratory between five single-celled freshwater phytoplankton species, competed together in pairs: *Chlorella vulgaris*, *Selenastrum capricornutum*, *Monoraphidium griffithii*, *Monodus subterraneus* and *Synechocystis* sp., all of which require phosphorus as an essential resource for their growth. The population densities of the different species were monitored over time as resources were continuously being added to the liquid medium, but their impact on their limiting resource (phosphorus) was also recorded. When any of the species was grown alone, it established a steady population density, reducing the phosphorus to a constant low concentration (Figure 8.1a). However, there was a clear rank order in the steady-state concentration to which phosphorus was reduced: *Synechocystis* < *Chlorella* < *Monoraphidium* < *Selenastrum* < *Monodus*. Then, when any two species were grown together, there was only one survivor – and that survivor was whichever species had previously reduced phosphorus to the lower level (Figure 8.1b).

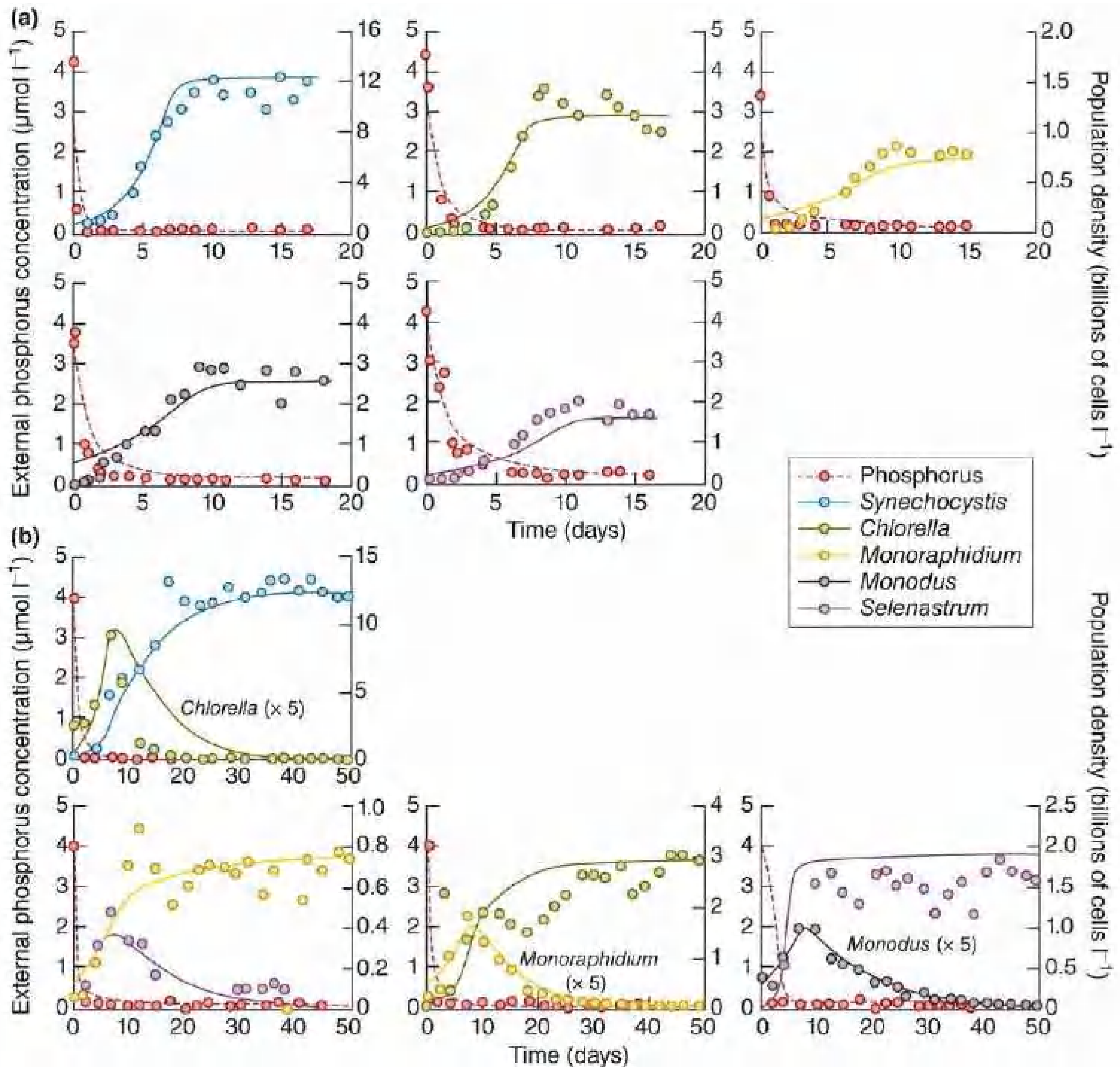


Figure 8.1 Competition between phytoplankton species for phosphorus: winners are those that reduce phosphorus concentrations to lower levels. (a) Each of five species (names and symbols shown in panel), when grown alone in a laboratory chemostat, establishes a stable population and maintains a resource, phosphorus (red dots), at a constant low level: *Synechocystis* (0.030 μmol l^{-1}) < *Chlorella* (0.059 μmol l^{-1}) < *Monoraphidium* (0.117 μmol l^{-1}) < *Selenastrum* (0.160 μmol l^{-1}) < *Monodus* (0.182 μmol l^{-1}). (b) With symbols retained from (a), it is apparent that when species are grown together in pairs, the only one to survive is the one that had previously reduced phosphorus to the lower level. (Densities have been multiplied by five in the cases indicated, so they can be visualised in the same figure. The lines are drawn simply to indicate general trends.)

Source: After Passarge *et al.* 2006).

Thus, although all species were capable of living alone in the laboratory habitat, when they competed, one species always excluded the other, because it was the more effective exploiter of their shared, limiting resource, reducing it to a level too low for the other species to survive. Indeed, a review of competition experiments with bacteria, phytoplankton and zooplankton found that this pattern – the species that reduces a limiting resource to the lowest level is the superior competitor – has been supported in almost every one of a substantial number of cases (Wilson *et al.*, 2007).

8.2.2 Competition among plant species for nitrogen

more efficient exploiters exclude less efficient ones

In other habitats and taxa, it is often not so easy to carry out the appropriate studies. However, one good example, with terrestrial plants competing for nitrogen, comes from the work of Tilman (Tilman & Wedin, 1991a , 1991b), who played the main role in elaborating this idea of superior competitors reducing resources to lower levels (see Tilman, 1982 ; but see MacArthur (1972) for its origin). Five grass species were grown alone in experimental conditions that gave rise in turn to a range of nitrogen concentrations. Two species, *Schizachyrium scoparium* and *Andropogon gerardi*, consistently reduced the nitrate and ammonium concentrations in soil solutions to lower values than those achieved by the other three species; and of these three, one, *Agrostis scabra*, left behind higher concentrations than the other two, *Agropyron repens* and *Poa pratensis*. Then, when *A. scabra* was grown with *A. repens*, *S. scoparium* and *A. gerardi*, the results, especially at low nitrogen concentrations where nitrogen was most likely to be limiting, were very much in line with the pattern we've already seen (Figure 8.2). The species that could reduce nitrogen to the lowest concentration always won; *A. scabra* was always competitively displaced.

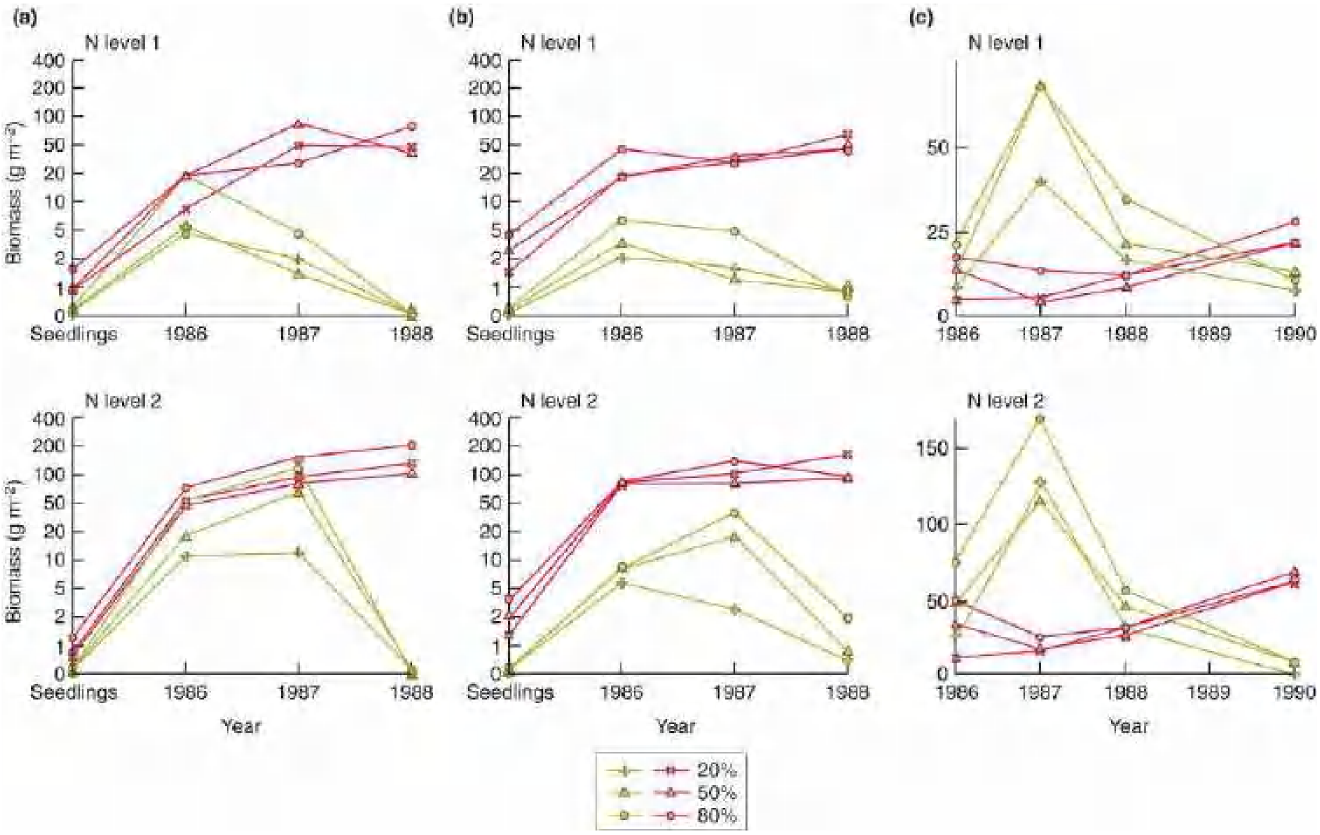


Figure 8.2 **In competition between grass species, the winner was the one that reduced nitrogen concentrations to lower levels.** The results of competition experiments in which *Agrostis scabra* (green lines) was competitively displaced by (a) *Schizachyrium scoparium*, (b) *Andropogon gerardi* and (c) *Agropyron repens* (red lines), at each of two nitrogen levels (both low). This occurred whether *A. scabra* represented 20%, 50% or 80% of the initial seed sown. In each case, *A. scabra* reduced nitrate and ammonium to lower concentrations (see text). Displacement was least rapid in (c) where the differential in concentration was least marked.

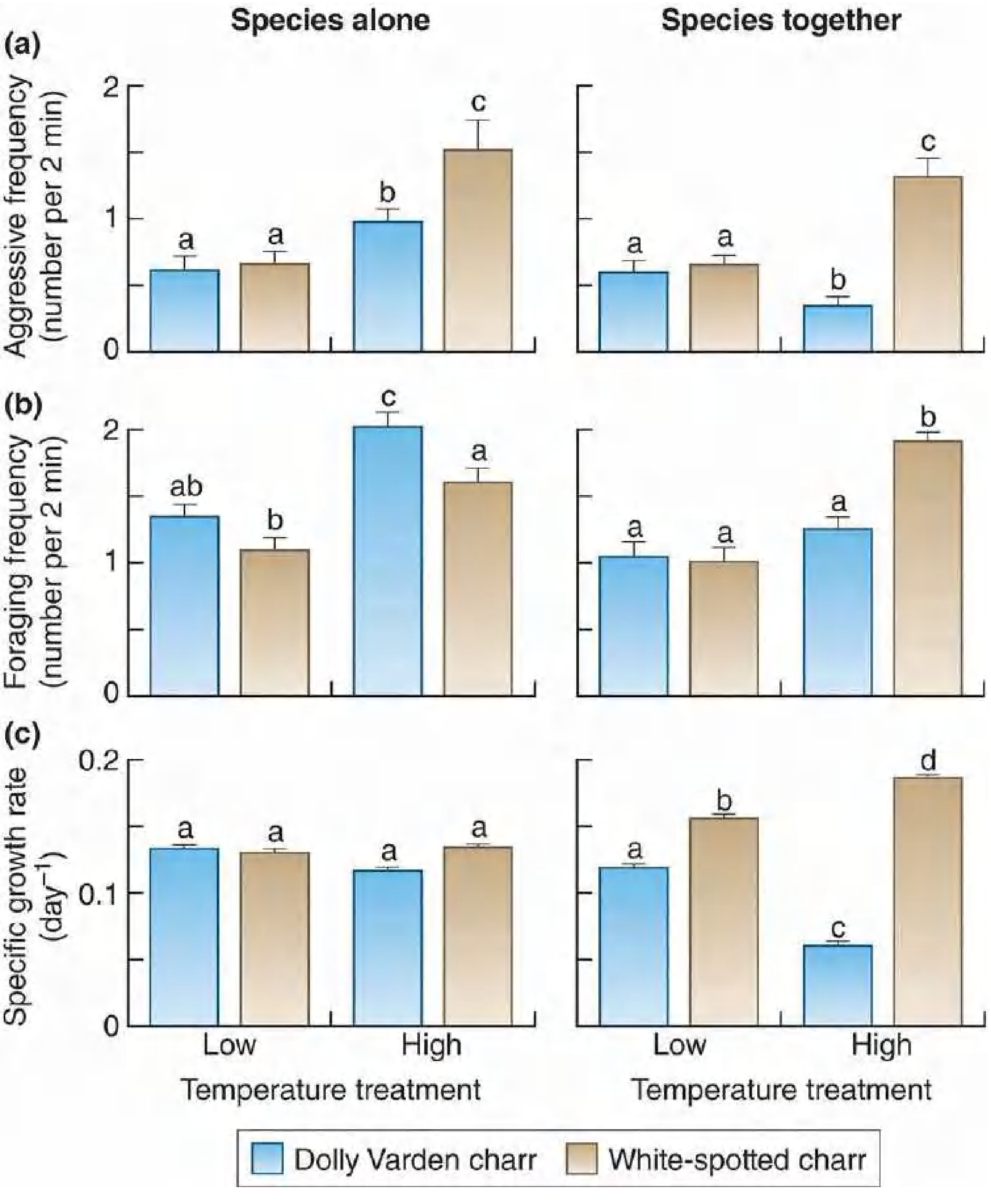
Source: After Tilman & Wedin (1991b).

A similar result has been obtained, too, for the nocturnal, insectivorous gecko *Hemidactylus frenatus*, an invader of urban habitats across the Pacific basin, where it is responsible for population declines of the native gecko *Lepidodactylus lugubris* (Petren & Case, 1996). The diets of the two geckos overlap substantially and insects are a limiting resource for both. The invader is capable of depleting insect resources in experimental enclosures to lower levels than the native gecko, and the native suffers reductions in body condition, fecundity and survivorship as a result.

8.2.3 Coexistence and exclusion of competing salmonid fishes

The Dolly Varden charr (*Salvelinus malma*) and the white-spotted charr (*S. leucomaenis*) are two morphologically similar and closely related species of salmonid fish found in many streams on Hokkaido Island, Japan. The Dolly Varden are distributed further upstream than the white-spotted charr. Water temperature, which has profound consequences for fish ecology, increases downstream. In streams where they occur together, there is a zone of overlap at intermediate altitudes.

To try to understand these distributions, Taniguchi and Nakano (2000) carried out experiments in laboratory streams. Higher temperatures (12°C as compared with 6°C) led to increased aggression in both species when they were tested alone, but these temperatures led to the Dolly Varden being *less* aggressive when white-spotted charr were also present (Figure 8.3a) and hence enhancing its competitive ability relative to that of the Dolly Varden. Consequently, Dolly Varden charr failed to obtain favourable foraging positions when white-spotted charr were present at the higher temperature and so foraged far less effectively (Figure 8.3b). Also, when alone, neither species' growth rates were influenced by temperature. But when both species were present, growth of Dolly Varden charr decreased with increasing temperature, whereas that of white-spotted charr increased (Figure 8.3c). Thus, the growth rate of the Dolly Varden was much lower than that of white-spotted charr at the higher temperature.


Species alone
Species together
(a)
Aggressive frequency (number per 2 min)
2
1
0
a
a
b
c
a
a
b
c
(b)
Foraging frequency (number per 2 min)
2
1
0
ab
b
c
a
a
a
a
b
(c)
Specific growth rate (day^{-1})
0.2
0.1
0
a
a
a
a
a
b
c
d
Low
High
Temperature treatment
Low
High
Temperature treatment
Dolly Varden charr
White-spotted charr

[Figure 8.3](#) **When together, white-spotted charr perform consistently better than Dolly Varden charr at higher ('upstream') temperatures.** (a) Frequency of aggressive encounters initiated by individuals of each of the two fish species (Dolly Varden charr, blue histograms, and white-spotted charr, brown histograms) during a 72-day experiment in artificial stream channels with two replicates each. Left, either 50 Dolly Varden or 50 white-spotted charr alone (allopatry); right: 25 of each species together (sympatry); (b) foraging frequency; and (c) specific growth rate in length. Different letters indicate means are significantly different from each other.

Source: After Taniguchi & Nakano (2000).

These results all support the idea that the lower altitudinal boundary of Dolly Varden charr in the Japanese streams was due to temperature-mediated competition favouring white-spotted charr, which were more aggressive, foraged more effectively, and grew far faster. The Dolly Varden, on the other hand, did not outcompete the white-spotted charr in any of the experiments, even at the lower temperatures. Nonetheless, further results in this study indicate that the survival was very poor in white-spotted charr at lower temperatures even in the absence of the Dolly Varden. Hence, it seems that the upstream limits of the white-spotted charr are determined simply by the temperatures there being too low for them.

8.2.4 Some general observations

These initial examples already illustrate several points of general importance.

1. Competing species often coexist at one spatial scale but are found to have distinct distributions at a finer scale of resolution. Here, the fishes coexisted in the same stream, but each was more or less confined to its own altitudinal zone (which had different temperatures).
2. Species are often excluded by interspecific competition from locations at which they could exist perfectly well in the absence of interspecific competition. Here, Dolly Varden charr can live in the white-spotted charr zone – but only when there are no white-spotted charr there. Similarly, *Chlorella* can live in laboratory cultures – but only when there were no *Synechocystis* there (and so on down the line).

   fundamental and realised niches

3. We can describe this by saying that the conditions and resources provided by the white-spotted charr zone are part of the *fundamental* niche of Dolly Varden charr (see Section 2.2 for an explanation of ecological niches) in that the basic requirements for the existence of Dolly Varden charr are provided there. But the white-spotted charr zone does not provide a *realised* niche for Dolly Varden when white-spotted charr are present. Likewise, the laboratory cultures provided the requirements of the fundamental niches of both *Chlorella* and *Synechocystis* (and the other three phytoplankton species), but when *Chlorella* and *Synechocystis* were both present, provided a realised niche for only *Synechocystis*.
4. Thus, a species' fundamental niche is the combination of conditions and resources that allow that species to exist, grow and reproduce when considered in isolation from any other species that might be harmful to its existence; whereas its realised niche is the combination of conditions and resources that allow it to exist, grow and reproduce in the presence of specified other species that might be harmful to its existence – especially interspecific competitors. Even in locations that provide a species with the requirements of its fundamental niche, that species may be excluded by another, superior competitor that denies it a realised niche there. But competing species can coexist when both are provided with a realised niche by their habitat. For the charr, the stream as a whole provided a realised niche for both species.

5. Indeed, the fish provide an example of a pattern that has been observed so frequently in studies of interspecific competition that it has been elevated to the status of a principle: the *competitive exclusion principle* or Gause's Principle (since his laboratory experiments were so influential in establishing it). It can be stated as follows:
    - If two competing species coexist in a stable environment, then they do so as a result of niche differentiation.
    - If, however, there is no such differentiation, or if it is precluded by the habitat, then one competing species will eliminate or exclude the other.

    In fact, further studies have generated versions of the Principle that extend beyond just two competing species, contending that *n* competing species require at least *n* separate niches (or *n* different resources) for their coexistence (see Armstrong & McGehee, 1980 .) We will see additional apparent examples of the Principle in operation below, and we will also see how the Principle emerges from simple mathematical model of interspecific competition (Section 8.4.1). But as we revisit the Principle, it will be as important to understand what it does *not* say as what it does – and that it by no means tells us the whole story about interspecific competition and coexistence.
6. Finally, the fish study also illustrates another point that will recur throughout this chapter, and throughout this book: the importance of experimental manipulation if we wish to discover what is really going on in a natural population – 'nature' may need to be prodded to reveal its secrets.

8.2.5 Coexistence of competing diatoms

Returning to our set of examples, another experimental study of competition amongst single-celled phytoplankton (in this case diatoms) looked at species coexisting on not one but two shared, limiting resources. The two species were *Asterionella formosa* and *Cyclotella meneghiniana*, and the resources, which were both capable of limiting the growth of both diatoms, were silicate and phosphate. *Cyclotella* was the more effective exploiter of silicate (reducing its concentration to a lower level), but *Asterionella* was the more effective exploiter of phosphate. Hence, in cultures where there were especially low supplies of silicate, *Cyclotella* excluded *Asterionella* (Figure 8.4): such cultures failed to provide a realised niche for *Asterionella*, the inferior competitor there. Conversely, in cultures where there were especially low supplies of phosphate, *Asterionella* excluded *Cyclotella*. However, in cultures with relatively balanced supplies of silicate and phosphate, the two diatoms coexisted (Figure 8.4): when both species were provided with sufficient supplies of a resource on which they were inferior, there was a realised niche for both.

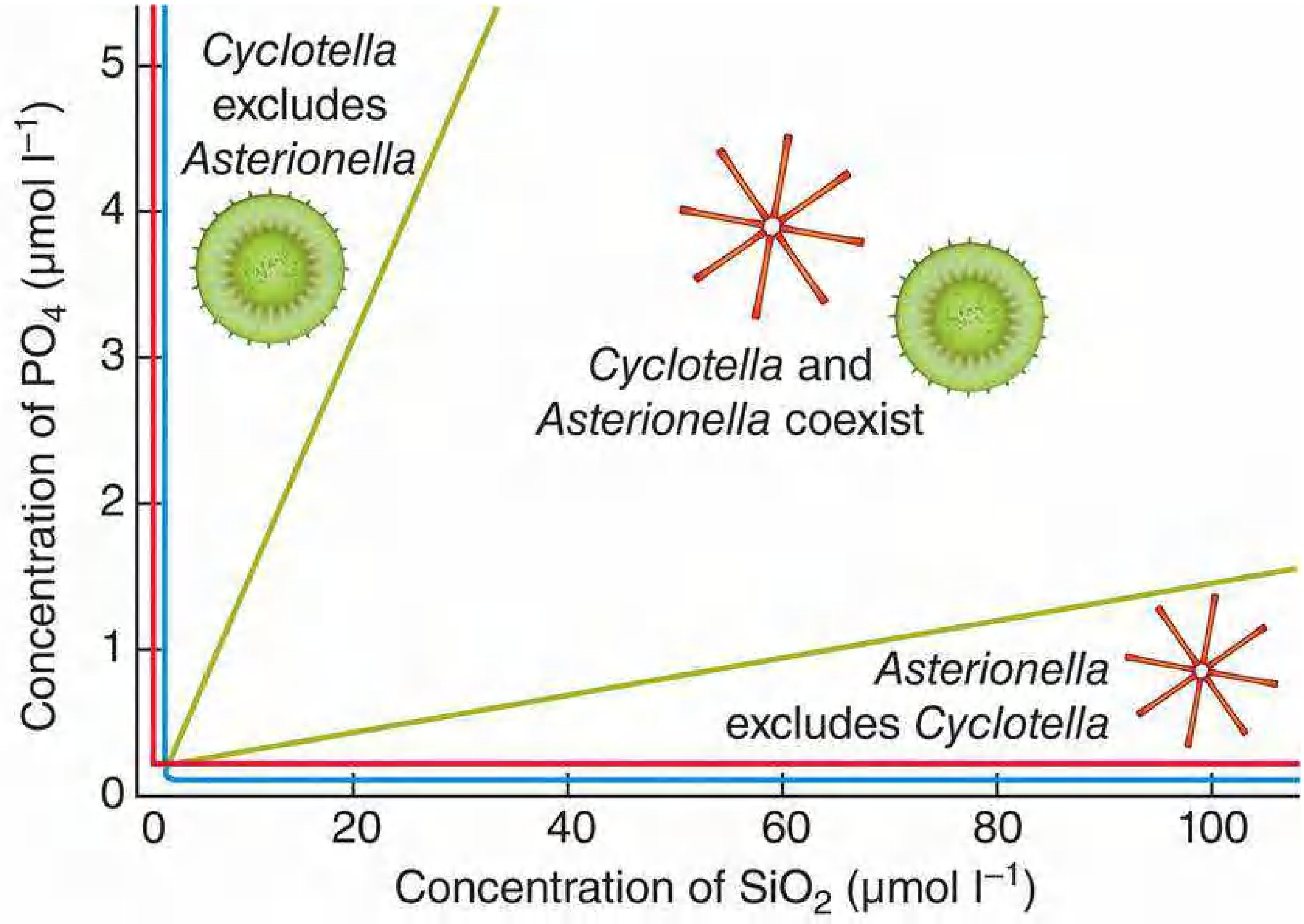


Figure 8.4 When two diatom species compete for two resources, each persists only in its realised niche. *Asterionella formosa* and *Cyclotella meneghiniana* coexist when there are roughly balanced supplies of silicate (SiO_2) and phosphate (PO_4), but *Asterionella* excludes *Cyclotella* when there are especially low supplies of phosphate, whereas *Cyclotella* excludes *Asterionella* when there are especially low supplies of silicate. Hence the red lines define the fundamental niche for *Cyclotella*, and the blue lines define the fundamental niche for *Asterionella*; while the lower green and vertical red lines define the realised niche for *Cyclotella*, and the upper green and horizontal blue lines define the realised niche for *Asterionella*.

Source: After Tilman (1982).

8.2.6 Coexistence of competing birds

It is not always so easy to identify the *niche differentiation* (i.e. differentiation of their realised niches) or the 'differential resource utilisation' that allows competitors to coexist. Ornithologists, for example, are well aware that closely related species of birds often coexist in the same habitat. For example, four *Sylvia* species (warblers) occur together on the Mediterranean islands of Corsica and Sardinia: Marmora's warbler (*Sylvia sarda*), the Dartford warbler (*S. undata*), the Sardinian warbler (*S. melanocephala*), and the subalpine warbler (*S. cantillans*). All look similar; all are insectivorous, with just the occasional use of berries or soft fruit; and all live in open, bush-dominated landscapes, using the bushes there for nesting. Yet, the closer we look at the details of the ecology of such coexisting species, the more likely we are to find ecological differences. In this case, for example, all four species are found in habitats with bushes of medium height, but where the bushes are smaller, Marmora's warblers are most often observed but subalpine warblers are largely absent, whereas the reverse is true where bushes are taller. And even within the particular habitat types, the species differ significantly in the different plant species and plant heights from which they prefer to forage (Martin & Thibault, 1996). Hence, it is tempting to conclude that such species compete but coexist by utilising slightly different resources in slightly different ways:

differential resource utilisation. In complex natural environments, however, such conclusions, while plausible, are difficult to prove.

coexistence through niche differentiation – and even competition – may be difficult to prove

Indeed, it is often not easy to prove even that the species compete. To do so, it is usually necessary to remove one or more of the species and monitor the responses of those that remain. This was done, for example, in a study of two very similar bird species, again warblers: the orange-crowned warbler (*Vermivora celata*) and the Virginia's warbler (*V. virginiae*), whose breeding territories overlap in central Arizona. On plots where one of the two species had been removed, the remaining species fledged between 78% and 129% more young per nest (Figure 8.5a). The enhanced performance was largely due to improved access to preferred nest sites and consequent decreases in the loss of nestlings to predators (Figure 8.5b). The typical effects of interspecific competition are revealed by experimentally eliminating that competition.

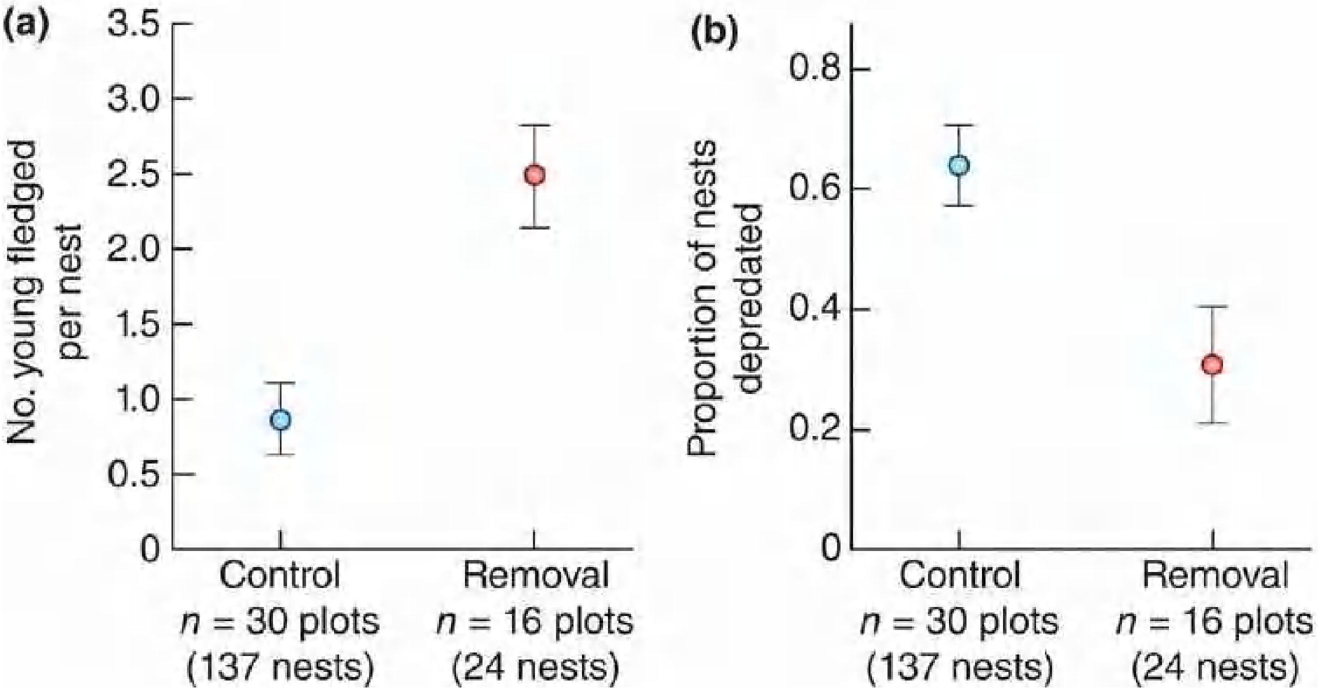


Figure 8.5 Warblers protected from interspecific competition fledge more young from their nests. Competition between orange-crowned warblers and Virginia's warblers, examined through a comparison of control plots (both species present) and plots from which one or other of the species was removed (males removed before pairing such that females neither settle nor nest). (a) The number of young fledged from nests (means ± SEs) was significantly higher ($P = 0.0004$) when one species was removed (data from both species combined). (b) The proportion of nests in which all offspring were taken by predators (means ± SEs) was significantly lower ($P = 0.002$) when one species was removed (data from both species combined).

Source: After Martin & Martin (2001).

8.2.7 Competition between unrelated species

The examples described so far have all involved pairs or groups of closely related species – phytoplankton, salmonid fish or birds. But competition may also occur between completely unrelated species, as shown for two examples in Figure 8.6. In the first (Figure 8.6a), removing sea urchins from experimental plots in the coastal waters of St Thomas, US Virgin Islands, increased the abundance of both parrotfish and surgeonfish, with which the sea urchins compete for their shared seaweed food (although one year later, the urchins had returned, and levels of food and fish abundance had returned to control levels). In the second example (Figure 8.6b), allowing insects to compete at natural densities with tadpoles of the Pine Barrens tree frog demonstrated that the competition from this very distantly related group was comparable in intensity to that provided by the much more closely related tadpoles of Fowler's toads. Close relatives are most similar, and hence, all other things being equal, more likely to compete. But clearly, interspecific competition can also occur across huge taxonomic divides.

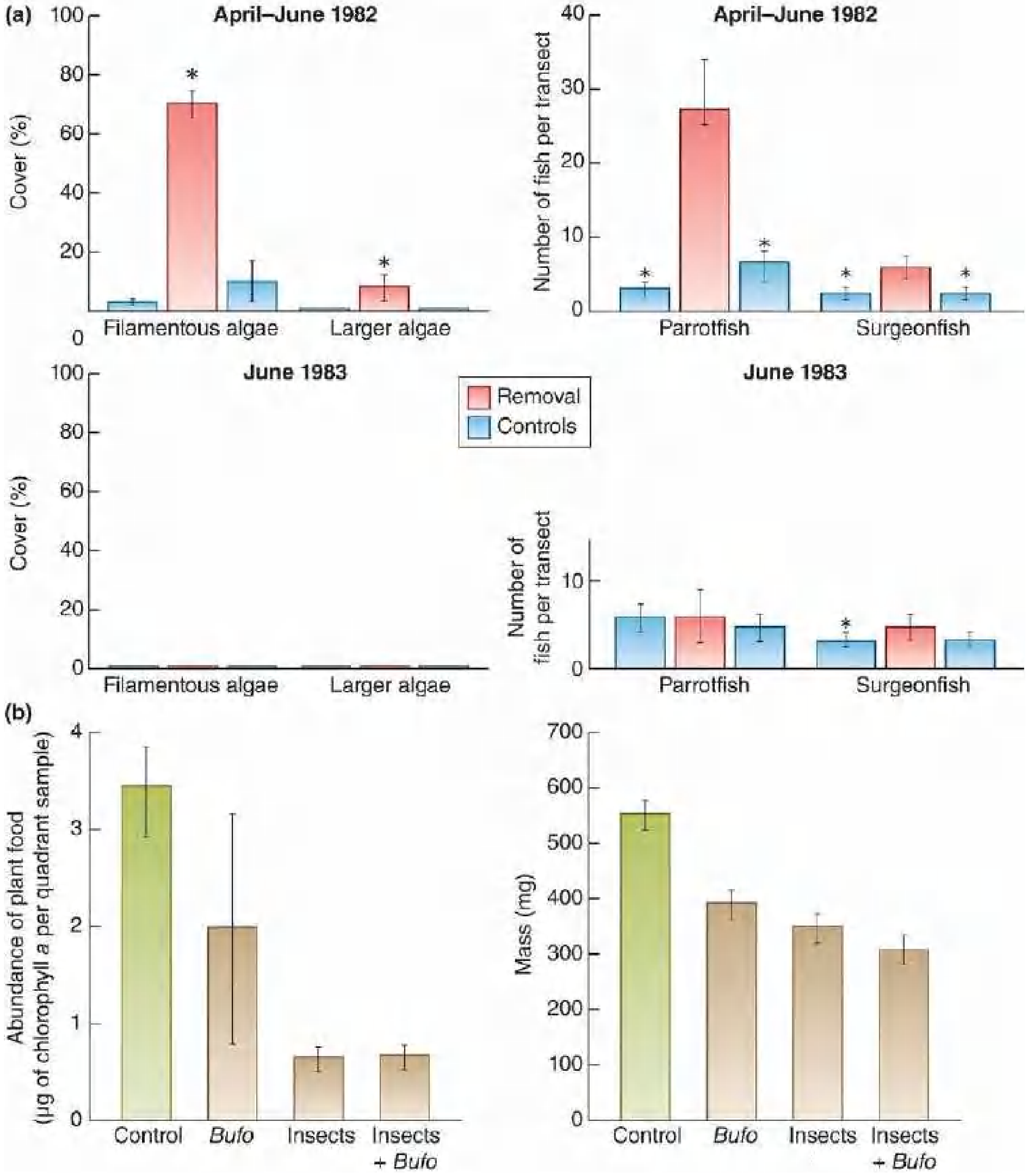

(a)
April–June 1982
Cover (%)
100
80
60
40
20
0
*
*
Filamentous algae
Larger algae
April–June 1982
Number of fish per transect
40
30
20
10
0
*
*
*
*
Parrotfish
Surgeonfish
June 1983
Cover (%)
100
80
60
40
20
0
Filamentous algae
Larger algae
🟥 Removal
🟦 Controls
June 1983
Number of fish per transect
10
0
*
Parrotfish
Surgeonfish
(b)
Abundance of plant food (µg of chlorophyll a per quadrant sample)
4
3
2
1
0
Control
Bufo
Insects
Insects + Bufo
Mass (mg)
700
600
500
400
300
200
100
0
Control
Bufo
Insects
Insects + Bufo

Figure 8.6 Competition between unrelated species: sea urchins with fish, and tadpoles with insects. (a) Competition between sea urchins and fish. When sea urchins, *Diadema antillarum*, were removed from experimental plots in St Thomas, US Virgin Islands, in February 1982, two months later amounts of their seaweed food had significantly increased (left panel) as had numbers of parrotfish and surgeonfish, which share the food resource with the urchins (right panel). But one year later urchins had reinvaded and reached 70% of their original abundance, and food and numbers of fish had reverted to control levels. Bars are 2 SEs. (b) Competition between insects and tadpoles. Tadpoles of the Pine Barrens tree frog, *Hyla andersonii*, were allowed to develop and metamorphose either alone in covered experimental tanks ('Control'), or in covered tanks to which tadpoles of Fowler's toad, *Bufo woodhousei fowleri* had been added ('*Bufo*'), or in uncovered tanks which a natural assemblage of insects could therefore colonise ('Insects'), or in uncovered tanks to which *Bufo* had been added ('Insects + *Bufo*'). The presence of *Bufo* reduced the abundance of the *Hyla* tadpoles' plant food (left panel) but the insects (and the insects and *Bufo* together) did so even more. Consequently (right panel) the *Hyla* tadpoles suffered more from competition with insects than they did from competition from *Bufo*, in terms of their mean mass at metamorphosis to froglets, and even more from the two combined. Bars are SEs.

Source: (a) After Hay & Taylor (1985). (b) After Morin *et al*. (1988).

8.3 Some general features of interspecific competition – and some warnings

8.3.1 Unravelling ecological and evolutionary aspects of competition

This series of examples has confirmed that individuals of different species can compete. This is hardly surprising. The field experiments with warblers and with sea urchins and fish, for example, have also confirmed that different species *do* compete in nature (i.e. there was a measurable interspecific reduction in abundance and/or fecundity and/or survivorship). We have also seen further apparent examples of the competitive exclusion principle in operation: that competing species may either exclude one another from particular habitats so that they do not coexist, or may coexist by utilising the habitat in slightly different ways.

But what about the coexisting Mediterranean warblers? We saw that the four species coexist and utilise the habitat in slightly different ways. But does this have anything to do with competition? It may do. Indeed, it may be that even if the species do not compete now, they coexist as a result of evolutionary responses to interspecific competition in the past. To understand how this could be the case, we note that when two species compete, individuals of one or both species may suffer reductions in fecundity and/or survivorship, as we have seen. The fittest individuals of each species may then be those that (relatively speaking) escape competition because they utilise the habitat in ways that differ most from those adopted by individuals of the other species. Natural selection will then favour such individuals, and eventually the population may consist entirely of them. The two species will evolve to become more different from one another than they were previously; they will compete less, and thus will be more likely to coexist.

coexisting competitors or the 'ghost of competition past'? ...

The trouble with this as an explanation for the warbler story is that there is no proof. We need to beware, in Connell's (1980) phrase, of uncritically invoking the 'ghost of competition past'. We cannot go back in time to check whether the species ever competed more than they do now. A plausible alternative interpretation is that the species have, in the course of their evolution, responded to natural selection in different but entirely independent ways. They are distinct species, and they have distinctive features. But they do not compete now, nor have they ever competed; they simply happen to be different. If all this were true, then the coexistence of the warblers would have nothing to do with competition. Alternatively again, it may be that competition in the past eliminated a number of other species, leaving behind only those that are different in their utilisation of the habitat: we can still see the hand of the ghost of competition past, but acting as an ecological force (eliminating species) rather than an evolutionary one (changing them).

The warbler story, therefore, and the difficulties with it, illustrate two important general points. The first is that we must pay careful, and separate, attention to both the ecological and the evolutionary effects of interspecific competition. The ecological effects are, broadly, that species may be eliminated from a habitat by competition from individuals of other species; or, if competing species coexist, that individuals of at least one species suffer reductions in survival and/or fecundity. The evolutionary effects appear to be that species differ more from one another than they would otherwise do, and hence compete less (but see Section 8.8).

... or simply evolution?

The second point, though, is that there are profound difficulties in invoking competition as an explanation for observed patterns, and especially in invoking it as an evolutionary explanation. An experimental manipulation (for instance, the removal of one or more species) can, as we have seen, indicate the presence of current competition if it leads to an increase in the fecundity or survival or abundance of the remaining species. But negative results would be equally compatible with the past elimination of species by competition, with the evolutionary avoidance of competition in the past, or with the independent evolution of noncompeting species. In fact, for many sets of data, there are no easy or agreed methods of distinguishing between these explanations (see also Chapter 16). Thus, in the remainder of this chapter when examining the ecological and, especially, the evolutionary effects of competition, we will need to be more than usually cautious.

8.3.2 A further warning: coexistence without niche differentiation?

It is tempting to write about competition, coexistence and exclusion by compiling a set of examples that all support a common narrative: namely, species compete with one another, and when competing species coexist, they do so as a result of niche differentiation. But ecologists have known for a very long time that there are studies that seem to run counter to this narrative and so call it into question. Probably the earliest, and certainly the most iconic example is to be found in Hutchinson's (1961) article 'The paradox of the plankton' – the paradox being that dozens of planktonic diatoms can coexist in a simple body of water with little apparent potential for niche differentiation: not enough different limiting resources or niche dimensions (Figure 8.7). Hutchinson did not claim that datasets like these prove the inapplicability of the competitive exclusion principle, and neither should we. But he identified a challenge that we should recognise, namely that our efforts to understand the role of interspecific competition and species coexistence in shaping ecological communities must go beyond simple species pairs to account for the coexistence of large numbers of apparently similar species in much more complex communities.

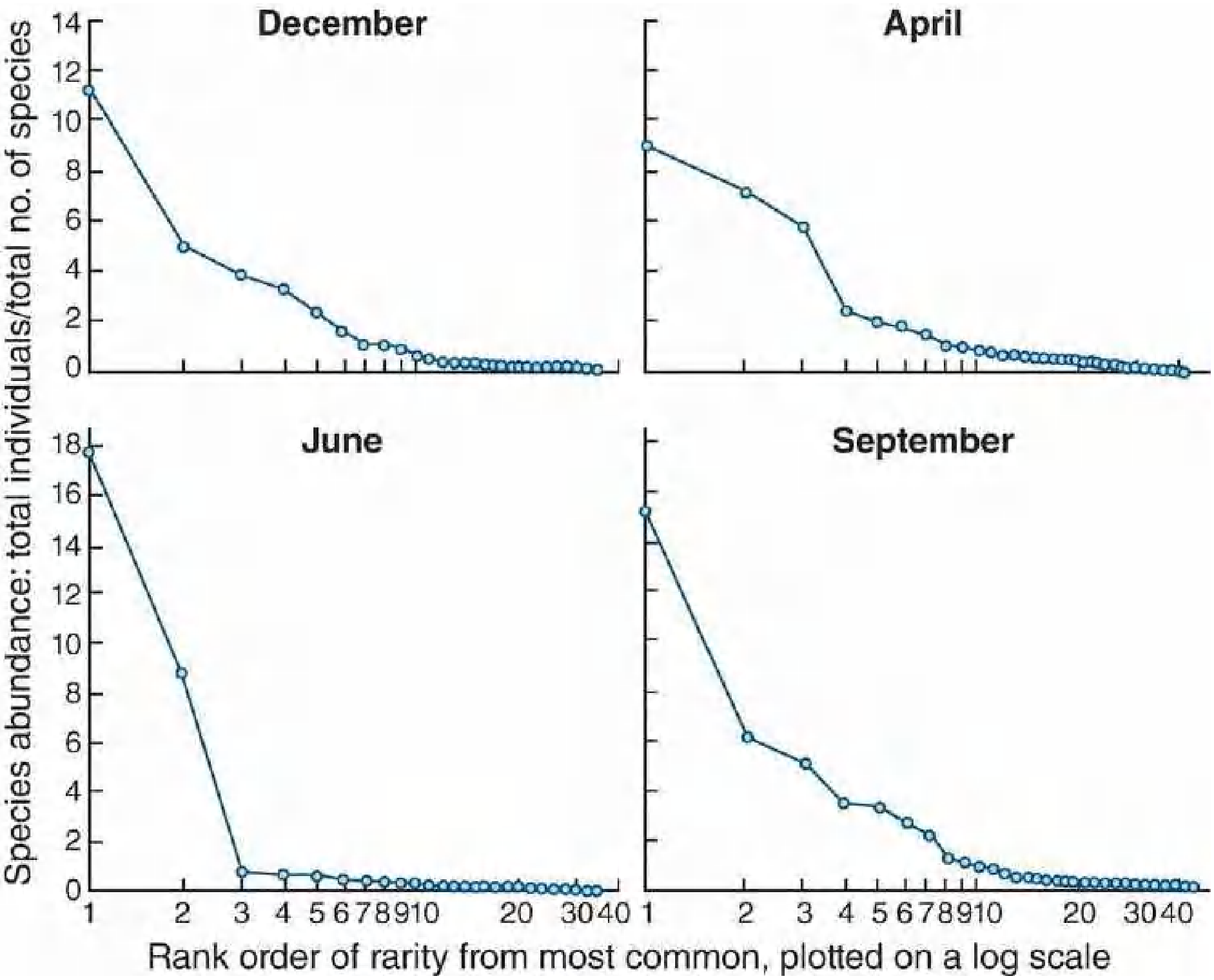


Figure 8.7 **The paradox of the plankton.** The influential dataset of planktonic diatoms from a lake in Denmark that Hutchinson used to challenge ecologists to explain the coexistence, in some cases, of large numbers of apparently similar species in apparently simple habitats. Not only were there consistently large numbers of species, throughout the year; the common species were commoner and the rare species rarer than normally expected when there is competitive exclusion based on niche differentiation.

Source: After Hutchinson (1961).

8.3.3 Exploitation and interference competition and allelopathy

interference and exploitation

As with intraspecific competition, a basic distinction can be made between interference and exploitation competition. With exploitation, individuals interact with each other indirectly, responding to a resource level that has been depressed by the activity of competitors. The phytoplankton species in Figure 8.1 provided a clear example of this. By contrast, the two species of charr provided an equally clear example of interference competition – the Dolly Varden charr, in particular, failing to obtain suitable foraging positions in the face of aggressive interactions with the white-spotted charr.

allelopathy

Interference, on the other hand, is not always as direct as this. Amongst plants, it has often been claimed that interference occurs through the production and release into the environment of chemicals that are toxic to other species but not to the producer (known as allelopathy). There is no doubt that chemicals with such properties can be extracted from plants, but establishing a role for them in nature, or that they have evolved *because of* their allelopathic effects, has proved difficult. Indeed, it is increasingly apparent that such chemicals often have other roles, too, including defence against herbivores, or in mediating decomposition or nutrient acquisition and cycling (Inderjit *et al*., 2011 .)

There are, though, certainly cases where the allelopathic effects on competitors of other species are apparent. In one interesting example, competition was studied off the coast of Fiji between the coral *Porites cylindrica* and the seaweeds *Sargassum polycystum* and *Galaxaura filimentosa*. Previous work had shown that *Galaxaura* produces terpenes that act to poison corals, whereas *Sargassum* has no such effects. In addition, though, Rasher and Hay (2014) showed not only that chemical extracts from *Galaxaura* adversely affected the growth of *Porites* (whereas those from *Sargassum* did not), but also that following eight days of competition with *Porites*, increased amounts of these allelochemicals were induced in *Galaxaura*, and the allelopathic effects on the coral were correspondingly greater, compared with treatments that were grown alongside coral skeleton rather than the living coral itself (Figure 8.8a). Moreover, the *Galaxaura* paid a price for producing these chemicals. The plants grown with *Porites* skeleton suffered relatively low levels of herbivory from surgeonfish (*Naso lituratus* and *N. unicornis*) – significantly lower than those suffered by the relatively unprotected *Sargassum* – as a result of the antiherbivore chemicals they produced. But the *Galaxaura* grown in competition with the live coral produced lower concentrations of these chemicals and were significantly more affected by the herbivores as a consequence (Figure 8.8b).

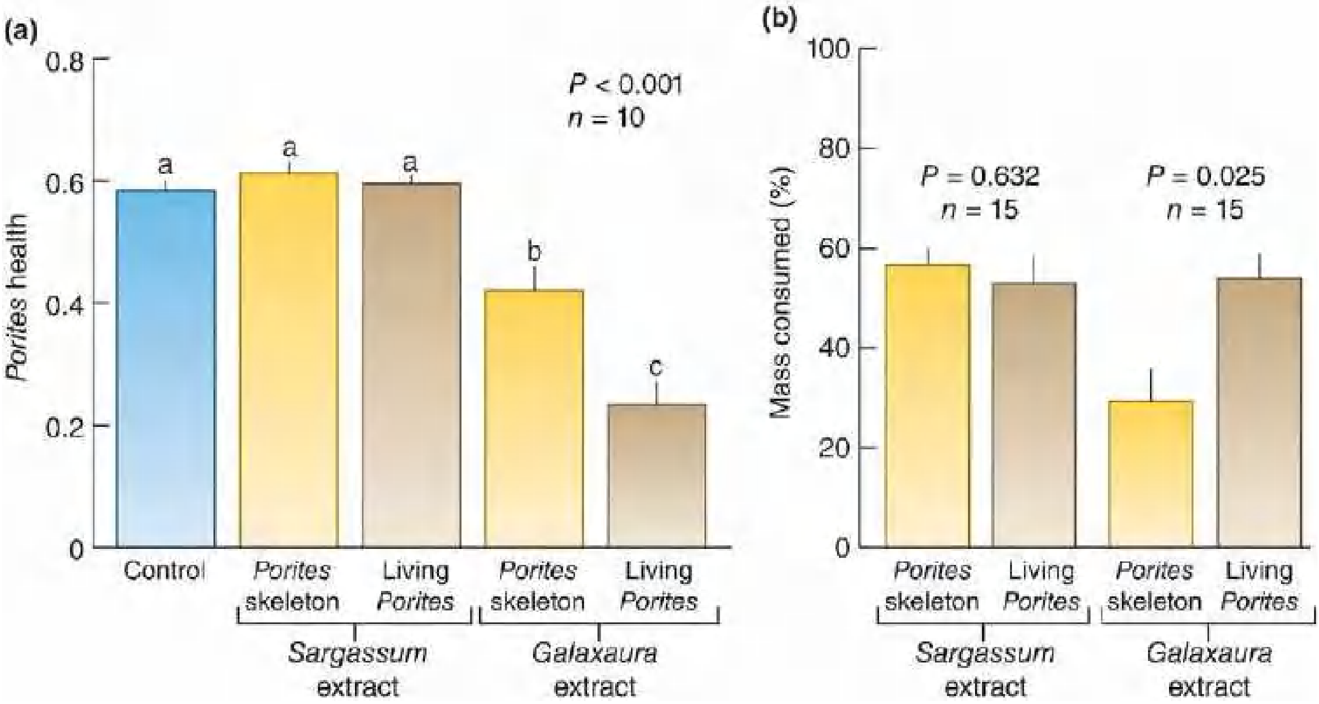


Figure 8.8 **Allelopathy – and its price in terms of increased predation – between seaweeds and coral.** (a) Effects of seaweed extracts (*Sargassum* or *Galaxaura*) on coral (*Porites*) health, where those seaweeds had previously been in contact with either live coral or coral skeleton. '*Porites* health' reflects the level of photosynthetic functionality, typically ranging from 0 to 0.83. Values above 0.5 are considered 'healthy'. Bars are SEs. Different letters indicate significant differences among treatments. (b) The price paid by *Galaxaura* for producing its allelochemicals: individuals doing so are as heavily attacked by surgeonfish as the relatively unprotected *Sargassum*. Bars are SEs.

Source: After Rasher & Hay (2014).

Amongst competing tadpole species, too, water-borne inhibitory products have been implicated as a means of interference (most notably, perhaps, an alga produced in the faeces of the common frog, *Rana temporaria*, inhibiting the natterjack toad, *Bufo calamita* (Beebee, 1991 ; Griffiths *et al*., 1993)), but here again their importance in nature is unclear (Petranka, 1989). Of course, the production by fungi and bacteria of allelopathic chemicals that inhibit the growth of potentially competing microorganisms is widely recognised – and exploited in the selection and production of antibiotics.

8.4 The Lotka–Volterra model of interspecific competition

Having surveyed a range of studies, we turn in this and following sections to a number of models of competition, which further clarify and extend some of the lessons we have learned. We begin here with the classic 'Lotka–Volterra' model (Volterra, 1926 ; Lotka, 1932).

8.4.1 The Lotka–Volterra model

The Lotka–Volterra model of interspecific competition is an extension of the logistic equation described in Section 5.7. As such, it incorporates all of the logistic's shortcomings, but a useful model can nonetheless be constructed, shedding light on factors that may determine the outcome of competitive interactions.

The logistic equation:

$$\frac{dN}{dt} = rN\frac{(K-N)}{K} \tag{8.1}$$

contains within the brackets a term responsible for the incorporation of intraspecific competition. The basis of the Lotka–Volterra model is the replacement of this term by one that incorporates both intra- and interspecific competition.

α: the competition coefficient

The population size of one species can be denoted by N_1, and that of a second species by N_2. Their carrying capacities and intrinsic rates of increase are K_1, K_2, r_1 and r_2, respectively. Suppose, then, that 10 individuals of species 2 have, between them, the same competitive, inhibitory effect on species 1 as does a single individual of species 1. The total competitive effect on species 1 (intra- and interspecific) will then be equivalent to the effect of ($N_1 + N_2/10$) species 1 individuals. The constant (1/10 in the present case) is called a competition coefficient and is denoted by α_{12} ('alpha-one-two'). It measures the per capita competitive effect on species 1 of species 2. Thus, $\alpha_{12} < 1$ means that individuals of species 2 have less inhibitory effect on individuals of species 1 than individuals of species 1 have on others of their own species, whilst $\alpha_{12} > 1$ means that individuals of species 2 have a greater inhibitory effect.

Lotka–Volterra model: a logistic model for two species

The crucial element in the model is the replacement of N_1 in the bracket of the logistic equation with a term capturing the inhibitory effects of species 1 and species 2, i.e.:

$$\frac{dN_1}{dt} = r_1N_1\frac{(K_1-(N_1+\alpha_{12}N_2))}{K_1} \tag{8.2}$$

or:

$$\frac{dN_1}{dt} = r_1N_1\frac{(K_1-N_1-\alpha_{12}N_2)}{K_1} \tag{8.3}$$

and in the case of the second species:

$$\frac{dN_2}{dt} = r_2N_2\frac{(K_2-N_2-\alpha_{21}N_1)}{K_2}. \tag{8.4}$$

These two equations constitute the Lotka–Volterra model.

behaviour of the Lotka–Volterra model is investigated using 'zero isoclines'

To appreciate the properties of this model, we must ask the question: under what circumstances does each species increase or decrease in abundance? In order to answer this, it is necessary to construct diagrams in which all possible combinations of species 1 and species 2 abundance can be displayed (i.e. all possible combinations of N_1 and N_2). These will be diagrams (Figures 8.9 and 8.10), with N_1 plotted on the horizontal axis and N_2 plotted on the vertical axis, such that there are low numbers of both species towards the bottom left, high numbers of both species towards the top right, and so on. Certain combinations of N_1 and N_2 will give rise to increases in species 1 and/or species 2, whilst other combinations will give rise to decreases in species 1 and/or species 2. Crucially, there must also therefore be ‘zero isoclines’ for each species (lines along which there is neither an increase nor a decrease), dividing the combinations leading to increase from those leading to decrease. Moreover, if a zero isocline is drawn first, there will be combinations leading to an increase on one side of it, and combinations leading to a decrease on the other.

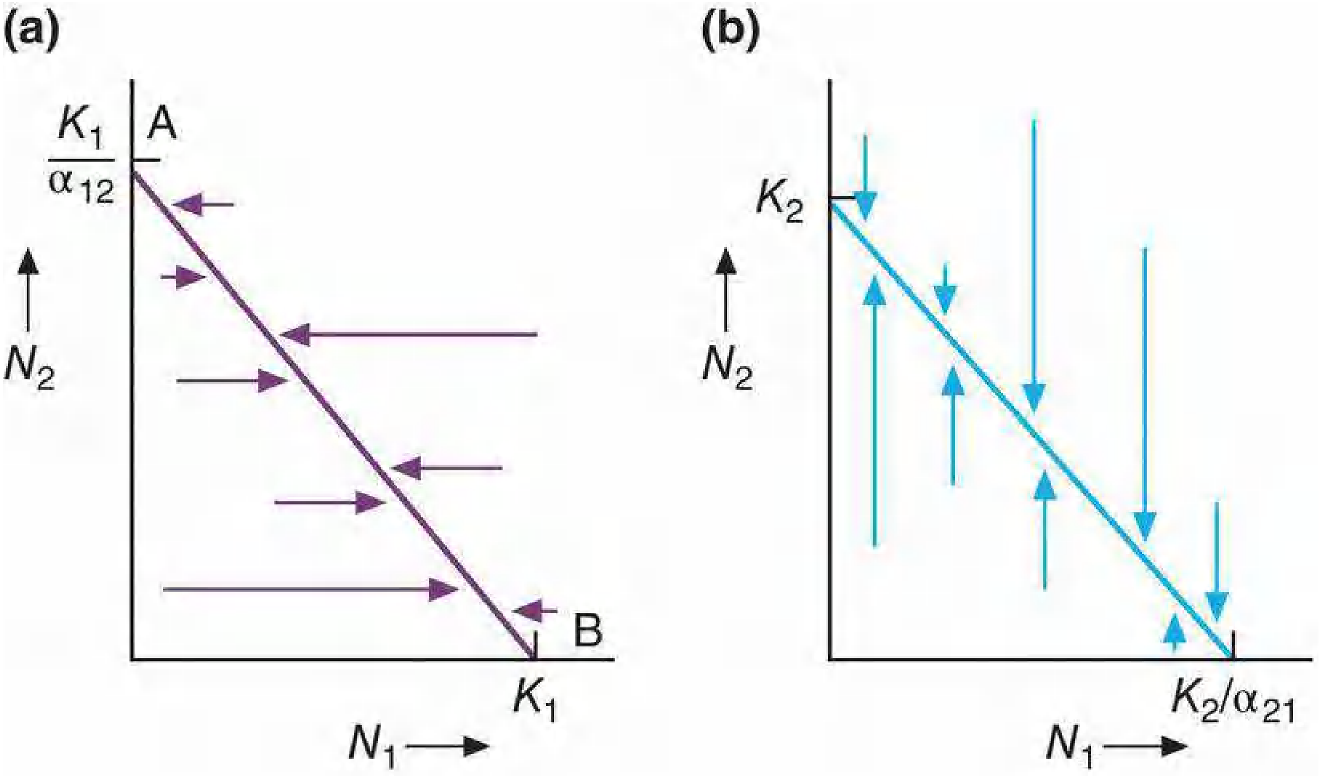


Figure 8.9 The zero isoclines generated by the Lotka–Volterra competition equations. (a) The N_1 zero isocline: species 1 increases below and to the left of it, and decreases above and to the right of it. (b) The equivalent N_2 zero isocline.

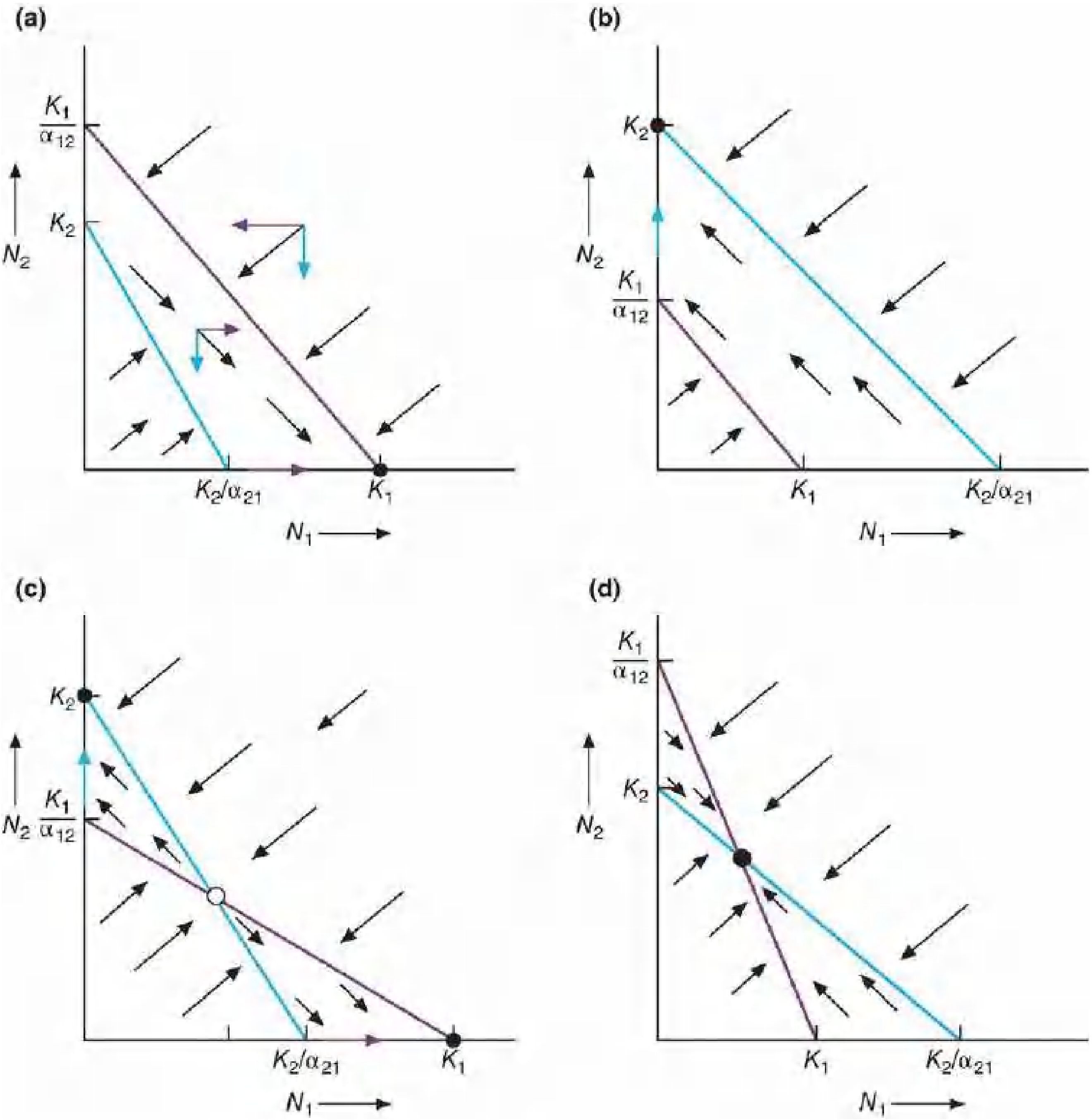


Figure 8.10 The outcomes of competition generated by the Lotka–Volterra competition equations for the four possible arrangements of the N_1 and N_2 zero isoclines. Vectors, generally, refer to joint populations, and are derived as indicated in (a). The solid circles show stable equilibrium points. The open circle in (c) is an unstable equilibrium point. For further discussion, see the text.

In order to draw a zero isocline for species 1, we can use the fact that on the zero isocline $dN_1/dt = 0$ (by definition), that is (from Equation 8.3):

$$r_1 N_1 (K_1 - N_1 - \alpha_{21} N_2) = 0. \tag{8.5}$$

This is true (trivially) when the intrinsic rate of increase (r_1) is zero or when the population size (N_1) is zero, but – much more interestingly – it is also true when:

$$K_1 - N_1 - \alpha_{21} N_2 = 0, \tag{8.6}$$

which can be rearranged as:

$$N_1 = K_1 - \alpha_{21}N_2. \tag{8.7}$$

The line is therefore the zero isocline for species 1, as shown in Figure 8.9a. Below and to the left of this, the numbers of both species are relatively low, and species 1, subjected to only weak competition, increases in abundance (the arrows in the figure, representing this increase, point from left to right, since N_1 is on the horizontal axis). Above and to the right of the line, the numbers are high, competition is strong and species 1 decreases in abundance (arrows from right to left). Based on an equivalent derivation, Figure 8.9b has combinations leading to an increase and decrease in species 2, separated by a species 2 zero isocline, with arrows, like the N_2 axis, running vertically.

four ways in which the two zero isoclines can be arranged

Finally, in order to determine the outcome of competition in this model, it is necessary to fuse Figures 8.9a and b, allowing the behaviour of a joint population to be predicted. In doing this, it should be noted that the arrows in Figure 8.9 are actually vectors, with a strength as well as a direction. Figure 8.10 shows that there are, in fact, four different ways in which the two zero isoclines can be arranged relative to one another, and the outcome of competition will be different in each case. The different cases can be defined and distinguished by the intercepts of the zero isoclines. For instance, in Figure 8.9a:

$$\frac{K_1}{\alpha_{12}} > K_2 \quad \text{and} \quad K_1 > \frac{K_2}{\alpha_{21}} \tag{8.8}$$

i.e.:

$$K_1 > K_2\alpha_{12} \quad \text{and} \quad K_1\alpha_{21} > K_2. \tag{8.9}$$

strong interspecific competitors outcompete weak interspecific competitors

The first inequality ($K_1 > K_2\alpha_{12}$) indicates that the inhibitory intraspecific effects that species 1 can exert on itself are greater than the interspecific effects that species 2 can exert on species 1. The second inequality, however, indicates that species 1 can exert more of an effect on species 2 than species 2 can on itself. Species 1 is thus a strong interspecific competitor, whilst species 2 is a weak interspecific competitor; and as the vectors in Figure 8.10a show, species 1 drives species 2 to extinction and attains its own carrying capacity. The situation is reversed in Figure 8.10b. Hence, Figures 8.10a and b describe cases in which the environment is such that one species invariably outcompetes the other.

when interspecific competition is more important than intraspecific, the outcome depends on the species' densities

In Figure 8.10c:

$$K_2 > \frac{K_1}{\alpha_{12}} \quad \text{and} \quad K_1 > \frac{K_2}{\alpha_{21}} \tag{8.10}$$

i.e.:

$$K_2\alpha_{12} > K_1 \quad \text{and} \quad K_1\alpha_{21} > K_2. \tag{8.11}$$

Thus, individuals of both species compete more strongly with individuals of the other species than they do amongst themselves. This will occur, for example, when each species produces an allelochemical that is toxic to the other species but is harmless to itself, or when each species is aggressive towards or even preys upon individuals of the other species, more than individuals of its own species. The consequence, as the figure shows, is an unstable equilibrium combination of N_1 and N_2 (where the isoclines cross), and two stable points. At the first of these stable points, species 1 reaches its carrying capacity with species 2 extinct; whilst at the second, species 2 reaches its carrying capacity with species 1 extinct. Which of these two outcomes is actually attained is determined by the initial densities: the species which has the initial advantage will drive the other species to extinction.

when interspecific competition is less important than intraspecific, the species coexist

Finally, in Figure 8.10d:

$$\frac{K_1}{\alpha_{12}} > K_2 \quad \text{and} \quad \frac{K_2}{\alpha_{21}} > K_1 \tag{8.12}$$

i.e.:

$$K_1 > K_2\alpha_{12} \quad \text{and} \quad K_2 > K_1\alpha_{21}. \tag{8.13}$$

In this case, both species have less competitive effect on the other species than they have on themselves. The outcome, as Figure 8.10d shows, is a stable equilibrium combination of the two species, which all joint populations tend to approach. The species coexist, stably.

8.4.2 Lessons from the Lotka–Volterra model

support for the competitive exclusion principle

The Lotka–Volterra model confirms aspects of what we have already seen in our series of actual examples. It is able to generate a range of possible outcomes – the predictable exclusion of one species by another, exclusion dependent on initial densities, and stable coexistence – and these outcomes can be aligned clearly with biologically reasonable circumstances. The model, therefore, in spite of its simplicity and its failure to address many of the complexities of the dynamics of competition in the real world, serves a useful purpose. Figure 8.10a and b describes cases in which a strong interspecific competitor invariably outcompetes a weak interspecific competitor, or, we might say, in which the weak interspecific competitor lacks a realised niche when in competition with the stronger competitor. We saw this, for example, with our charr and with our protists, previously. By contrast, Figure 8.10d describes the case in which both species compete more strongly with members of their own species than with those of the other species, and they coexist, stably, as a consequence. This is clearly reminiscent of those examples in which two species coexisted where there was differentiation of their realised niches: the charr on a broader geographic scale, for example, or the warblers. Between them, these outcomes capture the essence of the competitive exclusion principle.

mutual antagonism: the outcome is probable rather than definite

Finally, Figure 8.10c describes a situation in which interspecific competition is, for both species, a more powerful force than intraspecific competition: both species are positively harmful to the other. This is reminiscent of the examples of allelopathy we saw above, except that now those harmful effects are reciprocated. A classic example of such a situation is provided by work on two species of flour beetle, *Tribolium confusum* and *T. castaneum*, carried out by Park (1962), whose experiments in the 1940s, 1950s and 1960s were amongst the most influential in shaping ideas about interspecific competition. He reared the beetles in simple containers of flour, which provided fundamental and often realised niches for both species. But as well as eating the flour, the beetles preyed upon each other, and crucially here, beetles of both species ate more individuals of the other species than they did of their own. Figure 8.9c suggests that because species are affected more by inter- than intraspecific competition, the outcome is strongly dependent on the relative abundances of the competing species. The small amount of interspecific aggression displayed by a rare species will have relatively little effect on an abundant competitor; but the large amount of aggression displayed by an abundant species might easily drive a rare species to local extinction. Moreover, if abundances are finely balanced, a small change in relative abundance will be sufficient to shift the advantage from one species to the other. The outcome of competition will then be unpredictable – either species could exclude the other, depending on the exact densities that they start with or attain. Table 8.1 shows that this was indeed the case with Park's flour beetles. There was always only one winner, and the balance between the species changed with climatic conditions. Yet at all intermediate climates the outcome was probable rather than definite. Even the inherently inferior competitor occasionally achieved a density at which it could outcompete the other species.

Table 8.1 Competition between *Tribolium confusum* and *T. castaneum* in a range of climates. One species is always eliminated and climate alters the outcome, but at intermediate climates the outcome is nevertheless probable rather than definite.

Source: After Park (1954).

Climate	**Percentage wins**	
	T. confusum	**T. castaneum**
Hot–moist	0	100
Temperate–moist	14	86
Cold–moist	71	29
Hot–dry	90	10
Temperate–dry	87	13
Cold–dry	100	0

what the Lotka–Volterra model leaves out

We must remember, however, that the Lotka–Volterra model incorporates no niches, as such, nor any explicit consideration of resources, which we saw in the phytoplankton studies, for example, were so important in determining the outcome of competition. The focus in the Lotka–Volterra model, rather, is on the relative strengths of intra- and interspecific competition. These may reflect patterns in the differentiation of niches or in the differential utilisation of resources, but they need not do so. Interspecific competition is a process that is often associated, ecologically and evolutionarily, with a particular pattern (niche differentiation), but the process and the pattern are not inextricably linked. Niche differentiation can arise through other processes, and the outcome of interspecific competition may depend on things other than the differentiation of niches, as we also see below.

8.5 Consumer-resource models of competition

We will be looking in some detail at the interactions between consumers and their resources (for example, predators and their prey) in later chapters. But we need to turn now to models of interspecific competition that explicitly consider consumer-resource dynamics. This serves to remind us that the clear divisions that we have imposed in order to divide this text into chapters and provide a linear structure – for example, separating out competition and predation – are not ones that are recognised by the organisms themselves.

8.5.1 A model for a single resource

We start by considering two species competing for a single limiting resource (Tilman, 1982 , 1990). Different models, based on varying details in the mechanism of exploitation, can be constructed, but the following simple model links the dynamics of two competing consumers, abundances N_1 and N_2, to those of their shared abiotic resource, amount R. (We retain here the convention of calling this R, which should not be confused with the net reproductive rate, also called R, discussed in Section 4.7.1 and elsewhere.)

$$\frac{dN_1}{dt} = \frac{g_1 N_1 R}{(R + C_1)} - m_1 N_1 \tag{8.14}$$

$$\frac{dN_2}{dt} = \frac{g_2 N_2 R}{(R + C_2)} - m_2 N_2 \tag{8.15}$$

$$\frac{dR}{dt} = a(R_{\max} - R) - \frac{\left(\dfrac{g_1 N_1 R}{R + C_1}\right)}{Y_1} - \frac{\left(\dfrac{g_2 N_2 R}{R + C_2}\right)}{Y_2}. \tag{8.16}$$

The first terms in the consumer Equations 8.14 and 8.15 say that the consumer species grow at a rate that increases towards a maximum rate, g_i, as R increases, reaching half that rate when the amount of resource is C_i (where i is either 1 or 2, depending on the species concerned). Thus, C_i is highest in consumers that require the most resource in order to grow rapidly. The second terms describe their loss from the system at per capita mortality rate, m_i. In the resource Equation 8.16, the amount of resource increases as a result of being supplied to the system at a rate, a, up to a maximum amount, $R_{\max}$. Resource is also depleted by the two consumers, at a rate that depends on the positive terms in their growth rates, converted from number of consumer individuals to amount of resource by dividing by Y_i, the number of individuals of species i produced per unit of resource.

the importance of low R^* confirmed – and a possible need for trade-offs

Ignoring any details of calculations, if we take one consumer species at a time, then each reaches a stable equilibrium abundance when the amount of resource equilibrates at R^*, where this is given by:

$$R_i^* = m_i C_i / (g_i - m_i). \tag{8.17}$$

But it is not possible to have stable coexistence of both species and the resource. Only the species that reduces the amount of resource to the lowest value of R^* survives, confirming in a mathematical model what we have seen in Section 8.2.1 for actual examples.

Equation 8.17 also tells us that successful exploitative competitors (low R^*_i) are those that combine resource-utilisation efficiency (low C_i), low rates of loss (low m_i) and high rates of increase (high g_i). On the other hand, it may not be possible for an organism to combine, say, low

C_i and high g_i. A plant's growth, for example, will be most enhanced by putting its matter and energy into leaves and photosynthesis – but to enhance its nutrient-utilisation efficiency it would have to put these into roots. A lioness will be best able to subsist at low densities of prey by being fleet-footed and manoeuvrable – but this may be difficult if she is often heavily pregnant. Understanding successful exploitative competitiveness, therefore, may require us ultimately to understand how organisms trade off features giving rise to low values of R^* against features that enhance other aspects of fitness.

8.5.2 A model for two resources

the zero net growth isocline: a niche boundary

We now extend the model in Equations 8.14–8.17 to include not one but two resources, X and Y. This time, we refer to the two consumer species as A and B. Details of the equations and their analysis can be found in Tilman (1982 , 1986); here we just focus on results. Beginning with just one of the consumer species, we can set its net growth rate to zero and so define its zero net growth isocline (ZNGI) when utilising two essential resources (see Section 3.8). Because the resources are essential, the ZNGI is rectangular (Figure 8.11), with the arms defined by X_1 * and Y_1 *, respectively (though similar arguments could be made for other resource-pairs and hence shapes for the ZNGI). This isocline is the boundary between resource combinations that are sufficient to allow the species to survive and reproduce, and resource combinations that are not. It therefore also represents the boundary of the species' niche in these two dimensions. The system settles at an equilibrium at which both consumer population size and the resource levels remain constant. Population size is constant (by definition) at all points on the isocline, but there is only one point on the isocline where both resource levels are also constant (point S * in Figure 8.11). This point is thus the two-resource equivalent of R * for one resource. It represents a balance between the consumption of the resources by the consumer (taking the resource concentrations towards the bottom left of the figure) and the natural renewal of the resources (taking the concentrations towards the top right). Indeed, in the absence of the consumer, resource renewal would take the resource concentrations to the 'supply point', shown in the figure.

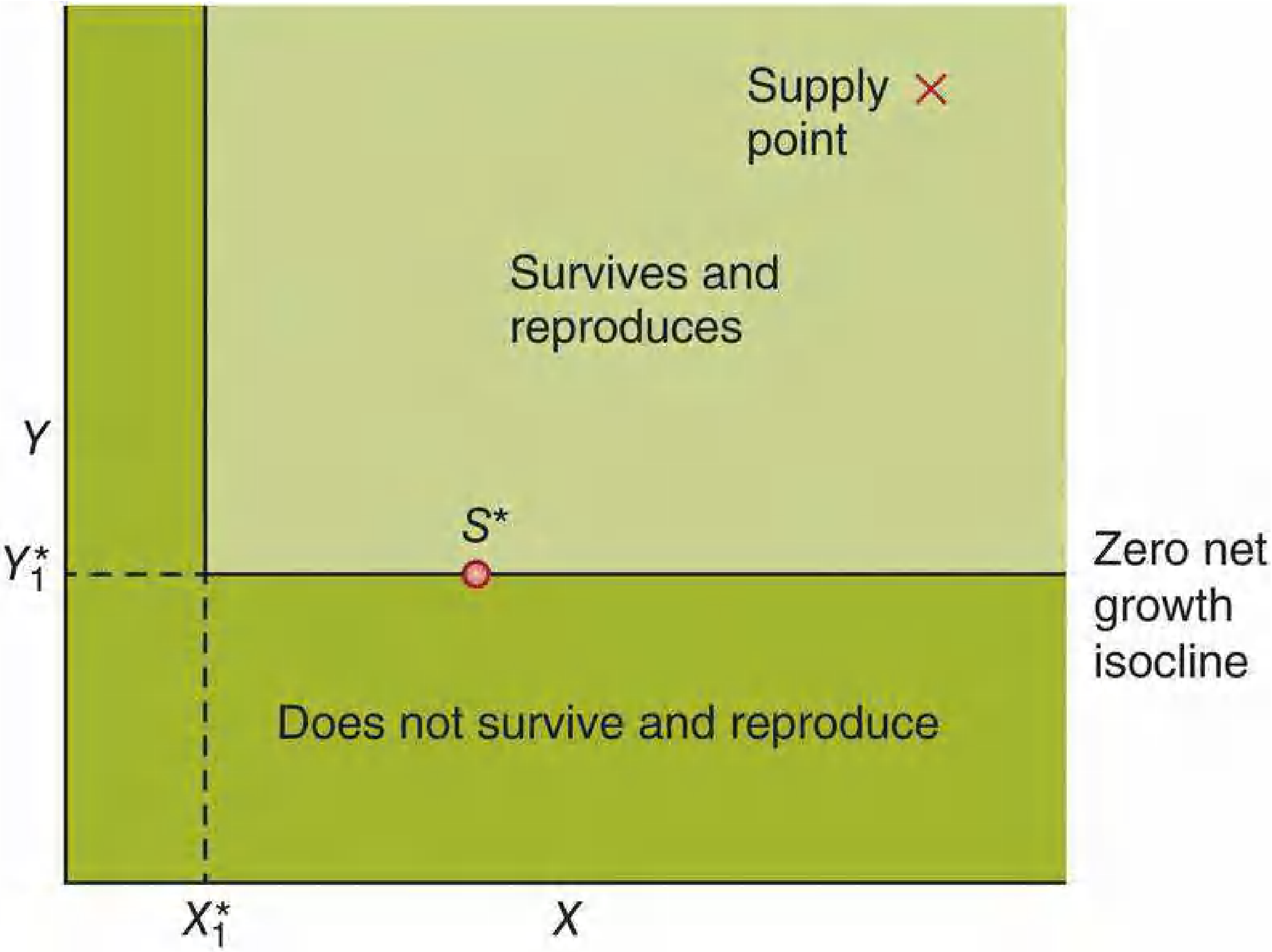


Figure 8.11 The zero net growth isocline of a species potentially limited by two resources, divides resource combinations on which the species can survive and reproduce, from those on which it cannot. The isocline is rectangular in this case because the two resources, X and Y, are essential (see Section 3.8.1). Point S^* is the only point on the isocline at which there is also no net change in resource concentrations (consumption and resource renewal are equal and opposite). In the absence of the consumer, resource renewal would take the resource concentrations to the 'supply point' shown.

To move from intra- to interspecific competition, it is necessary to superimpose the isoclines of two species on the same diagram (Figure 8.12) much as we did with the Lotka–Volterra isoclines. The two species will have different consumption rates, but there will still be a single supply point. The outcome depends on the position of this supply point.

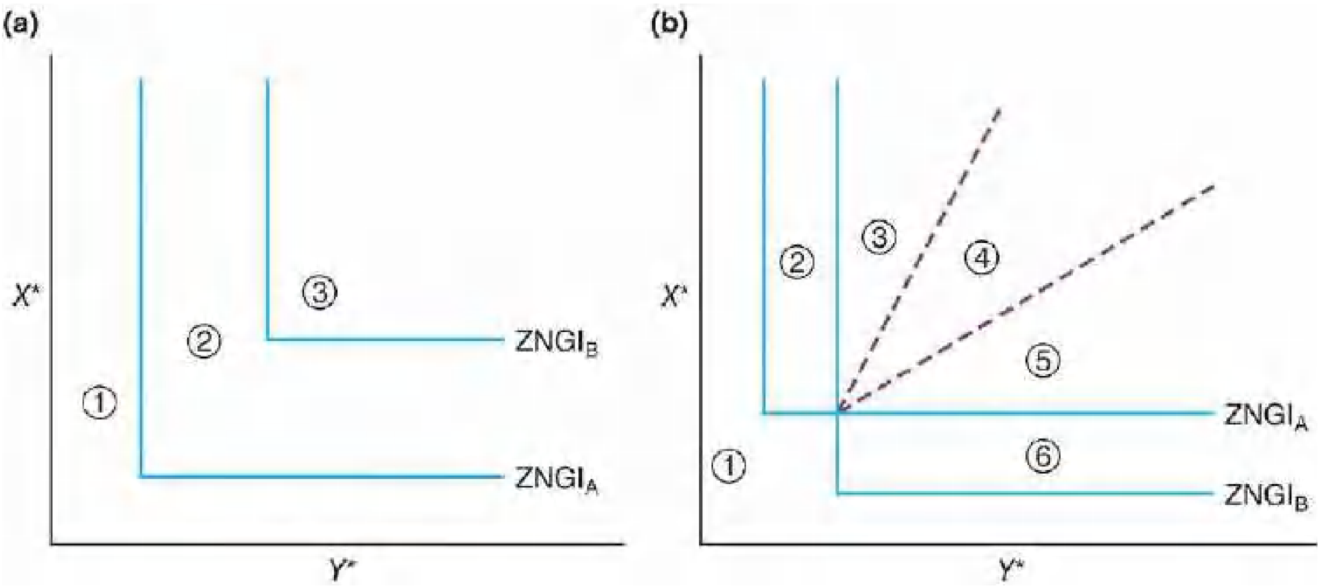


Figure 8.12 Competitive exclusion and coexistence in models with zero net growth isoclines. (a) Competitive exclusion: the isocline (zero net growth isocline, ZNGI) of species A lies closer to the resource axes than the isocline of species B. If the resource supply point is in region 1, then neither species can exist. But if the resource supply point is in regions 2 or 3, then species A reduces the resource concentrations to a point on its own isocline (where species B cannot survive and reproduce): species A excludes species B. (b) Potential coexistence of two competitors limited by two essential resources. The isoclines of species A and B overlap, leading to six regions of interest. With supply points in region 1 neither species can exist; with points in regions 2 and 3, species A excludes species B; and with points in regions 5 and 6, species B excludes species A. Region 4 contains supply points lying between the limits defined by the two dashed lines. With supply points in region 4 the two species coexist. For further discussion, see text.

a superior and an inferior competitor

In Figure 8.12a, the isocline of species A is closer to both axes than that of species B. There are three regions in which the supply point might be found. If it was in region 1, below the isoclines of both species, then there would never be sufficient resources for either species and neither would survive. If it was in region 2, below the isocline of species B but above that of species A, then species B would be unable to survive and the system would equilibrate on the isocline of species A. If the supply point was in region 3, then this system too would equilibrate on the isocline of species A. Analogous to the one-resource case, species A would competitively exclude species B because of its ability to exploit both resources down to levels at which species B could not survive. Of course, the outcome would be reversed if the positions of the isoclines were reversed.

In Figure 8.12b the isoclines of the two species overlap, and there are six regions in which the supply point might be found. Points in region 1 are below both isoclines and would allow neither species to exist; those in region 2 are below the isocline of species B and would only allow species A to exist; and those in region 6 are below the isocline of species A and would only allow species B to exist. Regions 3, 4 and 5 lie within the fundamental niches of both species. However, the outcome of competition depends on which of these regions the supply point is located in.

coexistence – dependent on the ratio of resource levels at the supply point

The most crucial region in Figure 8.12b is region 4. For supply points here, the resource levels are such that species A is more limited by resource X than by resource Y, whilst species B is more limited by Y than X. However, species A consumes more X than Y, whilst species B consumes more Y than X. Because each species consumes more of the resource that more limits its own growth, the system equilibrates at the intersection of the two isoclines, and this equilibrium is stable: the species coexist.

subtle niche differentiation – each species consumes more of the resource that more limits its own growth

This is niche differentiation, but of a subtle kind. Rather than the two species exploiting different resources, species A disproportionately limits itself by its exploitation of resource X, whilst species B disproportionately limits itself by its exploitation of resource Y. The result is the coexistence of competitors. By contrast, for supply points in region 3, both species are more limited by Y than X. But species A can reduce the level of Y to a point on its own isocline below species B's isocline, where species B cannot exist. Conversely, for supply points in region 5, both species are more limited by X than Y, but species B depresses X to a point below species A's isocline. Thus, in regions 3 and 5, the supply of resources favours one species or the other, and there again is competitive exclusion.

It seems then that two species can compete for two resources and coexist as long as two conditions are met.

1. The habitat (i.e. the supply point) must be such that one species is more limited by one resource, and the other species more limited by the other resource.
2. Each species must consume more of the resource that more limits its own growth.

As with other cases of coexistence by niche differentiation, the essence is that intraspecific competition is, for both species, a more powerful force than interspecific competition.

8.5.3 Models with complex dynamics

coexistence via complex dynamics, not niche differentiation

Up to this point, therefore, the consumer–resource approach to interspecific competition has supported the view that emerges from the Lotka–Volterra model, but also augmented it by stressing the importance of resource levels and by adding subtleties to our view of niche differentiation. But we have limited ourselves so far to consumer–resource interactions in which the abundances of each consumer–resource pair settle at a stable, equilibrium combination. The picture changes if we allow for more complex interactions. We can start by assuming that the resource is itself biotic. We can therefore think of it as a 'prey' species. We will retain the convention of calling its abundance *R*, and assume that it is self-limited in the manner of the logistic equation (Section 5.7), but that it is also consumed by two species, numbers N_1 and N_2, that therefore compete for the resource (Armstrong & McGehee, 1980). Hence,

$$\frac{dR}{dt} = rR\left(1 - \frac{R}{K}\right) - c_1 RN_1 - c_2 RN_2. \tag{8.18}$$

As in the logistic, *r* and *K* are the intrinsic growth rate and carrying capacity of the resource species, while c_1 and c_2 are the consumption rates of the two consumer species. Their own dynamics are simply a death rate, m_i, counteracted by a birth rate that depends on the amount of resource each consumes Equation 8.18 and their conversion efficiencies in turning resource into new consumers, e_i:

$$\frac{dN_1}{dt} = e_1c_1RN_1 - m_1N_1 \tag{8.19}$$

$$\frac{dN_2}{dt} = e_2c_2RN_2 - m_2N_2. \tag{8.20}$$

In these very basic equations, the consumers and their prey do settle at a simple, stable equilibrium. Hence, they provide yet another route for getting to the Lotka–Volterra, competitive exclusion outcome, namely that only one of the competing species survives, at a stable equilibrium abundance – that being the species that, alone, reduces the equilibrium abundance of the prey resource, R^*, to the lowest level. However, the outcome changes as soon as we modify the dynamics of just one of the competitors such that its interaction with the resource gives rise not to a simple, stable consumer–resource equilibrium, but to a dynamic pattern in their abundances – for example, replacing Equations 8.18 and 8.19 with:

$$\frac{dR}{dt} = rR\left(1 - \frac{R}{K}\right) - c_1\left(\frac{R}{R + G}\right)N_1 - c_2RN_2 \tag{8.21}$$

$$\frac{dN_1}{dt} = e_1c_1\left(\frac{R}{R + G}\right) - m_1R. \tag{8.22}$$

These changes reflect the reasonable assumption that as the amount of available resource increases, the consumption rate may level off (these ‘functional responses’ are examined more fully in Section 10.2). The dynamics of this modified interaction are shown, for particular values of the parameters, in Figure 8.13a. This time, both of the competing consumers coexist on the single resource, albeit at numbers that fluctuate over time. The reason for this is explained graphically in Figure 8.13b. Consumer 1 can invade a system comprising only consumer 2 and the resource, because (at these parameter values) it can reduce the amount of resource to a level, $R_1{}^*$, that is lower than the level to which consumer 2 can reduce it, $R_2{}^*$. But conversely, consumer 2 can invade a system comprising only consumer 1 and the resource, because the level to which consumer 2 can reduce the resource, $R_2{}^*$, is itself lower than the average level of resource when species 1 and the resource interact alone, $\overline{R}_1$. Hence, because $R_1{}^* < R_2{}^* < \overline{R}_1$, neither consumer can eliminate the other – and if neither competitor can exclude the other, they must necessarily coexist.

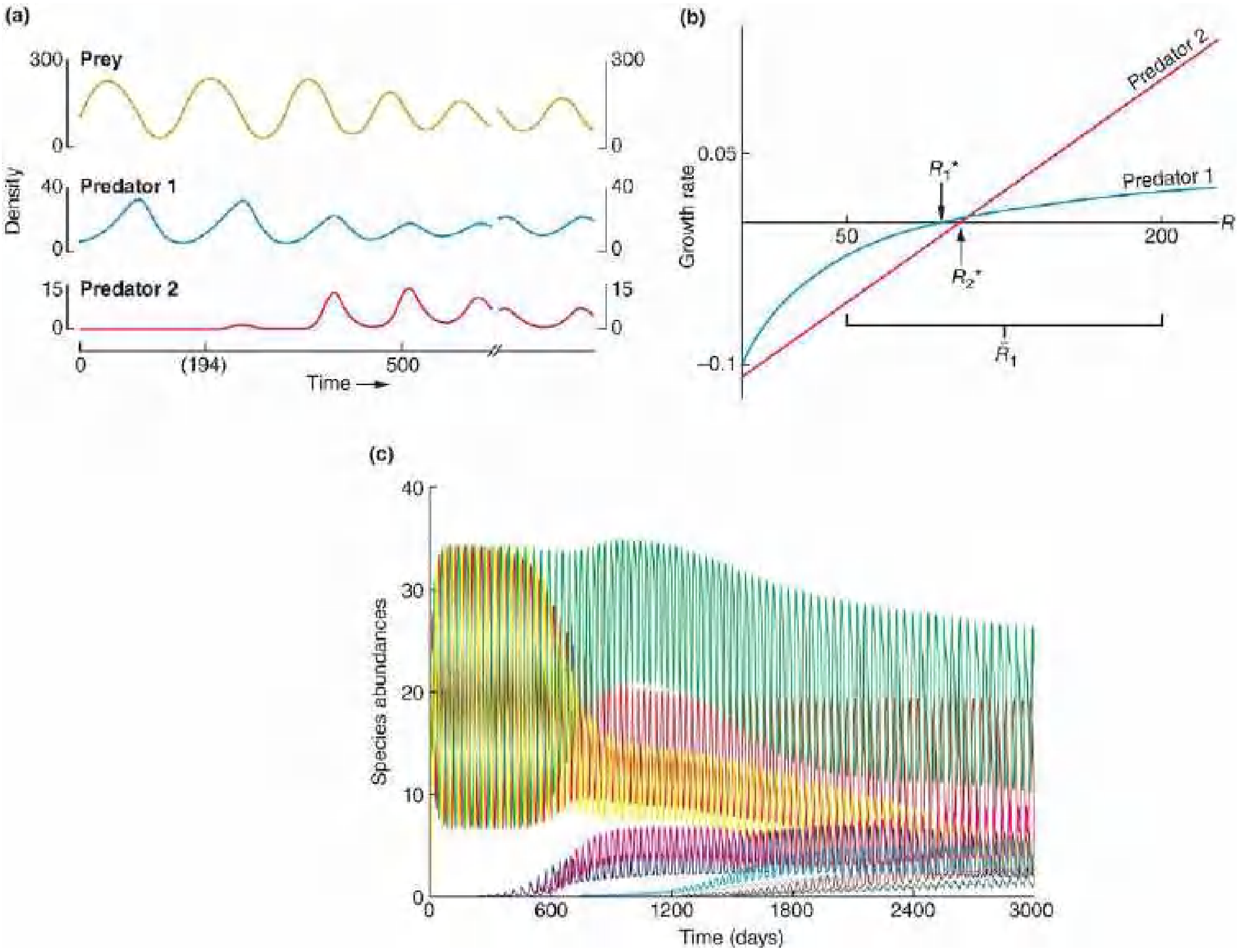


Figure 8.13 Coexistence of competitors sharing a resource is facilitated when they display complex dynamics. (a) The dynamics of two predator species competing for a shared prey species, with their dynamics described by Equations 8.21 and 8.22. The system was started with only predator 1, and a small number of predator 2 added at time point 194. Long term, the system settles to the consistent dynamic pattern shown to the right of the break-point in the time axis. (b) Population growth rates of predators 1 and 2 in the system shown in (a), over a range of prey densities, R. $R_1{}^*$ and $R_2{}^*$ are the levels to which predators 1 and 2, respectively, can reduce the prey and maintain positive growth. The bracket represents the range of prey abundances exhibited by predator 1 in the two species system, generating an average prey abundance $\overline{R}_1$. The two predators coexist because $R_1{}^* < R_2{}^* < \overline{R}_1$. (c) The dynamics of nine phytoplankton species (colour-coded) competing for three chemical resources, governed by consumer–resource equations that allow for non-linear interactions between consumers and resources and hence can generate these complex dynamics.

Source: (a, b) After Armstrong & McGehee (1980). (c) After Huisman & Weissing (1999).

In a similar way, Huisman and Weissing (1999) investigated models in which competing species utilised not biotic resources (‘prey’) but chemical nutrients, as phytoplankton would. This connects their model more directly to Hutchinson’s ‘paradox of the plankton’. Again, when they replaced the very simple Equations 8.18–8.20 with more realistic interactions between the phytoplankton and the nutrients, complex patterns of fluctuations were readily generated, including chaos (Section 5.6.5); and again, this allowed the coexistence of many more competitors than resources. An example of nine competitors coexisting on three resources is shown in Figure 8.13c.

8.5.4 Consumer–resource competition in practice

two competitors coexisting on two resources

We have already seen several examples – in plants, in geckos and in phytoplankton (see Sections 8.2.1 and 8.2.2) – in which the most fundamental prediction of these models is confirmed, where the winner between competing species is the one that can survive with their shared, limiting resource at the lowest level. We have also seen an example, again with phytoplankton, with not one but two shared, limiting resources, where in cultures with relatively balanced supplies of the two resources, the two competitors coexisted, since both species were provided with sufficient supplies of a resource on which they were inferior (see Figure 8.4).

Similar patterns have been demonstrated in plants, too. In an 11-year experiment on competition for nitrogen and for light amongst six species of prairie grasses, there were six pairwise combinations in which the species equilibrating with lower levels of both nitrogen *and* light (N^* and L^*) proved to be the superior competitor (Dybzinski & Tilman, 2007). However, there were two further combinations in which one species had the lower N^* and the other the lower L^*. In these, the species pairs coexisted. What's more, where there was exclusion, the rate at which it occurred was slower where the differences in N^* or L^* were smaller.

There are also rare examples in which this pattern is apparent when two species compete not for abiotic but for biotic resources. In one of these, returning again to plankton, two species of rotifers, *Brachionus rubens* and *B. calyciflorus*, competed for two species of alga, *Monoraphidium minutum* and *Chlamydomonas sphaeroides* (Rothaupt, 1988). When the rotifers were reared alone, *B. rubens* fed more effectively and grew faster on *Monoraphidium*, whereas *B. calyciflorus* fed more effectively and grew faster on *Chlamydomonas* (Figure 8.14a). In this case, both food resources are acceptable for both rotifers and are therefore perfectly substitutable (Section 3.8). Hence the zero net growth isoclines (ZNGIs) are single straight lines. According to the theory, the outcome should depend on the positions of the ZNGIs and the supply rate of new resource, and since in this case the ZNGIs cross, the potential for coexistence clearly exists at appropriate resource supply concentrations, though not at others (Figure 8.14b). In fact, when these predictions of theory were tested experimentally, they were confirmed in 11 out of 12 cases (Figure 8.14b).

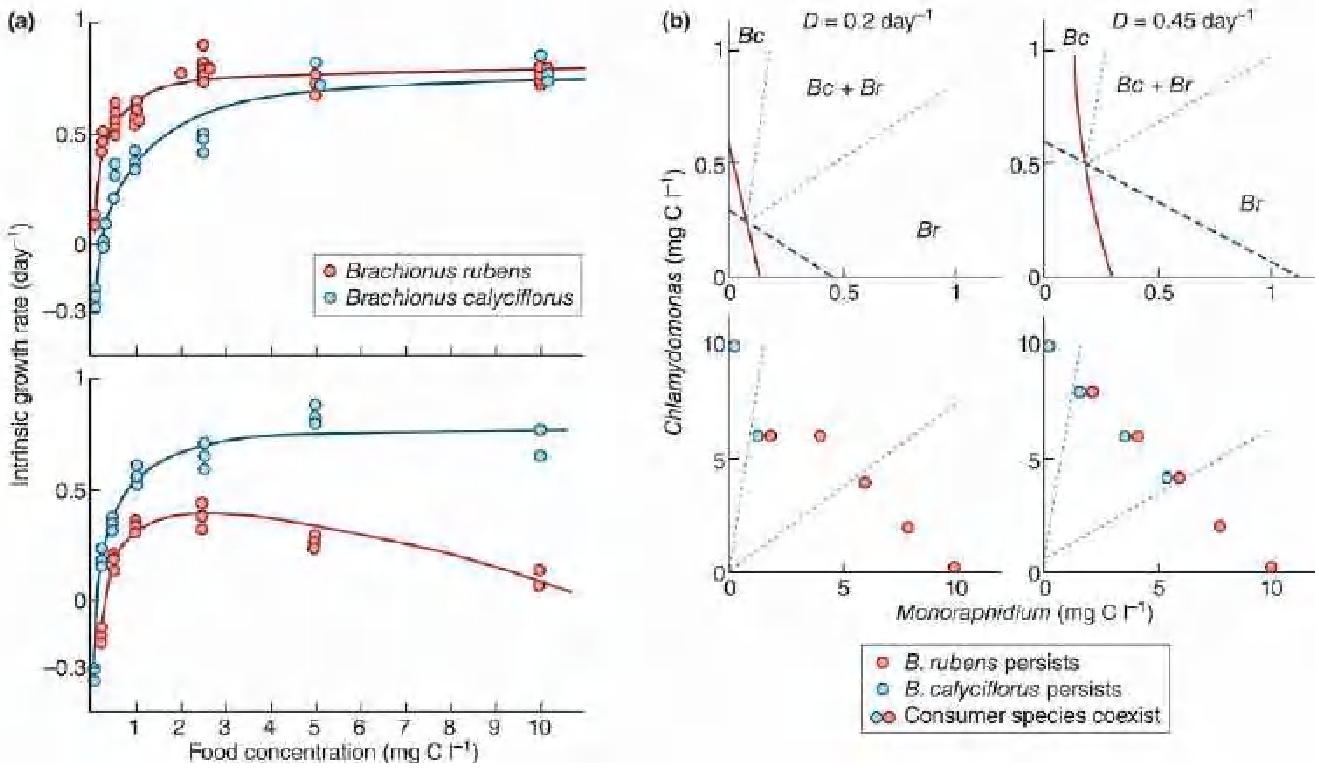


Figure 8.14 **Two species of rotifers competing for two species of alga coexist only at appropriate supply rates of the algae.** (a) Population growth rates of *Brachionus rubens* (red) and *B. calyciflorus* (blue) reared alone on *Monoraphidium minutum* (above) and *Chlamydomonas sphaeroides* (below) in experimental cultures. The curves were fitted by non-linear regression. The decline in growth rate of *B. rubens* at high concentrations of *Chlamydomonas* is caused by mechanical disruption of the feeding process. (b). Above, the predicted outcome of competition in the system at two different dilution rates: 0.2 and 0.45 of the culture medium being renewed with fresh resource each day. The solid red lines are the ZNGIs for *B. rubens* and the dashed blue lines those for *B. calyciflorus*; the dotted lines define the region within which supply points of resources are such that each consumer consumes most of the resource that most limits it (see Figure 8.12). The outcomes of competition thus predicted in the different regions are indicated. Below, those same regions with the outcome of competition experiments indicated: a red circle indicates that only *B. rubens* persists, a blue circle that only *B. calyciflorus* persists, and two circles side by side that the consumer species coexist. Predictions were confirmed in 11 out of 12 cases.

Source: After Rothaupt (1988).

trade-offs in grasses

We can find support, too, for the idea that trade-offs may be necessary between resource-use effectiveness (competitive ability) and other important life history characteristics. We do so by returning to the example of Tilman's grasses that we saw in Figures 8.2. Five species were chosen from various points in a typical old-field successional sequence in Minnesota, USA (Figure 8.15a; see Section 18.4 for a much fuller discussion of succession), and it is clear that the better competitors for nitrogen are found later in the sequence. These species, and *S. scoparium* and *A. gerardi* in particular, had higher root allocations, but their above-ground vegetative growth rates and reproductive allocations were lower (e.g. Figure 8.15b). In other words, they achieved their low values of R * by the high resource-utilisation efficiency given to them by their roots (low C_i, Equation 8.17), but appear to have paid for this through a reduction in growth and reproductive rates (lower g_i). In fact, over all the species, a full 73% of the variance in the eventual soil nitrate concentration was explained by variations in root mass (Tilman & Wedin, 1991a). This successional sequence therefore appears to be one in which fast growers and reproducers are replaced by efficient and powerful exploiters and competitors.

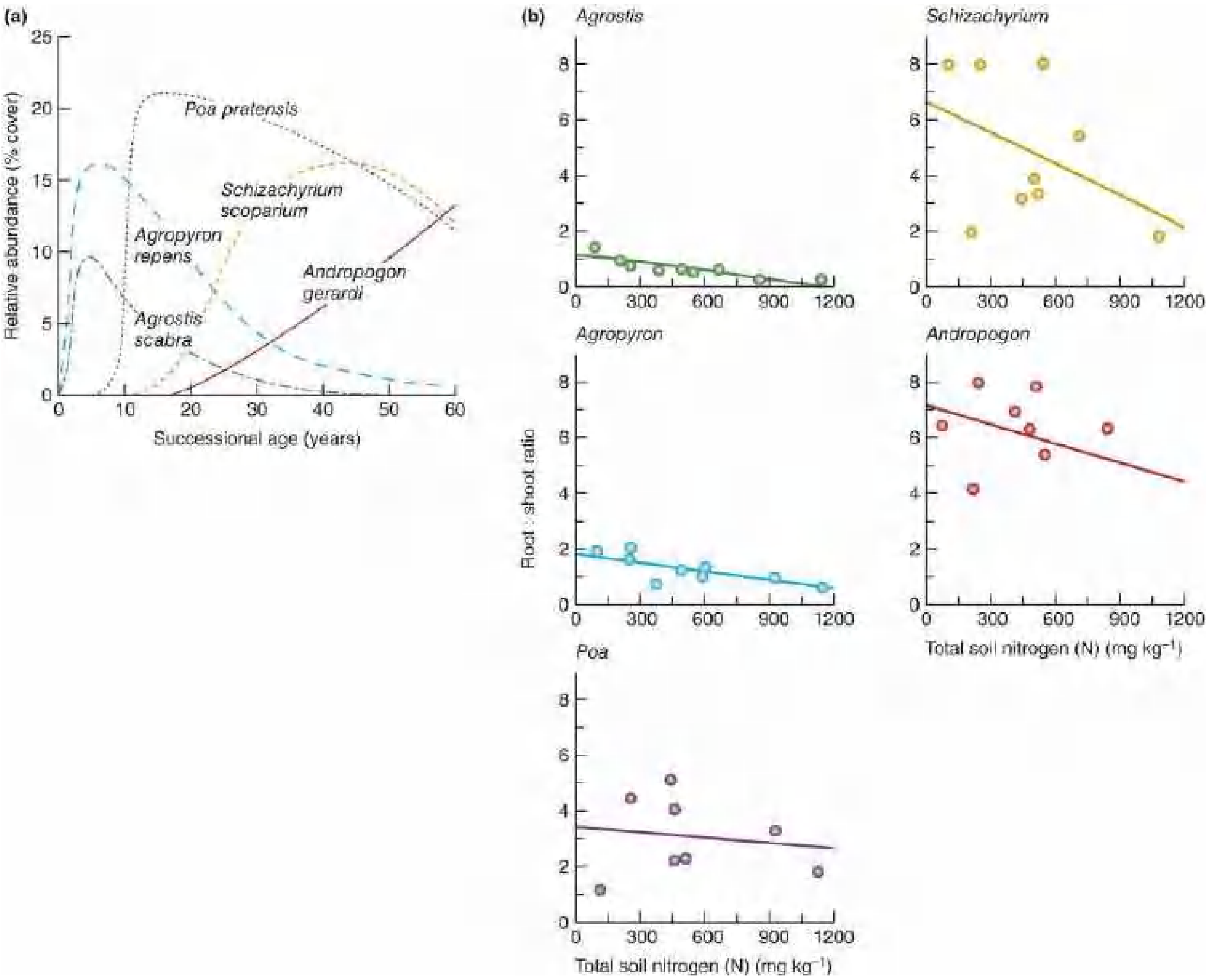


Figure 8.15 **Grasses exhibit a trade-off between resource-use effectiveness and growth and reproductive rates.** (a) The relative abundances of five grasses during old-field successions at Cedar Creek Natural History Area, Minnesota, USA. (b) The root : shoot ratios were generally higher in the later successional species and declined as soil nitrogen increased.

Source: After Tilman & Wedin (1991a).

the more limiting resources there are, the more species may coexist

Moving beyond species pairs, we can see that resource competition theory predicts that the number of coexisting species should increase with the number of resources that are at physiologically limiting levels, but also that more complex dynamics may lead to there being more coexisting competitor species than resources. Interlandi and Kilham (2001) tested these ideas directly in three lakes in the Yellowstone region of Wyoming, USA using an index (Simpson's index) of the species diversity of the phytoplankton there (diatoms and other species). If one species exists on its own, the index equals 1; in a group of species where biomass is strongly dominated by a single species, the index will be close to 1; when two species exist at equal biomass, the index is 2; and so on. The index is therefore always equal to or less than the number of species, and will be far less whenever species abundances are unevenly distributed. According to resource competition theory, this index should increase with the number of resources limiting growth. The spatial and temporal patterns in phytoplankton diversity in the three lakes for two years, 1996 and 1997, are shown in Figure 8.16a. The principal limiting resources for phytoplankton growth were nitrogen, phosphorus, silicon and light, and so the levels of these were measured at the same depths and times that the phytoplankton were sampled. It was noted where and when any of the potential limiting factors actually occurred at levels below threshold limits for growth. Species diversity did indeed increase with the number of limiting resources (Figure 8.16b), but equally pertinent, the number of coexisting competitor species frequently exceeded four (Figure 8.16a), the number of limiting resources. This, therefore, provides a rare example of field data supporting the theory.

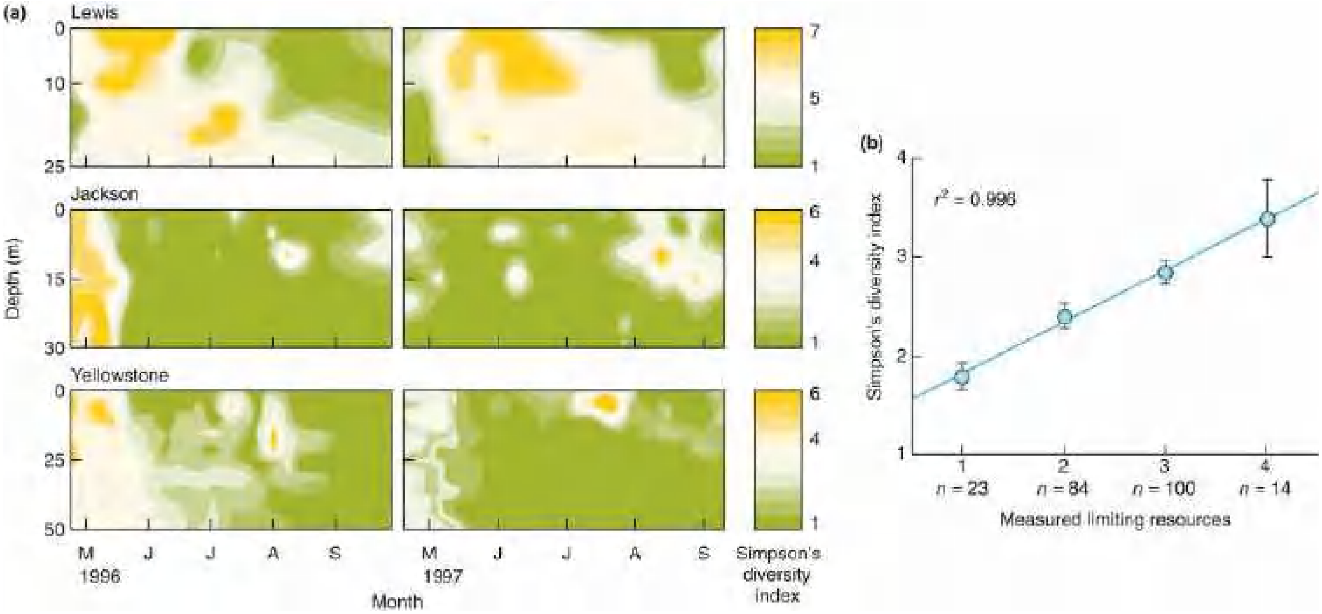


Figure 8.16 The diversity of competing phytoplankton species increases with the number of limiting resources. (a) Variation in phytoplankton species diversity (Simpson's index) with depth in two years in three large lakes in the Yellowstone region, USA. Shading indicates depth–time variation in a total of 712 discrete samples: dark yellow areas denote high species diversity, and green areas denote low species diversity. (b) Phytoplankton diversity (Simpson's index; mean ± SE) associated with samples with different numbers of measured limiting resources in 221 samples from those in (a). The number of samples (*n*) in each limiting resource class is shown.

Source: After Interlandi & Kilham (2001).

competition for one resource may influence competition for another

Finally, it is worth noting that whereas resource competition theory tends to focus on resources considered separately, even when they are subsequently combined in resource-pairs, in practice competition for one resource often affects the ability of an organism to exploit another resource. For example, Buss (1979) pointed out that in interactions between species of bryozoa (colonial, modular animals), there is an interdependence between competition for space and for food. Competition for space affects feeding in that when a colony of one species contacts a colony of another species, it interferes with the self-generated feeding currents upon which bryozoans rely. But equally, competition for food affects competition for space in that a colony short of food will have a greatly reduced ability to grow into and over its neighbours.

root and shoot competition

Comparable examples are found amongst rooted plants. If one species invades the canopy of another and deprives it of light, the suppressed species will suffer directly from the reduction in light energy that it obtains, but this will also reduce its rate of root growth, and it will therefore be less able to exploit the supply of water and nutrients in the soil. This in turn will reduce its rate of shoot and leaf growth. Repercussions flow backwards and forwards between roots and shoots. A number of workers have attempted to separate the effects of canopy and root competition by an experimental design in which two species are grown: (i) alone; (ii) together; (iii) in the same soil, but with their canopies separated; and (iv) in separate soil with their canopies intermingling. One example is a study of maize (*Zea mays*) and pea plants (*Pisum sativum*) (Semere & Froud-Williams, 2001). In full competition, with roots and shoots intermingling, the biomass production of maize and peas respectively (dry matter per plant, 46 days after sowing) was reduced to 59% and 53% of the 'control' biomass when the species were grown alone. When only the roots intermingled, pea plant biomass production was still reduced to 57% of the control value, but when just the shoots intermingled, biomass production was only reduced to 90% of the control (Figure 8.17). These results indicate, therefore, that soil resources (mineral nutrients and water) were more limiting than light, a common finding in the literature (Snaydon, 1996). They also support the idea of root and shoot competition combining to generate an overall effect, in that the overall reduction in plant biomass (to 53%) was close to the product of the root-only and shoot-only reductions (90% of 57% is 51.3%).

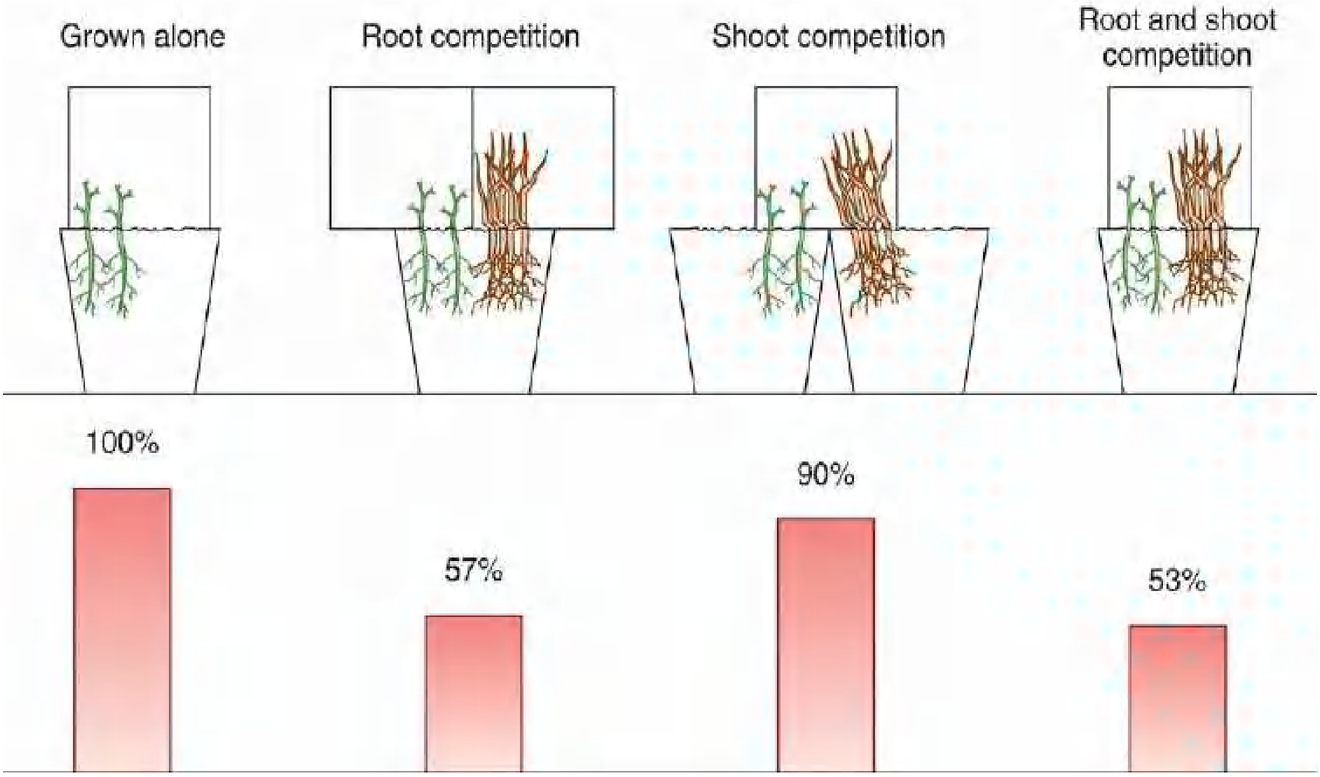


Figure 8.17 Root and shoot competition between maize and pea plants. Above are the experimental plants used, below are the dry weights of pea plants after 46 days as a percentage of those achieved when grown alone.

Source: Data from Semere & Froud-Williams (2001).

8.5.5 Spatial and temporal separation of niches

In practice, even when consumer–resource differences are at the heart of competitor coexistence, the resources used by ecologically similar species are separated spatially. Differential resource utilisation will then express itself as either microhabitat differentiation between the species (e.g. different species of fish feeding at different depths), or even a difference in geographic distribution. Alternatively, the availability of the different resources may be separated in time; for example, different resources may become available at different times of the day or in different seasons. Differential resource utilisation may then express itself as a temporal separation between the species.

We must remember, however, that resource partitioning is only one means of niche differentiation, and the coexistence of competitors may not be inevitably bound up with their use of resources. Niches can also be differentiated on the basis of conditions. Two species may use precisely the same resources, but if their ability to do so is influenced by environmental conditions (as it is bound to be), and if they respond differently to those conditions, then each may be competitively superior in different environments. This too can express itself as a microhabitat differentiation, or a difference in geographic distribution or a temporal separation, depending on whether the appropriate conditions vary on a small spatial scale, a large spatial scale or over time. Of course, in a number of cases (especially with plants) it is not easy to distinguish between conditions and resources (see Chapter 3). Niches may then be differentiated on the basis of a factor (such as water) that is both a resource and a condition. We return to the spatial and temporal separation of niches when we discuss the role of competition in shaping community structure in Chapter 16.

8.6 Models of niche overlap

8.6.1 Combining niche overlap and competitive similarity – a route to 'neutral' coexistence

The consumer–resource models that we considered in the previous section share with the Lotka–Volterra model one key characteristic, namely a focus on the outcome of the interaction rather than the speed with which that outcome is attained. Hence, when we say that two species coexist, we mean that they persist indefinitely, whether at a simple equilibrium or with some more complex dynamics. But we can gain further insight if we switch the focus onto the speed of competitive exclusion. From this perspective, species pairs would not simply be divided into two groups: coexisting and not coexisting. Instead, they would be spread along a continuum from rapid to infinitely slow exclusion, which has the advantage of allowing us to acknowledge that very slow exclusion may be indistinguishable, in practice, from coexistence. The following simple consumer–resource competition model allows us to do this (Carmel *et al.*, 2017). It follows the same lines as those previously, but assumes, for simplicity, that both consumers suffer the same mortality rate, m (as they would, for example, in an aquatic 'chemostat' system, where liquid drains away at a constant rate), that the consumers convert resource into new consumers with perfect efficiency (the e's in Equations 8.19 and 8.20, for example, are equal to 1), that the input rate of the two resources into the system is the same, b, and that the natural loss rates of resources from the system are negligible compared with the losses to consumption. Note, though, that if these assumptions are relaxed, the important results are just the same, but not so clear. The model comprises equations for two consumers, A and B, and two resources, R_1 and R_2:

$$\frac{dA}{dt} = c_{A1}R_1A + c_{A2}R_2A - mA \tag{8.23}$$

$$\frac{dB}{dt} = c_{B1}R_1B + c_{B2}R_2B - mB \tag{8.24}$$

$$\frac{dR_1}{dt} = b - c_{A1}R_1A - c_{B1}R_1B \tag{8.25}$$

$$\frac{dR_2}{dt} = b - c_{A2}R_2A - c_{B2}R_2B. \tag{8.26}$$

We can then define *niche overlap* between the consumers as:

$$\frac{c_{A1}/c_{A2}}{c_{B1}/c_{B2}} \text{ or } \frac{c_{B1}/c_{B2}}{c_{A1}/c_{A2}},$$

whichever is lower.

This makes sense, as c_{A1}/c_{A2} and c_{B1}/c_{B2} represent the degree to which each species specialises on the two resources, such that if they show the same preference (same niche) the overlap takes its maximum value of 1, whereas if both species specialise on only one of the resources, then c_{A1} and c_{B2}, for example, are both zero, as is the niche overlap.

We can further define a new quantity, the *competitive similarity* between the two species, as:

$$\frac{c_{A1}c_{A2}}{c_{B1}c_{B2}} \text{ or } \frac{c_{B1}c_{B2}}{c_{A1}c_{A2}},$$

whichever is lower.

This, too, makes sense as $c_{A1}\,c_{A2}$, for example, reflects the competitive ability of species A integrated over the two resources, so that the competitive similarity between the two species would be 1 if their competitive abilities are the same, but would fall to a value close to zero if their abilities were very different.

The insight that we get from this and related models is that the rate of competitive exclusion depends on both the degree of niche overlap and the degree of competitive similarity, as set out in Figure 8.18. This rate of exclusion is at or close to its maximum when the degree of niche overlap is high and the species differ considerably in their competitive ability (zone A in Figure 8.18). This describes a strong competitor excluding a weak competitor in the absence of substantial niche differentiation, precisely as we have seen in the Lotka–Volterra and other consumer–resource models. Similarly, as before, when niche overlap is zero so that the species do not, in fact, compete, long-term coexistence is more or less guaranteed, even if the species differ markedly in their notional competitive ability (effectively zero similarity in competitive ability; zone B in Figure 8.18). But the overall picture now goes beyond this.

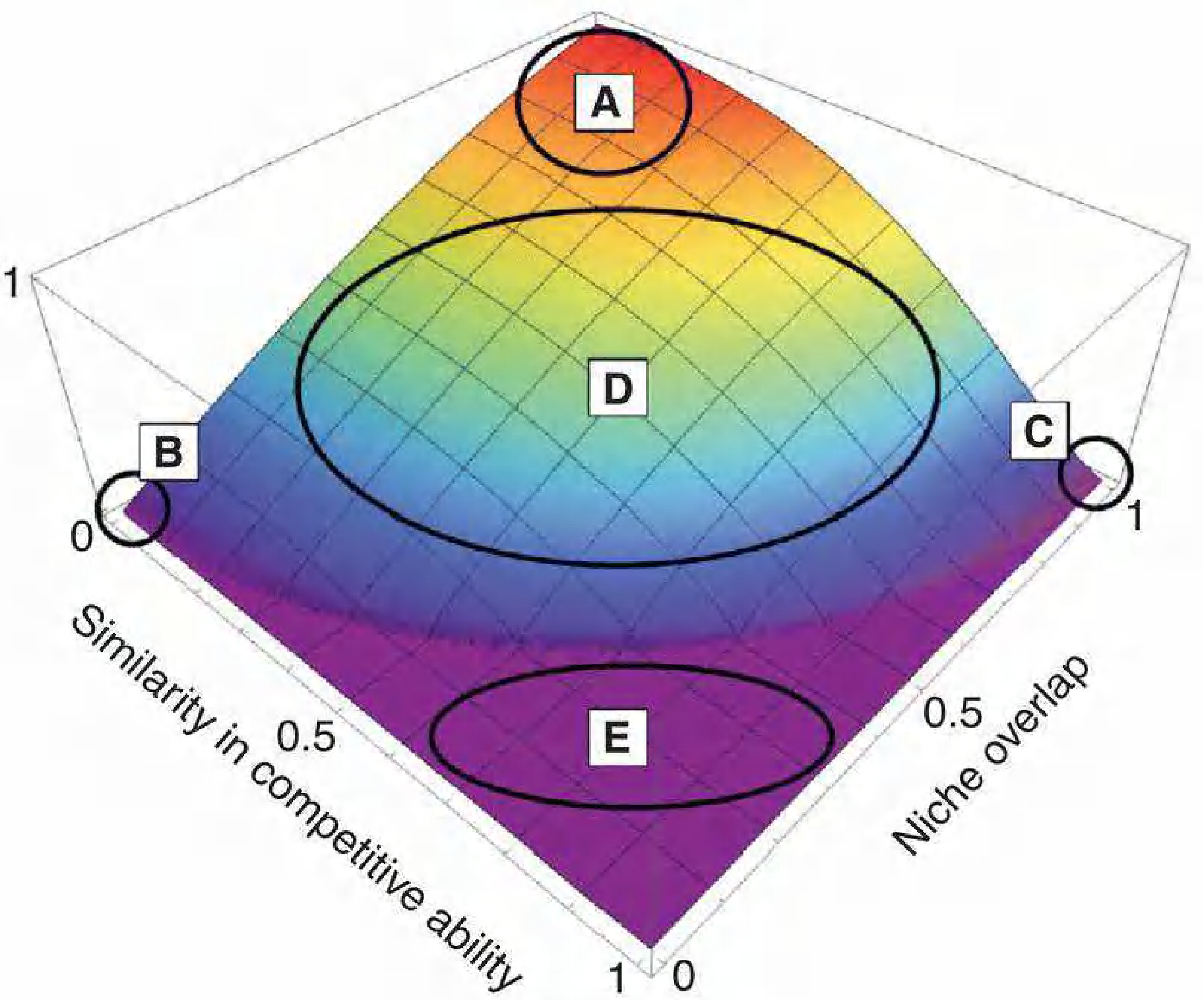


Figure 8.18 Competitor coexistence depends on both niche overlap and competitive similarity. Exclusion rate (indicated by the changing colours) of an inferior species – where values close to zero imply competitor coexistence – as a function both of niche overlap and the similarity in competitive ability for competition between two species governed by Equations 8.23–8.24. The regions, A–E are described in the text.

Source: After Carmel *et al*. (2016).

neutral competition

For example, suppose that niche overlap is complete (or nearly so) but the species are effectively the same in their competitive abilities (zone C in Figure 8.18). Previously we have concluded that one species must exclude the other, because only one competitor species can persist within a single undifferentiated niche. But now we can see that the rate of exclusion would be so slow (in the limit, zero) that these species would, in effect, coexist in spite of sharing the same niche. The scenario of species occupying the same niche and having equivalent competitive ability – put simply, being the same – is referred to as 'neutral' competition. It has received considerable attention in discussions of the role of interspecific competition in structuring communities (see especially Hubbell (2001)) and we will examine it in that context in Chapter 16. For now, though, we can see that the coexistence of two species that seem likely to compete with one another cannot, of itself, be taken as an indication that their niches must be differentiated from one another. Niche overlap provides the *potential* for competitive exclusion, but whether that potential is realised depends on the degree of competitive similarity between the species. From this new perspective, the transition from zone A through zones D and E in Figure 8.18 takes us from the Lotka–Volterra extremes of exclusion to coexistence, but now in a gradual way and with a dependence on both niche overlap and competitive similarity

8.6.2 A model of limiting similarity

It is easy to imagine that scientific progress is made by providing answers to questions. In fact, progress often consists of replacing one question with another, more pertinent, more challenging question. The relationship between niche overlap and coexistence is a case in point.

We have seen that the Lotka–Volterra model and the competitive exclusion principle imply that *any* amount of niche differentiation will allow the stable coexistence of competitors. Hence, in an attempt to discover whether this was 'true', the question 'do competing species need to be different in order to coexist stably?' greatly exercised the minds of ecologists through the middle of the 20th century (Kingsland, 1985). We can see now, however, that the question is badly put, since it leaves the precise meaning of 'different' undefined. Indeed, we saw above that 'coexistence' is itself a matter of degree, and that it depends on both the degree of niche overlap and the similarity of competitive abilities. A more pertinent question, therefore, would be 'how much niche differentiation is needed for stable coexistence?'

> how much niche differentiation is needed for coexistence?

One influential attempt to answer this question, based on the Lotka–Volterra model, was initiated by MacArthur and Levins (1967) and developed by May (1973). With hindsight, their approach is certainly open to question, but we can learn most about the 'limiting similarity problem' by first examining their approach and then looking at the objections to it. Here, as so often, the models can be instructive without being 'right'.

Imagine three species competing for a resource that is unidimensional and is distributed continuously; food size is a clear example. Each species has its own realised niche in this single dimension, which can be visualised as a resource utilisation curve (Figure 8.19). The consumption rate of each species is highest at the centre of its niche and tails off to zero at either end, and the more the utilisation curves of adjacent species overlap, the more the species compete. Indeed, by assuming that the curves are 'normal' distributions (in the statistical sense), and that the different species have similarly shaped curves, the Lotka–Volterra competition coefficient (applicable to both adjacent species) can be expressed by the following formula:

$$\alpha = e^{-d^2/4w^2} \tag{8.27}$$

where w is the standard deviation (or, roughly, 'relative width') of the curves, and d is the distance between the adjacent peaks. Thus, α is very small when there is considerable separation of adjacent curves ($d/w >> 1$; Figure 8.19a), and approaches unity as the curves themselves approach one another ($d/w < 1$; Figure 8.19b).

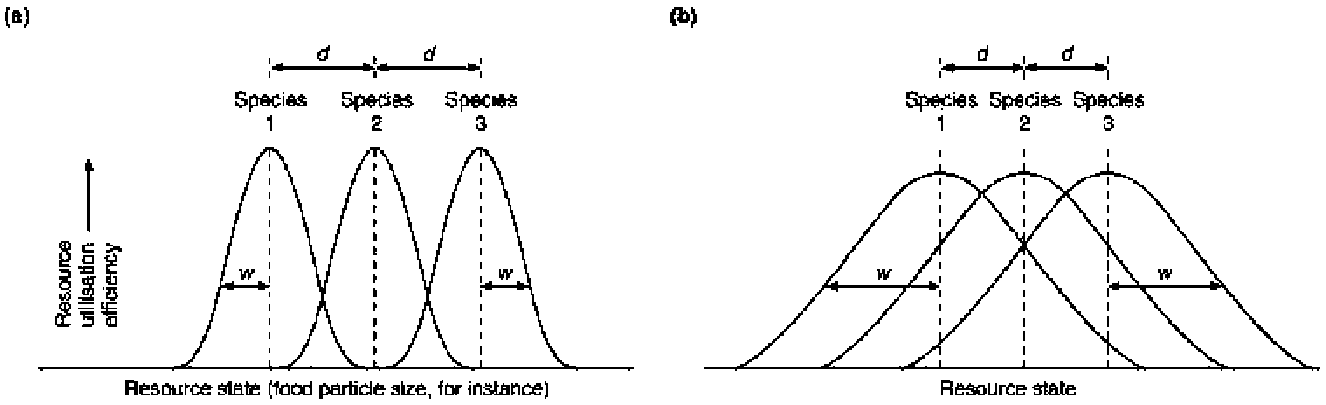


Figure 8.19 Resource utilisation curves for three species coexisting along a one-dimensional resource spectrum. *d* is the distance between adjacent curve peaks and *w* is the standard deviation of the curves. (a) Narrow niches with little overlap ($d > w$), i.e. relatively little interspecific competition. (b) Broader niches with greater overlap ($d < w$), i.e. relatively intense interspecific competition.

How much overlap of adjacent utilisation curves is compatible with stable coexistence? Assume that the two peripheral species have the same carrying capacity (K_1, representing the suitability of the available resources for species 1 and 3) and consider the coexistence, in between them on the resource axis, of another species (carrying capacity K_2). When d/w is low (α is high and the species are similar) the conditions for coexistence are extremely restrictive in terms of the $K_1 : K_2$ ratio; but these restrictions lift rapidly around the point at which, roughly, $d/w > 1$ (Figure 8.20). The conventional interpretation of this was that coexistence is possible when d/w is low, but only if the suitabilities of the environment for the different species are extremely finely balanced.

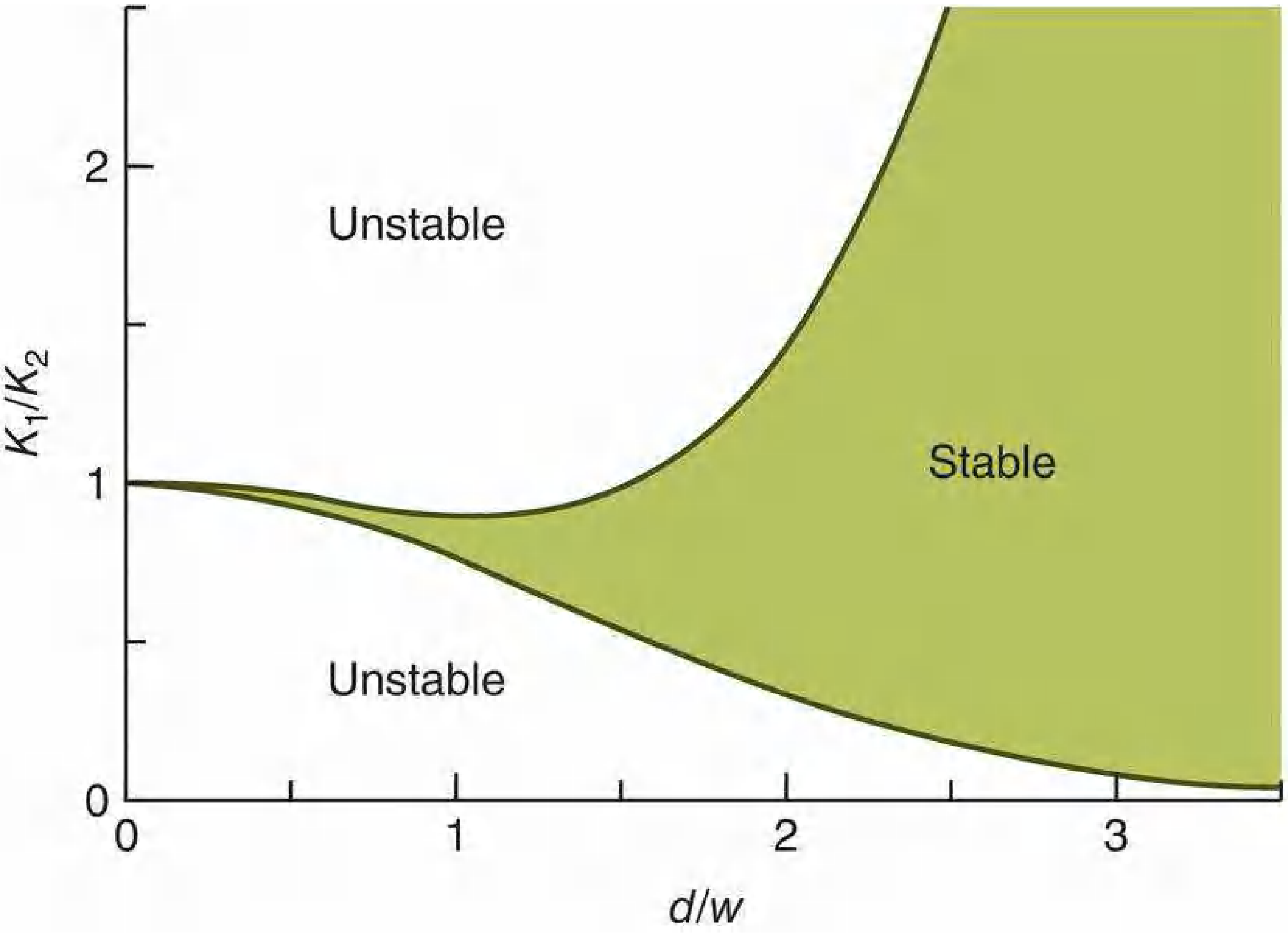


Figure 8.20 **The effect of niche similarity on competitor coexistence.** The range of habitat favourabilities (indicated by the carrying capacities K_1 and K_2, where $K_1 = K_3$) that permit a three-species equilibrium community with various degrees of niche overlap (d/w).

Source: After May (1973).

This model, then, suggests that there is a limit to the similarity of coexisting competitors, and that the limit is given by the condition $d/w > 1$. Are these the correct answers? In fact, it seems most unlikely that there is a universal limit to similarity, or even a widely applicable one that we could express in such a simple way as $d/w > 1$. Abrams (1976 , 1983), amongst others, has emphasised that models with competition in several dimensions, with alternative utilisation curves and so on, all lead to alternative limits to similarity, and often to much lower values of d/w being compatible with robust, stable coexistence. And, to reiterate, we saw above that coexistence depends on both the degree of niche overlap and the similarity in competitive abilities (implied by the narrowing, rather than the disappearance, of the stable region in Figure 8.20). We can see, therefore, that the question ‘do coexisting competitors need to be different?’ is unhelpful, and that the answer to the question ‘is there one identifiable limit to the similarity of coexisting competitors?’ is ‘no’. But it is reasonable to ask whether there are patterns in the similarities of coexisting competitors and whether we can make sense of these patterns, and we return to these questions in Section 16.2 when we look at the role of interspecific competition in structuring whole communities.

8.7 Heterogeneity, colonisation and pre-emptive competition

We have been assuming up to now that the environment is sufficiently constant for the outcome of competition to be determined by the degree of niche overlap and the competitive abilities of the competing species. In reality, such situations are far from universal. Environments are usually a patchwork of favourable and unfavourable habitats; patches are often only available temporarily;

and patches often appear at unpredictable times and in unpredictable places. Even when interspecific competition occurs, it does not necessarily continue to completion. Hence, systems do not necessarily reach equilibrium, and superior competitors do not necessarily have time to exclude their inferiors. Thus, an understanding of interspecific competition itself is not always enough. It is often also necessary to consider how interspecific competition is influenced by, and interacts with, an inconstant or unpredictable environment.

8.7.1 Unpredictable gaps: the poorer competitor is a better coloniser

'Gaps' of unoccupied space occur unpredictably in many environments. Fires, landslips and lightning can create gaps in woodlands; storm-force seas can create gaps on the shore; and voracious predators can create gaps almost anywhere. Invariably, these gaps are recolonised. But the first species to do so is not necessarily the one that is best able to exclude other species in the long term. Thus, so long as gaps are created at the appropriate frequency, it is possible for a 'fugitive' species and a highly competitive species to coexist. The fugitive species tends to be the first to colonise gaps; it establishes itself, and it reproduces. The other species tends to be slower to invade the gaps, but having begun to do so, it outcompetes and eventually excludes the fugitive from that particular gap.

One example of this is provided by four species of ant that occupy acacia trees in Kenya: *Crematogaster sjostedti*, *C. mimosae*, *C. nigriceps* and *Tetraponera penzigi*. The species are intolerant of one another. Fewer than 1% of trees are occupied by more than one species, and these cohabitations are only transient. Nonetheless, they coexist on a very fine spatial scale: all four are likely to be found within any 100 m^2 area. In battles amongst them for the occupation of trees, there is a clear competitive hierarchy: *C. sjostedti* > *C. mimosae* > *C. nigriceps* > *T. penzigi* (Palmer *et al.*, 2000). However, trees are also naturally and continually becoming available for recolonisation, for example following fire or elephant feeding or drought, and if we look at their abilities to colonise these newly available trees, by the expansion of mature colonies nearby, we see a pattern very different from the competitive hierarchy (Figure 8.21a), with *C. nigriceps* being the most effective, followed by *C. mimosae*. Moreover, the effectiveness of colonising these trees from a greater distance, with new 'foundress' queen ants, shows a perfect reversal of the competitive hierarchy: *T. penzigi* > *C. nigriceps* > *C. mimosae* > *C. sjostedti* (Figure 8.21b).

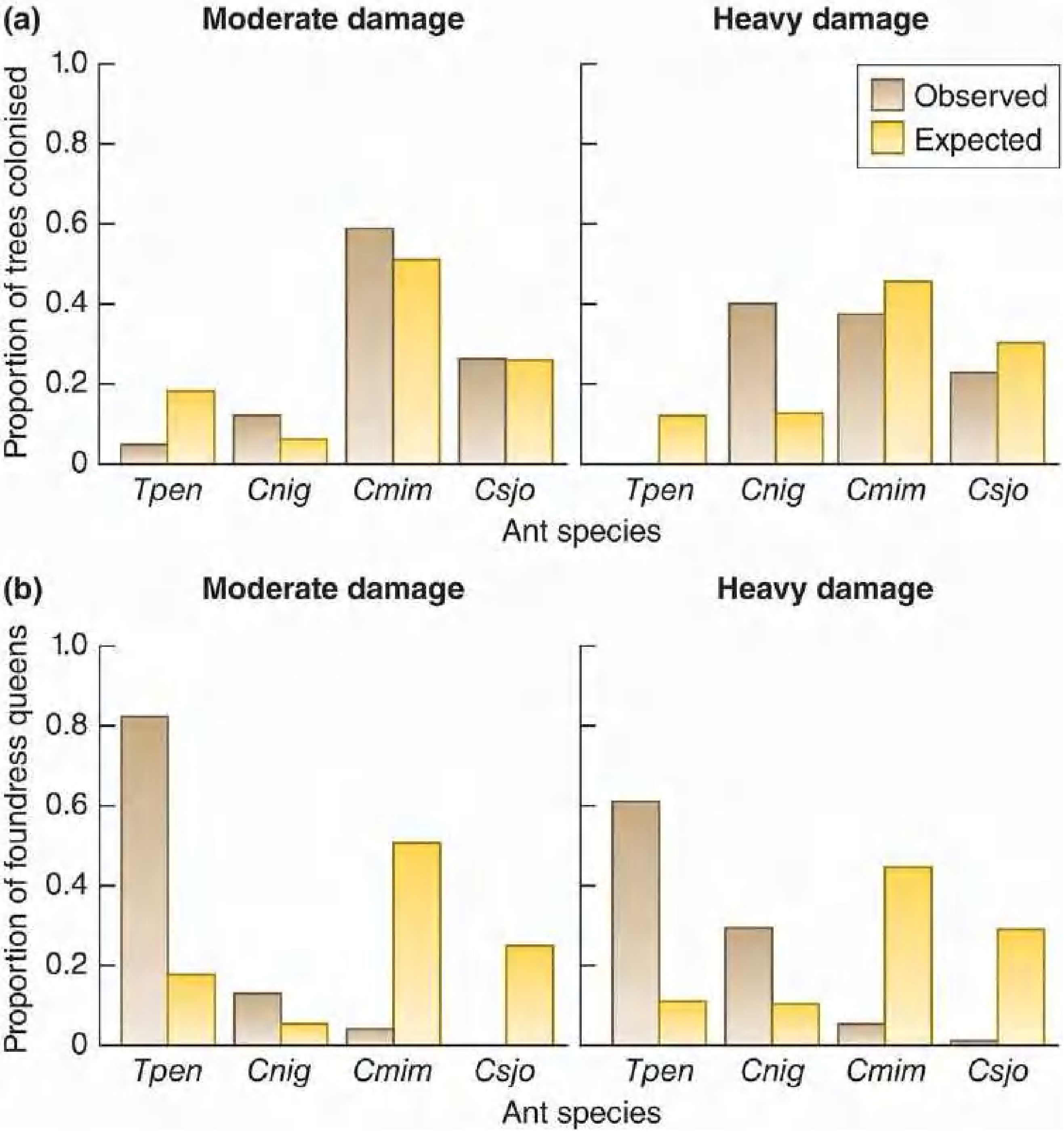


Figure 8.21 Ant species in Kenya vary greatly in their colonisation ability, allowing good competitors and good colonisers to coexist. The colonisation of damaged acacia trees, *Acacia drepanolobium*, by four species of acacia ants: *Tetraponera penzigi* (*Tpen*), *Crematogaster nigriceps* (*Cnig*), *C. mimosa* (*Cmim*) and *C. sjostedti* (*Csjo*), where damage was either moderate (ants absent, but plant growth intact) or heavy (ants absent and plants recovering from damage by regrowth). (a) Colonisation by expansion of nearby colonies. Observed proportions (out of 51 and 32 for moderately and heavily damaged tees, respectively) are compared with proportions expected from occupancies of trees in the near locality. (b) Colonisation by foundress queens. Observed proportions (out of 247 and 285 for moderately and heavily damaged trees, respectively) are compared with proportions expected from occupancies of trees in the near locality.

Source: After Stanton *et al.* (2002).

Thus, if new trees never became available, either as a result of damage, or simply via recruitment from saplings, it seems clear that *C. sjostedti* would steadily outcompete and ultimately exclude the other species. But because the species that were best in direct competition were worst at colonisation, and because trees were continuously becoming available for colonisation, all four species were able to coexist.

8.7.2 Unpredictable gaps: the pre-emption of space

Even in the absence of clear differences in colonisation ability, there may be a priority effect in determining the outcome of a competitive interaction. When two species compete from a common starting point, the result is usually predictable. But in the colonisation of unoccupied space, competition typically has a staggered start. Individuals of one species are likely to arrive, or germinate from the seed bank, in advance of individuals of another species. This, in itself, may be enough to tip the competitive balance in favour of the first species. If space is pre-empted by different species in different gaps, then this may allow coexistence, even though one species would always exclude the other if they competed 'on equal terms'.

first come, best served

We see this, for example, for two species of reef fish, *Thalassoma hardwicke* and *T. quinquevittatum*, in Figure 8.22. In experiments, repeated several times, three individuals of one of the species were introduced to a patch of reef occupied by three individuals of the other species, either more or less simultaneously or after five or 12 days. When the two species arrived at the same time, *T. quinquevittatum* was the superior competitor. But when the arrival was staggered, the competitive balance was either accentuated, or when *T. hardwicke* was the resident, reversed. There was a clear priority effect: whichever species arrived first fared best.

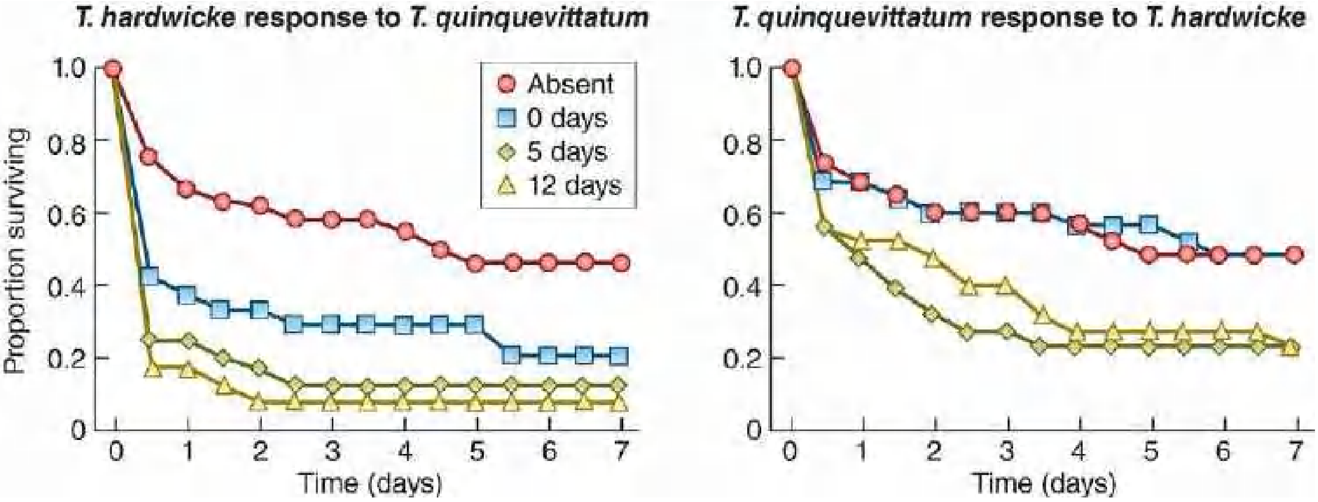


Figure 8.22 Competitive outcome determined by a priority effect. Survival time in two competing reef fish, *Thalassoma hardwicke* and *T. quinquevittatum* in Moorea, French Polynesia, when alone (red circle), when they arrived at the same time at a site (blue square) and when one arrived either 5 days or 12 days after the other.

Source: After Geange & Stier (2009).

In a review of priority effects, in interspecific competition and more generally, Fukami (2015) makes the useful distinction between niche pre-emption and niche modification. When species interactions are driven by the former, competitors may coexist, as we saw for the fish above, especially when their competitive abilities are themselves similar. With niche modification, by contrast, the species arriving first does not simply diminish the value of the niche for the late arriver, but changes it qualitatively. Hence, it may improve the habitat for subsequent species, for example by adding nutrients to the soil or new microhabitats. We discuss such facilitation in Chapter 13. Alternatively, it may deny late arrivers a niche altogether and so outcompete them.

8.7.3 Fluctuating environments

’the paradox of the plankton’ and storage effects

The argument that competitors may coexist because they live in a fluctuating environment that first favours one species, then the other, then the first, and so on, goes back at least as far as Hutchinson’s (1961) attempt to explain the ‘paradox of the plankton’. Chesson (2000), in particular, elaborated on the idea by stressing that these fluctuations may arise in two ways. They may be generated internally within the system by differences in the non-linear responses of the consumers to the resources and of the resources to the consumers, such that fluctuations of each are perpetuated within the system. This, in essence, is what we saw in the model of Huisman and Weissing (1999) in Figure 8.13c. Alternatively, the fluctuations may be in the environment external to and independent of the interactions within the system, such that there is a ‘storage effect’ with each species accumulating sufficient abundance through the good times to see it through the bad.

These effects, and their outcomes, are easier to describe than to demonstrate definitively with real data, but there are at least some supportive examples. We see one if we turn, once again, to an experimental system of diatoms we examined briefly in Figure 5.1 (Figure 8.23). *Cyclotella pseudostelligera* and *Fragilaria crotonensis* were maintained at a range of temperatures, and their numbers monitored, along with concentrations of their limiting resource, silicate, in order to parameterise a consumer–resource competition model similar to that described above, in Equations 8.18–8.20. The model was then used to simulate competition between them at 18°C, where *Cyclotella* was predicted to exclude *Fragilaria* (Figure 8.23a), at 24°C, where *Fragilaria* was predicted to exclude *Cyclotella* (Figure 8.23b), and with temperatures switching between 18 and 24 every 12 hours. Here, the two species were predicted to coexist, with the competitive balance shifting between the species (Figure 8.23c). These predictions were then supported by competition experiments between the species run over a number of days, representative replicates of which are also shown in Figure 8.23.

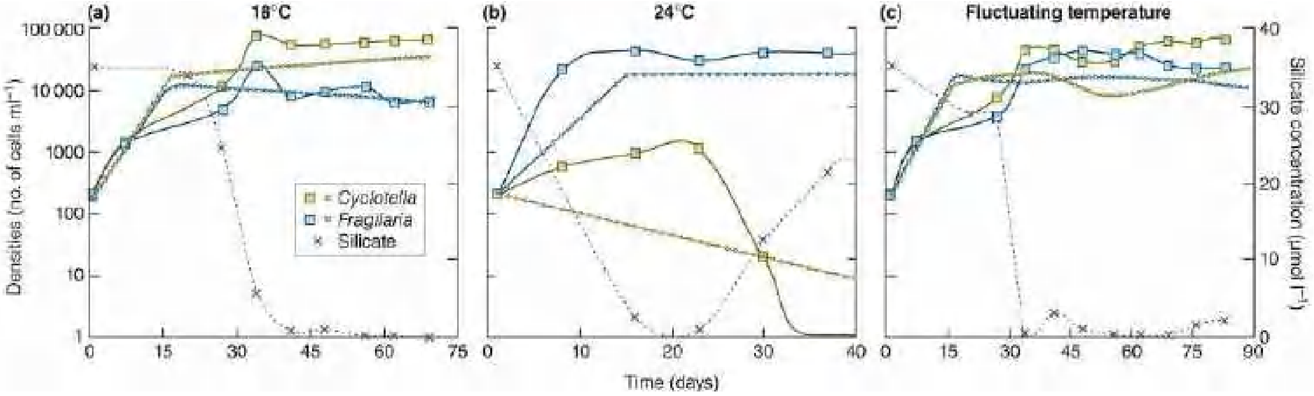


Figure 8.23 Diatom species coexist as a result of fluctuations in their environment. Representative replicates of competitive interactions between the diatoms *Cyclotella* (large green squares) and *Fragilaria* (large blue squares), competing for silicate. Also shown are the outcomes predicted by consumer–resource competition models parameterised for the system (small blue and green squares for *Cyclotella* and *Fragilaria*, respectively). (a) Run at 18°C. (b) Run at 24°C. (c) Run at temperatures switching between 18°C and 24°C every 12 hours.

Source: After Descamps-Julien & Gonzalez (2005).

Of the possible routes from fluctuation to coexistence in this case, there was no direct evidence for a role for interacting non-linear interactions, in that there was no clear pattern of correlated fluctuations in the silicate concentrations. However, the idea of there being a storage effect was supported, for example, by the 'compensatory dynamics' of the two species (when one went up, the other went down) both in the model output and in the real data (Figure 8.23c).

A supportive example from field data is provided by a study of two closely related shrubs, *Atriplex acanthocarpa* and *A. canescens*, coexisting in the Chihuahuan Desert in Mexico, where the most important fluctuating resource is water, arriving in rare and torrential pulses (Verhulst *et al,* 2008). *A. acanthocarpa* is usually found in dense patches consisting of thousands of small plants, many of which are dying or already dead. *A. canescens*, by contrast, is usually larger, much less inclined to be found in clumps, and without concentrations of dead or dying plants. Field monitoring showed that *A. canescens* had higher longevity and a longer generation time (more delayed reproduction); its survivorship followed a type 3 curve (Section 4.6.2), mortality declining with age, whereas for *A. acanthocarpa* mortality was relatively constant with age (a type 2 curve). Crucially, though, the species also differed in the responses of these demographic variables to water availability. Overall, in years of good recruitment, with plenty of water, *A. acanthocarpa*'s rate of population increase was greater than that of *A. canescens*. But in drier years, when both species declined in abundance, this rate of decline was also greater in *A. acanthocarpa* than in *A. canescens*. Thus, as water availability fluctuated, the more opportunistic *A. acanthocarpa* 'stored' abundance in the periods of plenty, while the more tolerant *A. canescens* maintained more of its stores in the periods of scarcity. As a result, they coexisted, and between them accounted for more than 90% of the vegetation cover in the study area.

8.7.4 Aggregated distributions

a clumped superior competitor adversely affects itself and leaves gaps for its inferior

A more subtle path to the coexistence of a superior and an inferior competitor on a patchy and ephemeral resource is based on the idea that the two species may have independent, aggregated (i.e. clumped) distributions over the available patches. This would mean that the powers of the superior competitor were mostly directed against members of its own species (in the high-density clumps), but that this aggregated superior competitor would be absent from many patches within which the inferior competitor could escape competition. An inferior competitor may then be able to coexist with a superior competitor that would rapidly exclude it from a continuous, homogeneous environment. Note, however, that whilst such coexistence of competitors has nothing to do with niche differentiation, it is linked to it by a common theme – that of species competing more frequently and intensively intraspecifically than they do interspecifically. Niche differentiation is one means by which this can occur, but temporary aggregations can give rise to the same phenomenon – even for the inferior competitor.

marine invertebrates in a field experiment

Certainly we can see this working in models (see, for example, the simple mathematical models of Atkinson and Shorrocks (1981) or Dieckmann *et al.* (2000) or the spatially explicit, patch dynamics model of Silvertown *et al.* (1992)). But there are also an increasing number of supportive examples from field or experimental systems. In one example, competition was followed amongst four species of marine invertebrates that settled naturally on artificial floating panels in Queensland, Australia: two encrusting bryozoan species, *Watersipora subtorquata* and *Celloporaria* sp., an ascidian, *Didemnum* sp., and an unidentified microcionid sponge. These were then transplanted such that their positions on such panels were either randomised or aggregated. The results are shown in Figure 8.24. When they were randomised, the ascidian was clearly the superior competitor, continuing to expand its coverage over the 28 days for which the panels were monitored. At the same time, the sponge fared reasonably well for two weeks but then stopped growing, while the two bryozoans also grew well for two weeks but then declined to near extinction. When the individuals within each species were aggregated, however, the strong competitor, the ascidian, had suffered a noticeable reduction in growth by the end, while two of the inferior competitors had either continued to grow well into the second half of the experiment (the sponge) or at least failed to suffer the same rate of decline (*Celloporaria*). Again, aggregation was favouring the coexistence of what would otherwise have been winners and losers in the competitive process.

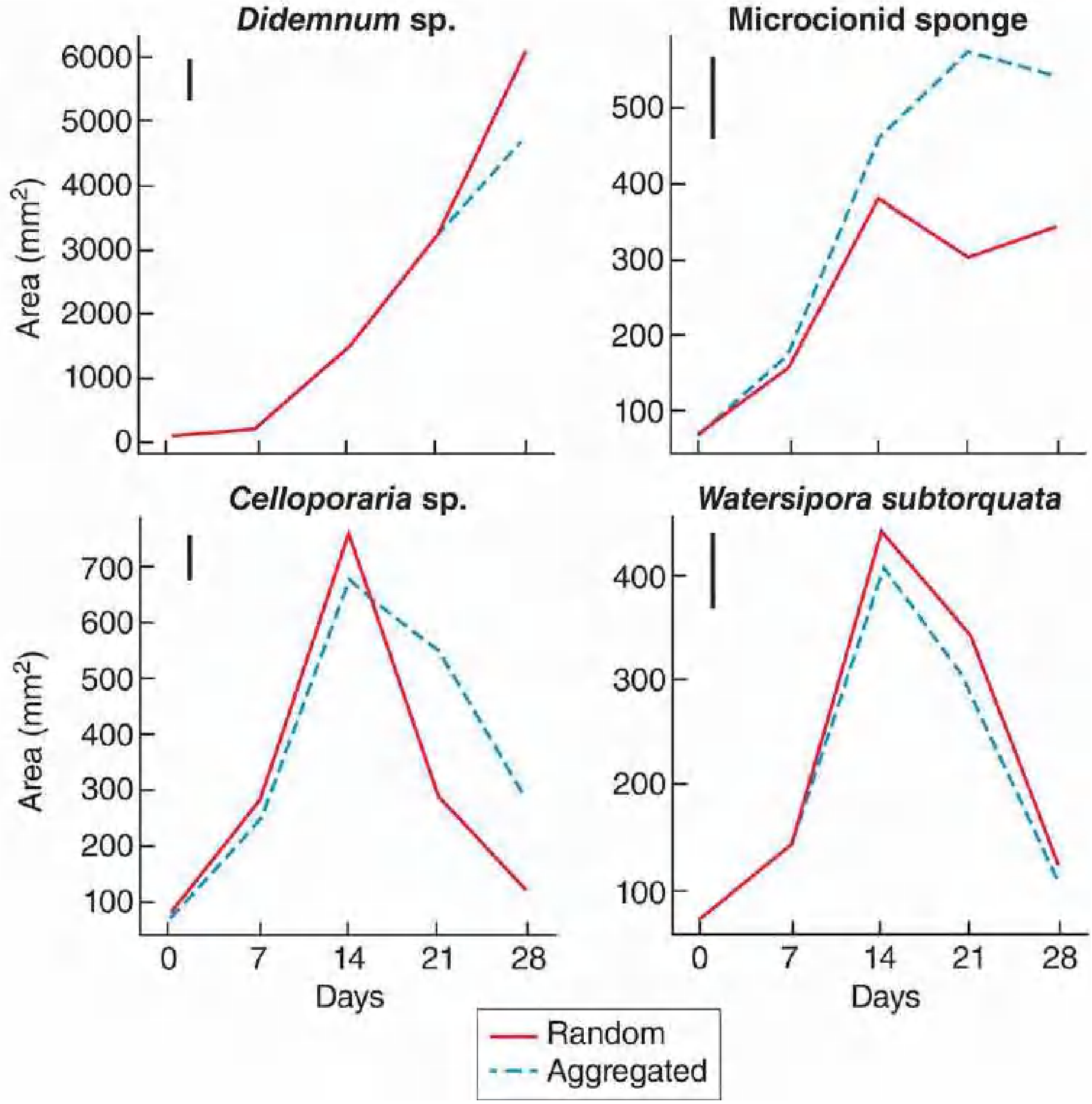


Figure 8.24 Coexistence of four marine invertebrate species is enhanced by aggregation. The total areas occupied by four marine species, as indicated, when they competed on floating panels in Queensland, Australia, having been established initially in a random or aggregated fashion. 95% error bars are shown to the top left of each panel.

Source: After Hart & Marshall (2009).

APPLICATION 8.1 Encouraging biodiversity in field margins

In an interesting application of this principle, Wassmuth *et al.* (2009) examined how best to protect and enhance wild plant diversity in field margins, where there are European agri-environment schemes in place encouraging farmers to do this. The problem is that sowing mixtures of wild plants in these margins often results in disappointingly low levels of biodiversity because of the competitive dominance of a few common species. They therefore looked at competitive outcomes amongst six wild annual species, chosen following recommendations for seed mixtures to be used in field margins, when their seeds were either dispersed randomly or in intraspecific aggregations. Plants were harvested during their flowering period, around 10 weeks after sowing. For all six species, the numbers harvested were greatest following the aggregated sowings (Figure 8.25a).

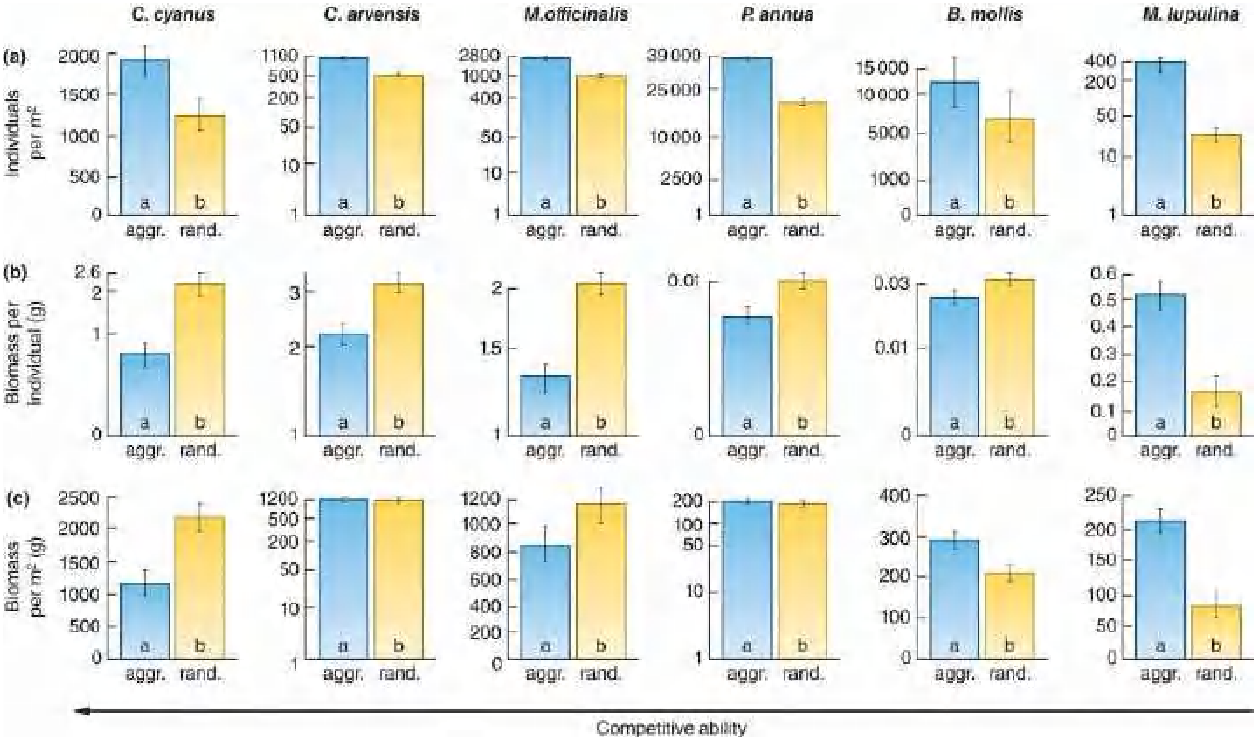


Figure 8.25 Coexistence of plant species in agri-environment schemes is enhanced by aggregation. The performances of six species of annual weeds when sown with their seeds aggregated (aggr.) or dispersed randomly (rand.). The species, arranged in order of their competitive ability from left (most competitive) to right (least) were *Centaurea cyanus*, *Calendula arvensis*, *Melilotus officinalis*, *Poa annua*, *Bromus mollis* and *Medicago lupulina*. (a) Numbers of individuals per m^2. (b) Biomass per individual. (c) Biomass per m^2. In each case, the values are the means after combining plants grown in monocultures and three- and six-species mixtures, but the patterns were consistent throughout. Bars are SEs. Different letters indicate significant differences ($P < 0.05$).

Source: After Wassmuth *et al.* (2009).

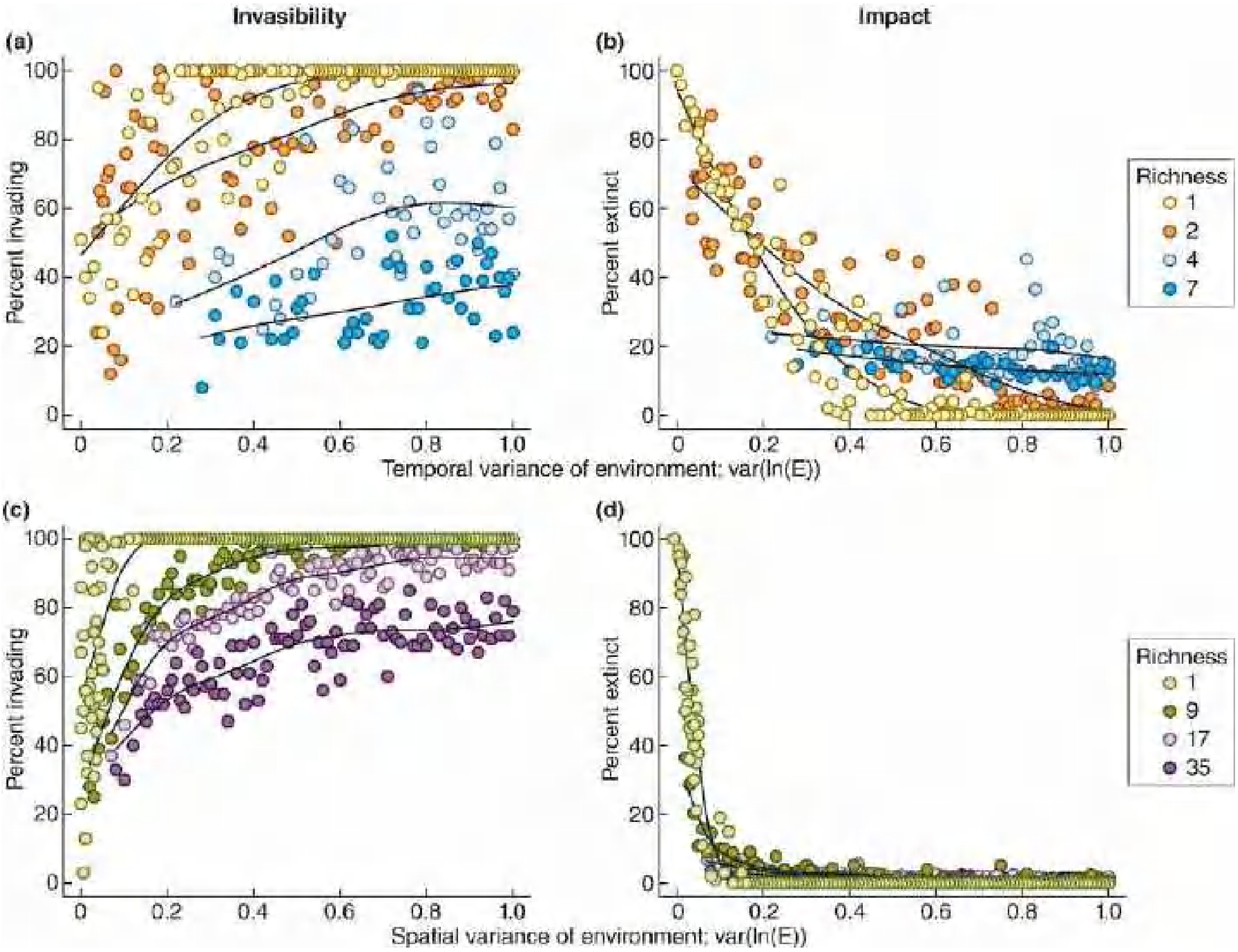
The role of competition in this – and in particular the shifting balance between intra- and interspecific competition with the switch from random to aggregated sowing – is apparent if we look at the sizes of individuals and the overall biomass in the contrasting treatments (Figure 8.25b, c). There was a clear trend for the more competitive species to be smaller and the least competitive species larger following the aggregated sowings, and hence for the more competitive species to contribute less and the least competitive species more to the overall biomass. Thus Wassmuth and her colleagues argue that agri-environment schemes designed to preserve and enhance biodiversity should consider spatial seeding strategies enhancing survival of less competitive, endangered plant species, and that new agricultural sowing technologies allowing intraspecific aggregation may be an important tool for the diversification of field boundary strips, fallows, and other elements of agricultural landscapes.

Repeatedly in this section, then, the heterogeneous nature of the environment can be seen to have fostered coexistence without there being a marked differentiation of niches. A realistic view of interspecific competition, therefore, must acknowledge that it often proceeds not in isolation, but under the influence of, and within the constraints of, a patchy, impermanent or unpredictable world. This reinforces a point that recurs throughout this book: heterogeneity (spatial, temporal or individual) can have a stabilising influence on ecological interactions.

heterogeneity often stabilises

APPLICATION 8.2 Invasion and heterogeneity; invasibility and impact

Alien invaders are an important problem for the conservation of communities (which they may alter) and individual species (which they may replace). Understanding what makes invasion more likely, and its effects more profound, may be critical in helping conservation biologists plan their resistance to such invasions. The consequences of any potential invasion are inevitably some combination of species exclusion (including competitive exclusion) and coexistence. Thus, the link between the ecology of invasion and of coexistence is clear (Shea & Chesson, 2002). The general tendency that we have seen for heterogeneity to enhance stability – stable coexistence in the present case – therefore prompted Melbourne *et al.* (2007) to ask whether environmental heterogeneity was important in determining the success of invasive species in establishing themselves in their new environment, and in determining their impact on the local community following the invasion. They investigated a simple model that determined a species' chance of invasion by whether its long-term rate of increase was positive, and measured its impact as the percentage of the resident community that became extinct following its invasion. These were then related to a combination of, first, the long-term average difference in the fitness of the invader and the resident community, and second, the level of variability in fitness, reflecting variability of the environment, which gave rise to a storage effect that could overcome any deficit in fitness. Without examining the model in detail, we can look at their results, summarised in Figure 8.26. Whether the variability was spatial or temporal, environmental variation increased the chances of invasion (it allowed invaders to establish when they might otherwise have been excluded) but then diminished the impact of an invasion (it allowed invaders and residents to coexist).



[**Figure 8.26**](#) **Environmental heterogeneity can enhance the coexistence of competitors through its effects on species invasion.** The effect of temporal variation ((a, b)) and spatial variation ((c, d)) in the environment (expressed as the variance in the natural log of species' birth rate, *E*) on the invasibility of model communities ((a, c)) and the impact on the model community following invasion ((b, d)). Results are shown for native 'communities' of 1 to 35 species ('richess'). In each case, 100 model communities were assembled and invasion of each was simulated 100 times. The percentage of these for which invasion was successful, and the mean percentage of species that were driven to extinction by the invasion, was then noted.

Source: After Melbourne *et al*. ([2007](#)).

Their results were in line with those of other, related modelling studies, and with the limited amount of available data. Invasions, for example, are often observed to give rise to a net increase in species diversity. But as the authors themselves remark, initial establishment is only the first step in the process of invasion. For successful invasion a species must also be able to spread once it has created its bridgehead, and this spread, too, will be influenced by environmental heterogeneity. Developing an understanding of the ecology of the whole process of invasion is an important challenge for the future.

8.8 Apparent competition: enemy-free space

two prey species attacked by a predator are, in essence, indistinguishable from two consumer species competing for a resource

We have seen that interspecific competition, pared down to its essentials, is an interaction between two species sharing a common resource. But imagine, instead, two species of prey, or host, attacked by a common predator or parasite. Both prey species are harmed by the predator, and the predator benefits from both species of prey. Hence, the increase in abundance that the predator achieves by consuming prey 1 increases the harm it does to prey 2. Indirectly, therefore, prey 1 adversely affects prey 2 and vice versa. These interactions are summarised in Figure 8.27, which shows that from the point of view of the two prey species, the signs of the interactions between them are indistinguishable from those that would apply to the indirect interaction of two species competing for a single resource (exploitation competition), except that in this case there appears to be no limiting resource. Holt (1977, 1984) therefore called such interactions 'apparent competition'. To strengthen the link further with conventional competition, Jeffries and Lawton (1985) referred to it as 'competition for enemy-free space', in that the persistence of prey species 1 will be favoured by avoiding attacks from the predator, which also attacks prey species 2. Hence, prey 1 can achieve this by occupying a habitat, or adopting a morphology or a behavioural pattern, that is sufficiently different from that of prey 2. In short, 'being different' will once again favour coexistence – but the difference will be not in the resource niche but in the way in which the predator is encountered or avoided.

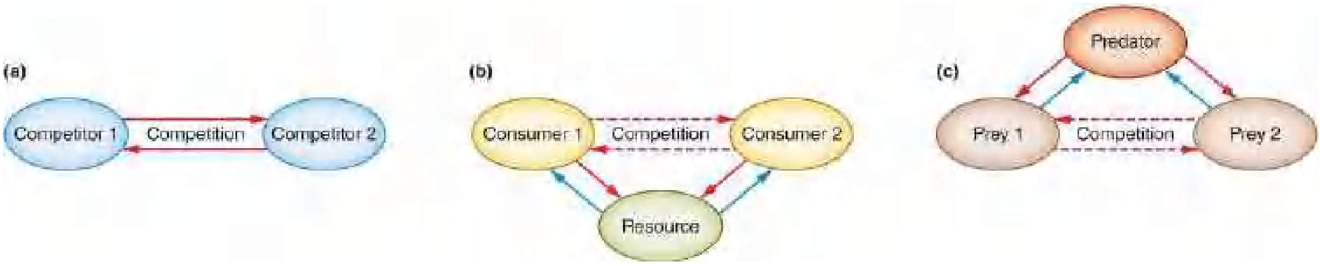


Figure 8.27 In terms of the signs of their interactions, competition and apparent competition are indistinguishable from one another. (a) Two competing species interfering directly (interference competition). (b) Two species consuming a common resource (exploitation competition). (c) Two species being attacked by a common predator ('apparent competition' for 'enemy-free space'). Red arrows indicate a negative effect, blue arrows a positive effect. Solid arrows indicate a direct effect, dashed arrows an indirect effect.

evidence for apparent competition in gastropods, bivalves and their predators

An early experimental demonstration of apparent competition for enemy-free space examined two groups of prey living on subtidal rocky reefs at Santa Catalina Island, California, USA (Figure 8.28). The first comprised three species of mobile gastropods; the second comprised sessile bivalves, dominated by the clam, *Chama arcana*. Both groups were preyed on by a lobster, an octopus and a whelk, although these predators showed a marked preference for the bivalves. In areas characterised by large boulders and much crevice space ('high relief') there were high densities of bivalves and predators, but only moderate densities of gastropods; whereas in low relief areas largely lacking crevice space ('cobble fields') there were apparently no bivalves, only a few predators but high densities of gastropods. Hence, the densities of the two prey groups were inversely correlated, but there was little in their feeding biology to suggest that they were competing for a shared food resource. On the other hand, when bivalves were experimentally introduced into cobble-field areas, the number of predators congregating there increased, the mortality rates of the gastropods increased, and the densities of the gastropods declined (Figure 8.28a, b). Increases in one prey group led to increased numbers of predators and a decline in the other prey group.

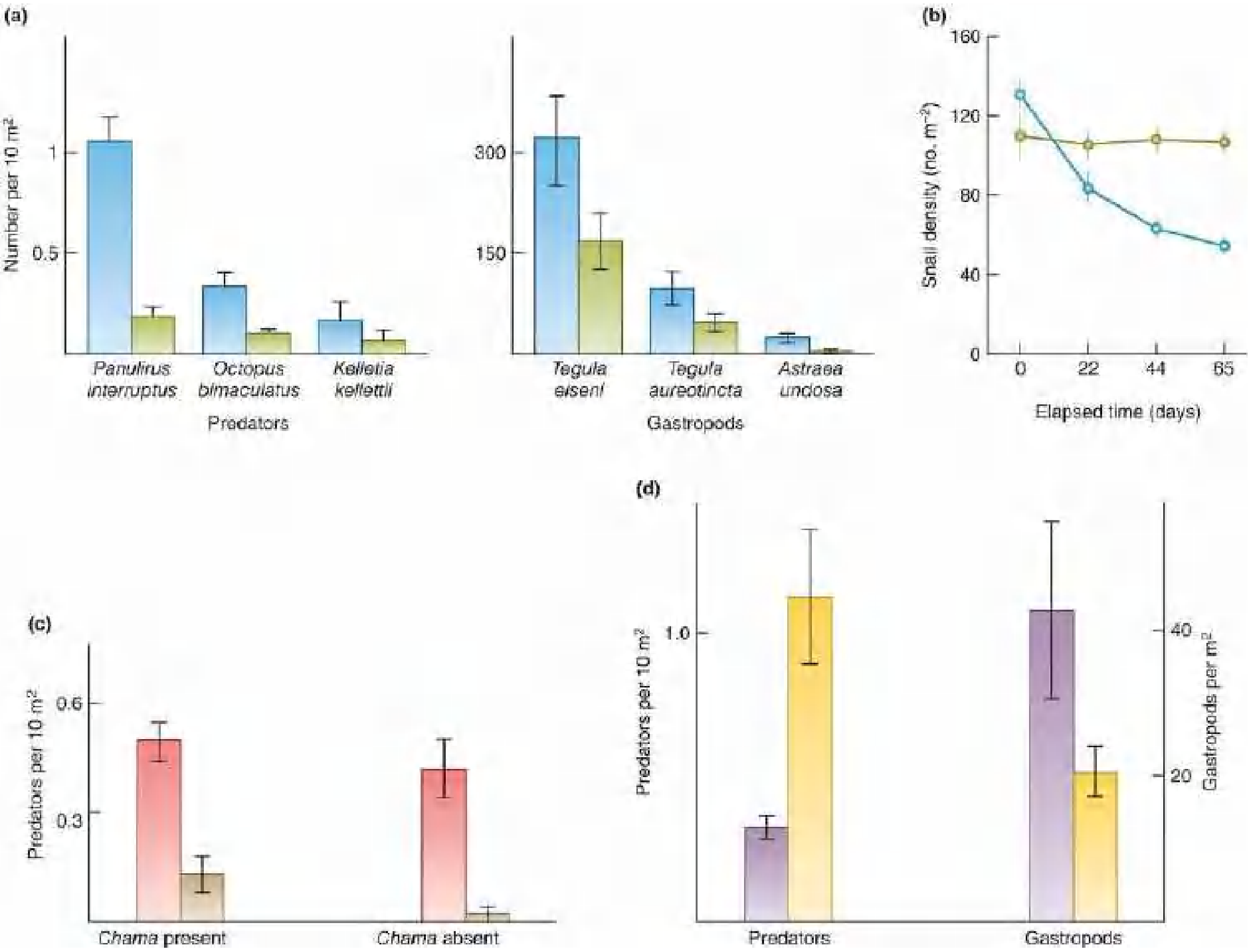


Figure 8.28 **Evidence for apparent competition for predator-free space at Santa Catalina Island, USA.** Three species of predators, a lobster (*Panulirus interruptus*), an octopus (*Octopus bimaculatus*) and a whelk (*Kelletia kelletii*), preyed on three species of mobile gastropods, *Tegula aureotincta*, *T. eiseni* and *Astraea undosa*, and on sessile bivalves, dominated by the clam *Chama arcana*. (a) Predator density (number per 10 m^2) and gastropod mortality increased (number of 'newly dead' shells per site) when bivalves were added to gastropod-dominated cobble sites (blue bars) relative to controls (green bars). (b) This led to a decline in gastropod density. (c) Predator density was higher (number per 10 m^2) at high (red bars) than at low (brown bars) gastropod-density cobble sites, both in the presence and absence of *Chama*. (d) Densities of predators were lower (number per 10 m^2) and densities of gastropods higher (number per m^2) at high-relief sites without *Chama* (purple bars) than at those with (yellow bars). Bars are SEs.

Source: After Schmitt (1987).

An exact reciprocation of the manipulation proved impossible, because the (mobile) gastropods were too inclined to move away once introduced, but cobble sites with naturally high densities of gastropods supported higher densities of predators, and had higher mortality rates of experimentally added bivalves than did sites with relatively low densities of gastropods (Figure 8.28c). And on the rare high relief sites without *Chama* bivalves, predator densities were lower, and gastropod densities higher, than was normally the case (Figure 8.28d). It seems clear that each prey group adversely affected the other through an increased number of predators, and hence increased predator-induced mortality.

In fact, as with so many subdivisions in ecology, the line between exploitation and apparent competition can be a blurred one. In Section 8.2.6, we examined competition between orange-crowned and Virginia warblers. We described it as competition for preferred nest sites (a resource they exploit) but pointed out that the main benefit of these was reduced loss of nestlings to predators. A good case could also be made, therefore, for calling this apparent competition – or for accepting that we apply classifications of the natural world because they can be useful, not because the world is that simple.

APPLICATION 8.3 Apparent competition threatens an endangered plant

Apparent competition may also be the means by which invasive species harm natives. In the coastal dune grasslands of northern California, for example, the European beachgrass, *Ammophila arenaria,* was introduced in the mid-1800s to help stabilise the dune systems, but it is particularly attractive to the seed-eating deer mouse, *Peromyscus maniculatus*, which attains higher densities and seed consumption rates in the vicinity of the grass. This in turn poses a threat to native species, including the endangered lupine, *Lupinus tidestromii*, since deer mice are major predators of their seeds. Indeed, in a study in Marin County, California, USA, Dangremond *et al.* (2010) showed that the probability of lupine seed consumption increased the closer the plant was to beachgrass plants, whether observed naturally or as a result of experimentally placing lupines at different distances from the grasses. This, and other demographic measurements, allowed them to estimate the rates of population increase of the lupines at their three field sites (Figure 8.29).

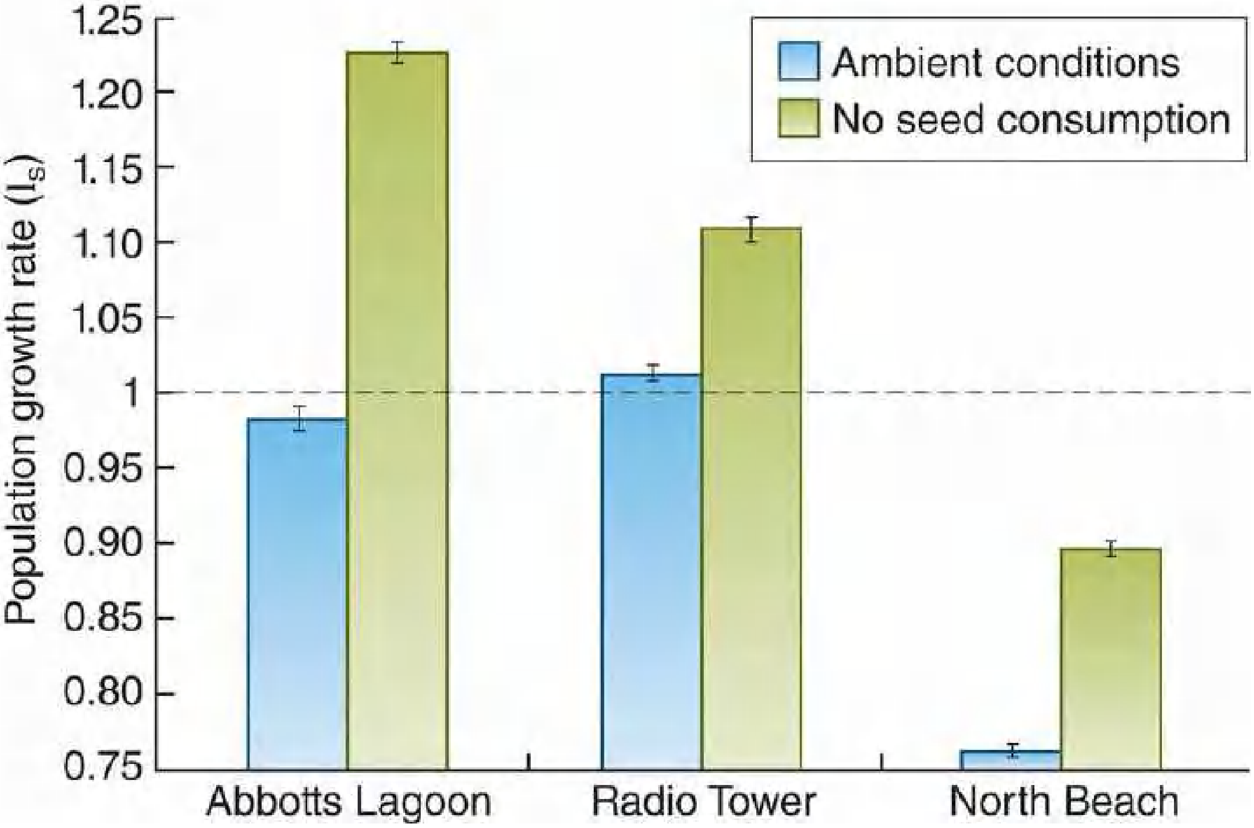


Figure 8.29 Apparent competition from beachgrass threatens lupine conservation. The estimated population growth rates in three populations of the lupine, *Lupinus tidestromii*, in California, USA, both naturally (where there is seed consumption by deer mice attracted to the site by the invasive grass, *Ammophila arenaria*) and on the assumption that there was no such seed consumption. Growth rates greater than one indicate a population than can increase in size; rates less than one indicate population decline. Bars are 95% CIs.

Source: After Dangremond *et al.* (2010).

Even if we ignore apparent competition, *A. arenaria* outcompetes *L. tidestromii.* Nonetheless, at all three sites seed predation by the mice, and hence apparent competition, had a serious effect on the lupine's rate of population increase. The likely longer-term consequences, though, differed between the sites (Figure 8.29). One, North Beach, seemed doomed to decline rapidly even in the absence of seed predation. A second, Radio Tower, appeared capable of sustaining itself even in the face of seed predation. But for the third, Abbotts Lagoon, seed predation took the population from sustainability into decline. For the local conservation of the lupine, control of the beachgrass is clearly essential – but not only those plants that compete with the lupine directly. Even at a distance, the grass will attract mice that may push the lupines over the edge into decline and extinction.

a reappraisal of plant competition

Taking a much broader view, Connell (1990) carried out a particularly revealing reappraisal of 54 published examples of field experiments on 'competition' in plants, where the original authors claimed to have demonstrated conventional interspecific competition in 50. In fact, in many of these, insufficient information had been collected to distinguish between conventional and apparent competition; and in a number of others, the information was available but ambiguous. For example, one study showed that removal of *Artemisia* bushes from a large site in Arizona led to much better growth of 22 species of herb. This was originally interpreted in terms of exploitative competition for water (Robertson, 1947): removal of *Artemisia* made more water available to the herbs. However, in the absence of *Artemisia*, they also experienced greatly reduced grazing pressure from deer, rodents and insects, for which the *Artemisia* bushes were not only a source of food but also a place of shelter. The outcome is therefore equally likely to have resulted from reduced apparent competition.

distinguishing pattern and process

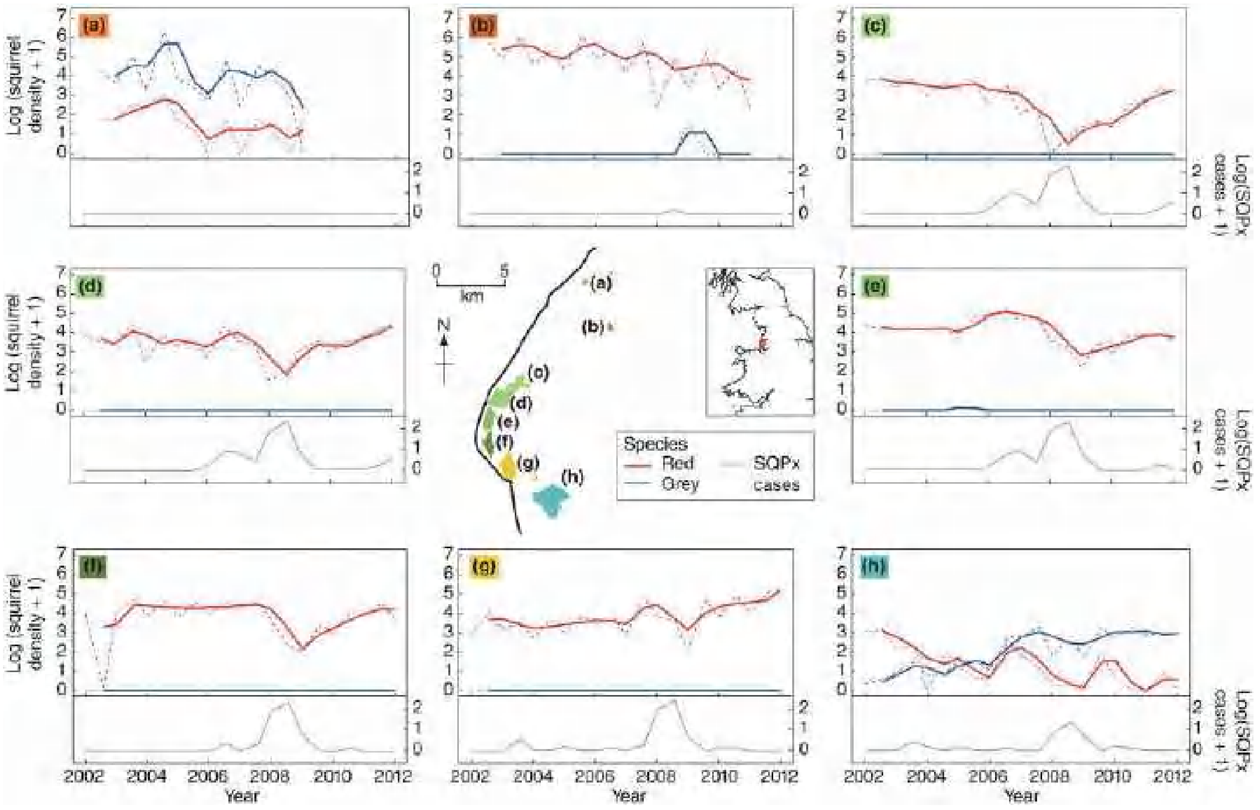
This emphasises that the relative neglect of apparent competition in the past has been unwarranted, but also re-emphasises that the distinction is important within interspecific competition between pattern on the one hand, and process or mechanism on the other. In the past, patterns of niche differentiation, and also of increased abundance of one species in the absence of another, have been interpreted as evidence of competition too readily. Now we can see that such patterns can arise through a wide variety of processes, and that a proper understanding requires that we study both the processes and the patterns.

APPLICATION 8.4 Red and grey squirrels, and squirrelpox virus

Apparent competition may equally be driven not only by shared predators but by shared parasites or pathogens. Indeed, many pathogens are relatively benign in their natural hosts but can be highly virulent when transmitted to other species which they more rarely, or have not previously, infected. We see this in the particular case of *zoonotic* pathogens: infectious agents that circulate naturally in a wildlife reservoir, where they do little or no harm, but can also be passed on to humans where they may cause serious disease or even be fatal. Examples include rabies, HIV, Ebola and COVID-19. We look at zoonotic infections again in Chapter 12. Mostly these do not involve apparent competition, but in those cases that do, shared pathogens can pose a conservation threat, as in the case of red and grey squirrels in the UK and their shared squirrelpox virus.

The red squirrel (*Sciurus vulgaris*) is native to the UK and many other parts of Europe. The grey squirrel (*S. carolinensis*) is a North American species that was introduced to the UK on several occasions from the late 1800s. Since that time, and at an accelerating rate, the grey squirrel has displaced the red, spreading northwards into northern England and Scotland, leaving only a few residual populations of red squirrels in southern and central England. For many years this was attributed to the larger size and more aggressive nature of the grey squirrel, and so to the conventional scenario of an invasive species outcompeting a resident. Since the middle of the 20th century, however, it has become increasingly apparent that conventional competition is only a part, and probably a small part, of the story. The grey squirrel is the host of squirrelpox virus, though it is rare to see any symptoms of infection in greys or to detect any ill effects. The virus, though, can be passed from greys to reds, and transmitted between reds, and in these cases the infection is frequently fatal, preceded by weeping sores on the skin, lethargy and reduced appetite.

Nonetheless, it is not easy to demonstrate the effects of this apparent competition at the population level – the transmission of squirrelpox virus from greys to reds in a region and the subsequent decline in the population of reds. An example, though, is shown in Figure 8.30. The Sefton coast, just north of Liverpool in north-west England. is home to one of the few remaining sets of red squirrel populations in southern Britain. There, they are surrounded by grey squirrels that are the subject of a control programme, which is only partially successful in the sense that contact between reds and greys is greatly reduced but not eliminated. Figure 8.30, then, shows estimates of the numbers of red squirrels in a series of nearby populations from 2002 to 2012. In some of these they coexisted directly with grey squirrels, and in each population the appearance of squirrelpox virus cases was noted.



[Figure 8.30](#) **Red squirrel populations are threatened by squirrelpox (SQPx) virus infection that they acquire from invasive grey squirrels.** The central insets show the study area in north-west England and the location of eight study populations. The surrounding plots show the time series of the numbers of both squirrel species and of squirrelpox cases in red squirrels, as indicated, counted at twice-yearly standard transects converted to densities by accounting for the area surveyed. Dashed lines are the individual estimates, solid lines the two-point moving averages. The numbers of red squirrelpox cases come from postmortem examination of carcasses submitted for examination and cover locations within 1500 m of the study area.

Source: After Chantrey *et al*. ([2014](#)).

Almost every time there was a decline in the numbers of reds, this was accompanied, and usually preceded, by the appearance of squirrelpox. In many cases there were no grey squirrels seen in the immediate vicinity – their role, it seems, is simply to introduce squirrelpox, following which transmission amongst reds can sustain the infection. Moreover, there is the suggestion of a wave of infection spreading through the populations – starting in the more central populations in late 2005 but not reaching the more peripheral populations until around two years later. It is also apparent, though, that having suffered their decline, most of the populations showed clear signs of recovery after the infection had passed through. If there is unrestricted contact between red and grey squirrels, then the elimination of reds locally, as a result of apparent competition mediated by squirrelpox virus, seems inevitable. But control of greys, and of the transmission of squirrelpox virus, seems capable, given sufficient effort, of saving the remaining reds even if it does not turn back the grey tide.

8.9 Ecological effects of interspecific competition: experimental approaches

We have discussed a number of reasons why interspecific competition may be easier to describe in theory than to demonstrate in practice. Coexistence may be the result of evolutionary avoidance in the past rather than ecological dynamics in the present; coexisting species may have niches that are differentiated from one another, but even in the absence of differentiation, they may coexist more or less indefinitely if their competitive abilities are closely matched; and so on. It is therefore important that we appreciate the range of approaches that have been taken in seeking evidence for competition in natural systems, and the pros and cons of those various approaches.

We have already noted in particular the difficulties in interpreting merely observational evidence, and it is for this reason that many studies have taken an experimental approach. For example, we have seen manipulative field experiments involving birds (Section 8.2.6), sea urchins and fish (Section 8.2.7), corals and seaweeds (Section 8.3.3) and gastropods and bivalves (Section 8.8), where the density of one or both species was altered. The fecundity, the survivorship, the abundance or the resource utilisation of the remaining species is subsequently monitored, and then compared either with the situation prior to the manipulation, or, far better, with a comparable control plot in which no manipulation has occurred. Such experiments have consistently provided valuable information, but they are typically easier to perform on some types of organism (e.g. those that are sessile) than they are on others.

The second type of experimental evidence has come from work carried out under artificial, controlled (often laboratory) conditions. Again, the crucial element is usually a comparison between the responses of species living alone and in combination. Such experiments have the advantage of being comparatively easy to perform and control, but they also have major disadvantages. Firstly, species are examined in environments that are different from those they experience naturally. Also, the simplicity of the environment may preclude niche differentiation because niche dimensions are missing that would otherwise be important. Nevertheless, these experiments can provide useful clues to the likely effects of competition in nature.

longer-term experiments

The most direct way of discovering the outcome of competition between two species under controlled conditions is to put them together and leave them to it. However, even the most one-sided competition is likely to take a few generations (or a reasonable period of modular growth) before it is completed, and this direct approach is therefore easier in some groups than in others. It has most frequently been applied to aquatic microorganisms (such as the phytoplankton species in Section 8.2.1) and to insects (such as the flour beetles in Section 8.4.2), but neither higher plants, nor vertebrates, nor large invertebrates lend themselves so readily to this approach (although we saw a plant example in Section 8.2.2). This may bias our view of the nature of interspecific competition.

single-generation experiments: substitutive experiments

Given these problems, the alternative approach in controlled settings, especially with plants, has often been to follow populations over just a single generation, comparing ‘inputs’ and ‘outputs’. A number of experimental designs have been used. In ‘substitutive’ experiments (Jolliffe, 2000), the effect of varying the proportion of each of two species is explored whilst keeping overall density constant. The amount of seed or the biomass of each species in each mixture is then monitored at the end of the experimental period. Such ‘replacement series’ may then be established at a range of total densities. In practice, however, most workers have used only a single total density, which is problematic, since it means that the effect of competition over several generations – when total density would inevitably alter – cannot be predicted (see Firbank & Watkinson, 1990).

Nonetheless, replacement series have often proved valuable. An early, influential study was that of de Wit *et al.* (1966) on competition between the grass *Panicum maximum* and the legume *Glycine javanica*, which often form mixtures in Australian pastures. *Panicum* acquires its nitrogen only from the soil, but *Glycine* acquires part of its nitrogen from the air, by nitrogen fixation, through its root association with the bacterium *Rhizobium* (see Section 13.11). The competitors were grown in replacement series with and without an inoculation of *Rhizobium*, which should provide *Glycine* with a resource (nitrogen in the air) unavailable to *Panicum*. The results are given both as replacement series and as 'relative yield totals' (Figure 8.31). The relative yield of a species is the ratio of its yield in the mixture to its yield alone in the replacement series, controlling for any absolute yield differences between the species. The relative yield total for a mixture is then the sum of the two relative yields. The replacement series (Figure 8.31a) show that both species, but especially *Glycine*, fared better (were less affected by interspecific competition) in the presence of *Rhizobium*. There is more to be learned, however, from examining the relative yield totals (Figure 8.31b). In the absence of *Rhizobium*, these never departed significantly from 1, indicating that a second species could only be accommodated by a compensatory reduction in the output of the first. But in the presence of *Rhizobium* the relative yield totals consistently exceeded 1, suggesting that niche differentiation was now possible such that the species yielded more between them than either could alone.

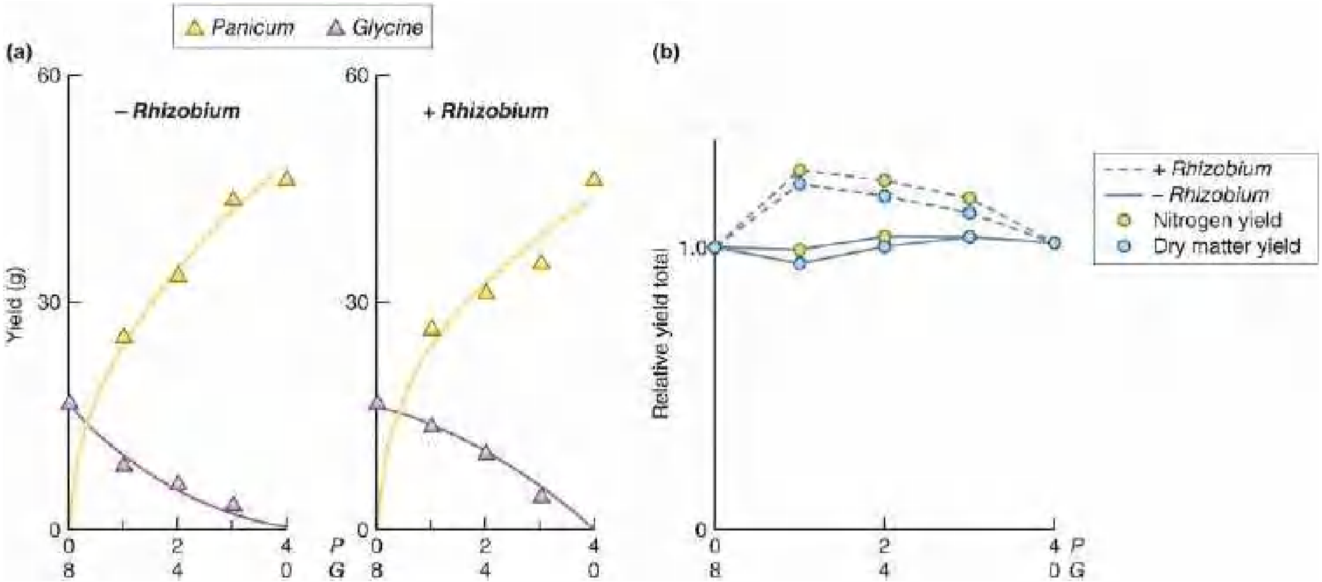


Figure 8.31 A substitutive experiment on interspecific competition between *Panicum maximum* (*P*) and *Glycine javanica* (*G*), in the presence and absence of *Rhizobium*: (a) replacement diagrams; (b) relative yield totals.

Source: After de Wit *et al.* (1966).

A substitutive design proved convenient, too, in seeking to understand the extent to which interactions between microbial species in natural communities were competitive or mutualistic (beneficial to both species – see Chapter 13). Thus, Foster and Bell (2012) isolated 72 bacterial species from water-filled tree-holes in a beech woodland, and then cultured them for seven days in a natural medium as the 72 monocultures and 180 randomly selected species-pairs, with five replicates each, established with an initial inoculation that always contained the same number of bacterial cells (hence the design was substitutive). The results were then expressed as the relationship between the observed overall productivity and the productivity that would be predicted simply by adding the productivities of the two species in monoculture (Figure 8.32). For a few, the observed value exceeded that predicted (indicative of a mutualistic interaction); for the vast majority the observed value was significantly lower than the prediction. For this natural community of microbes, at least, competition appears to be the dominant interaction between them.

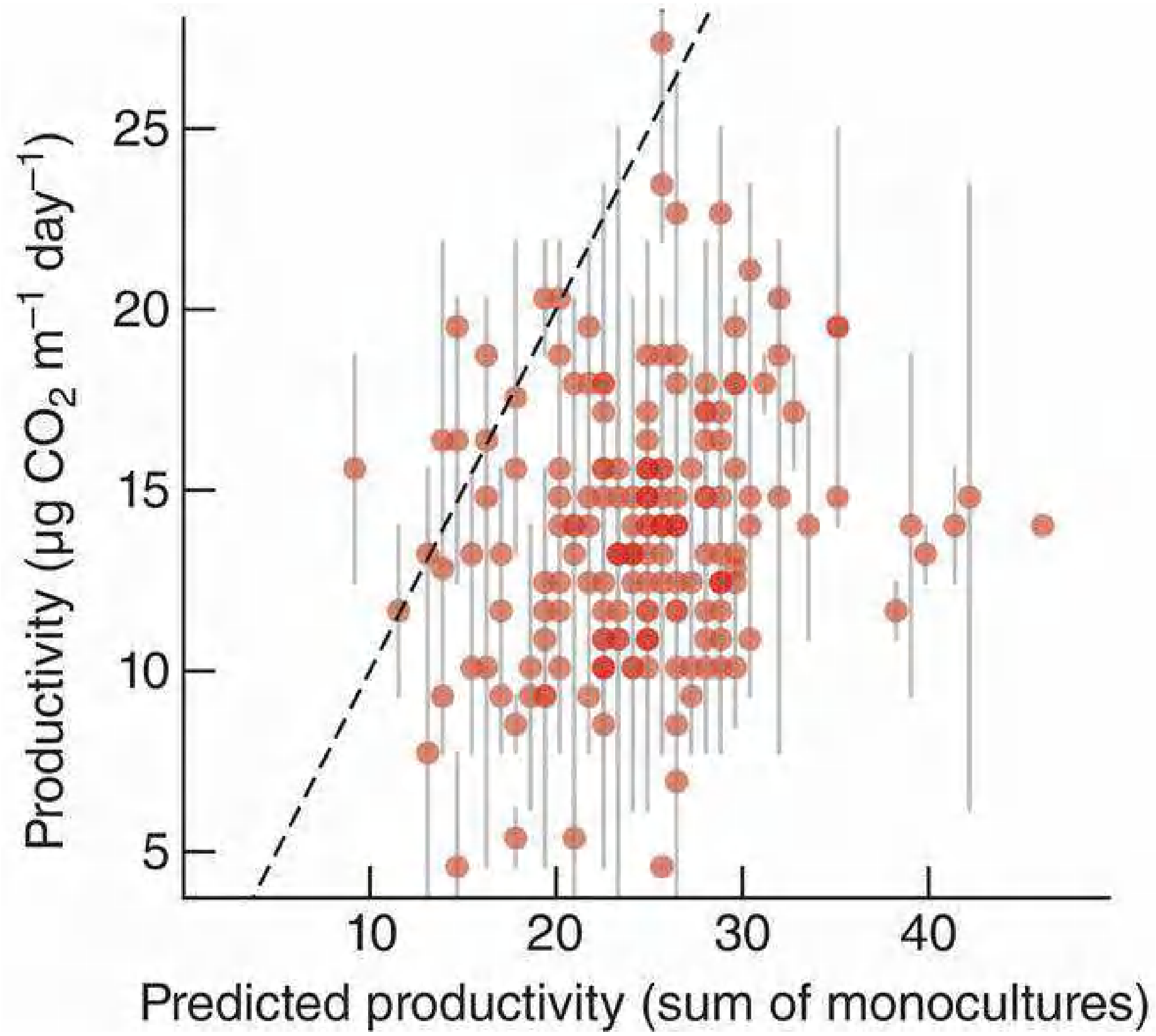


Figure 8.32 A substitutive design demonstrates competition among microbes in water-filled tree-holes. Observed and predicted productivities (measured as total respiration: μg CO_2 ml^{-1} day^{-1}) for 180 pairs of bacterial species isolated from natural tree-holes and competed against one another under natural conditions in a substitutive design experiment. (Shading is darker where points overlap.)The predicted productivities represent the sum of those achieved when each species in grown in monoculture. The black dashed line shows where observed and predicted probabilities are equal. Bars are SEs.

Source: After Foster & Bell (2012).

APPLICATION 8.5 Additive experiments help guide prairie restoration

A second popular approach in the past has been the use of an ‘additive’ design, in which one species (typically a crop) is sown at a constant density, along with a range of densities of one or more additional species (typically weeds). The justification for this is that it mimics the natural situation of a crop infested by weeds, and it therefore provides information on the likely effect on the crop of various levels of infestation (Firbank & Watkinson, 1990). A problem with additive experiments, however, is that overall density and species proportion are changed simultaneously. It has therefore proved difficult to separate the effect of the weeds themselves on crop yield from the simple effect of increasing total density (crop plus weeds). Nonetheless, the approach can prove useful, and an example is shown in Figure 8.33, describing the effects of additional species on the ‘target’ species, little bluestem, *Schizachyrium scoparium*, a grass growing right across the Great Plains of the USA, often used in prairie restoration. It was sown either alone, or in seed mixtures that contained either one other of three further species used in prairie restoration, or with all three of those species, or with additional native species not usually used in restoration. The establishment and spread of the target species were then followed over seven years. The percentage cover of little bluestem showed some tendency to decline when additional species were added to the sowing mixture, but the effect was not significant, and its cover seemed to recover to maximum levels in the richest mixtures, perhaps because the other species were competing more with one another than with little bluestem. Certainly, the experiment did not suggest that little bluestem needs to be ‘protected’ from competition with other species when it is used in prairie restoration.

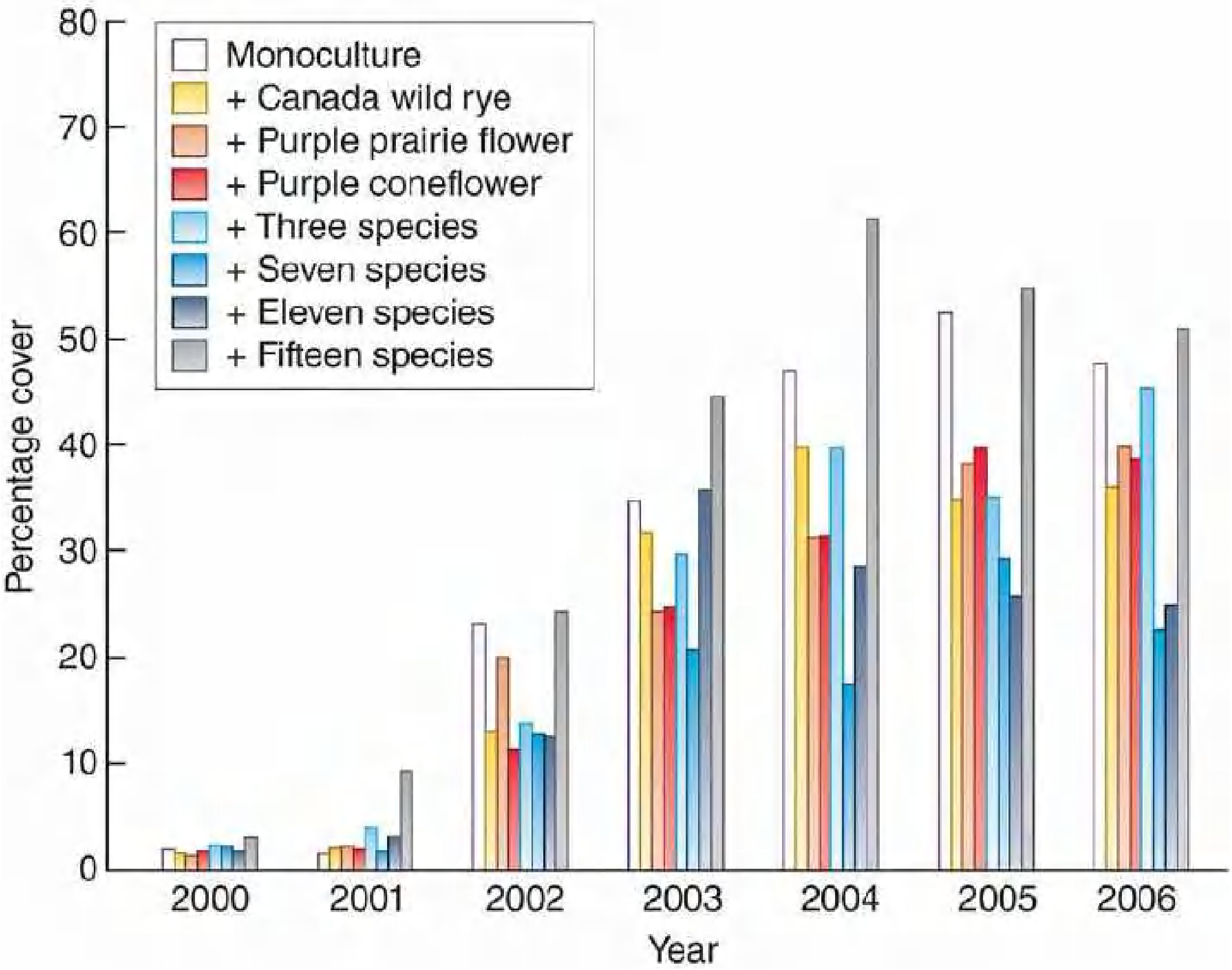


Figure 8.33 An additive design demonstrates little effect of other species on little bluestem. The percentage cover of the grass, little bluestem, *Schizacyrium scoparium*, sown in plots in Kansas in 2000 and followed for seven years, alone, and with addition of seeds of the species indicated. The three species mixture consisted of the three species listed above; the richer mixtures contained these plus other native species. Percentage cover did not differ significantly amongst the treatments.

Source: After Piper (2007).

APPLICATION 8.6 Response surface analysis of the consequences of intercropping

In substitutive designs the proportions of competitors are varied but total density is held constant, whilst in additive designs the proportions are varied but the density of one competitor is held constant. To try to overcome these problems, in a ‘response surface’ design, the densities of both species are varied independently, though this design has been used less often, perhaps, as suggested by Neumann *et al.* (2009), because of the difficulties of analysing the three dimensional surfaces generated, to which appropriate sets of competition equations have to be fitted. One example, though, is provided by the work of Neumann *et al.* (2009) themselves, who applied the approach in the context of ‘intercropping’, the cultivation of two crops in the same field with the aim of maximising overall output in a sustainable manner. Their two species were peas, *Pisum sativum*, and oats, *Avena sativa*, grown in experimental plots in Germany in alternating 7.8 m-long rows, 0.122 m apart, at densities of 40, 60, 70 and 80 seeds per m^2 for peas and 38, 75, 150 and 300 seeds per m^2 for oats in each of two years, 2002 and 2003. Results were expressed in terms of the nitrogen yield of the grain of both species at the end of the season, since this nitrogen yield is of particular commercial interest in the face of the ever-increasing costs of nitrogen fertilisers (Figure 8.34).

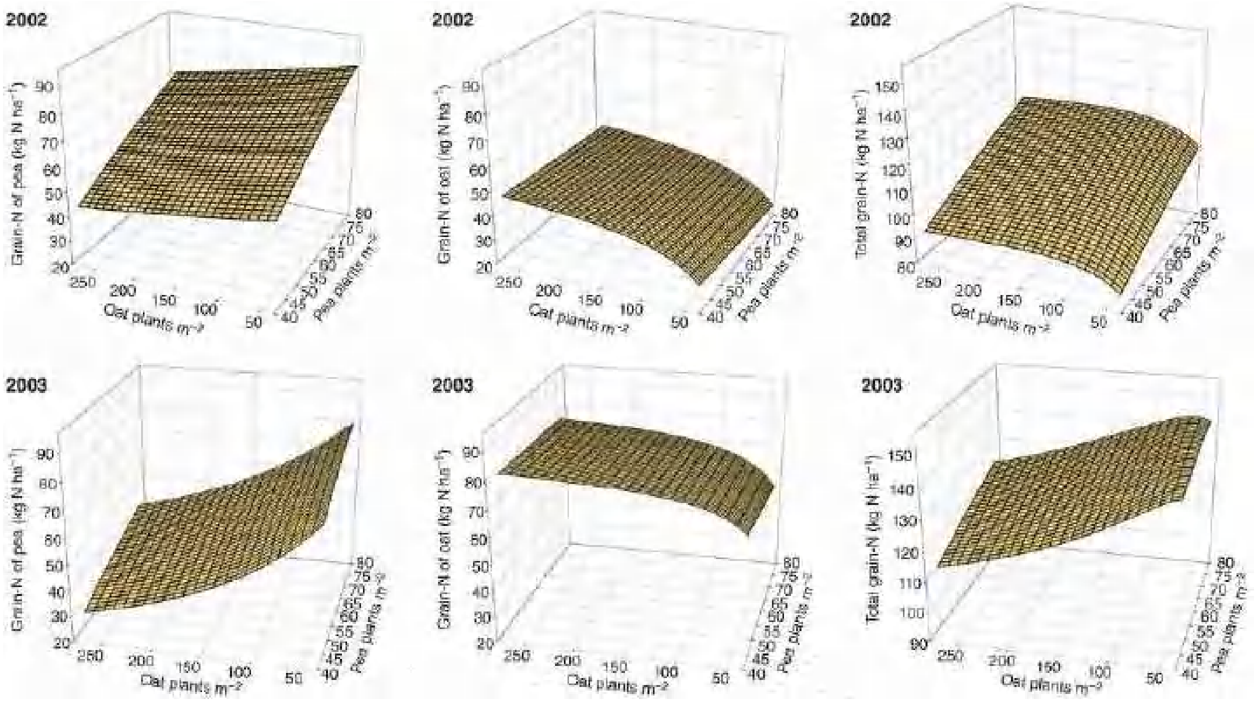


Figure 8.34 Response surface analysis demonstrates the effects of intercropping peas and oats. Response surfaces of the yields (in terms of nitrogen content of the harvestable grain) of peas, oats and the two combined, when they were sown at a range of densities in both 2002 and 2003. Details of how the surfaces were derived from the experimental data can be found in the original article.

Source: After Neumann *et al.* (2009).

It is apparent that the results differed between the two years. Yields from oats in particular were higher in 2003, probably as a result of a higher soil nitrogen content compared with 2002, to which the leguminous peas will have been less sensitive. This in turn, though, altered the intercropping mixture most likely to maximise total yield. In both years the recommendation would be to sow high densities of peas. Then, when soil nitrogen levels are low, as in 2002, this should be combined with high densities of oats. But when nitrogen levels are higher, as in 2003, sowing densities of oats should be much lower, since at higher densities the now more vigorous and competitive oats have a strongly detrimental effect on pea yields. Such recommendations could not be derived from simple additive or substitutive designs.

8.10 Evolutionary effects of interspecific competition

8.10.1 Natural experiments

pros and cons of natural experiments

We have seen that interspecific competition is commonly studied by an experimenter comparing species alone and in combination. Nature, too, often provides information of this sort: the distribution of certain potentially competing species is such that they sometimes occur together (sympatry) and sometimes occur alone (allopatry). These ‘natural experiments’ can provide additional information about interspecific competition, and especially about evolutionary effects, since the differences between sympatric and allopatric populations are often of long standing. The attractions of natural experiments are first that they are natural – they are concerned with organisms living in their natural habitats – and second, that they can be ‘carried out’ simply by observation – no difficult or impracticable experimental manipulations are required. They have the disadvantage, however, of lacking truly ‘experimental’ and ‘control’ populations. Ideally, there should be only one difference between the populations: the presence or absence of a competitor species. In practice, populations typically differ in other ways too, simply because they exist in different locations. Natural experiments should therefore always be interpreted cautiously.

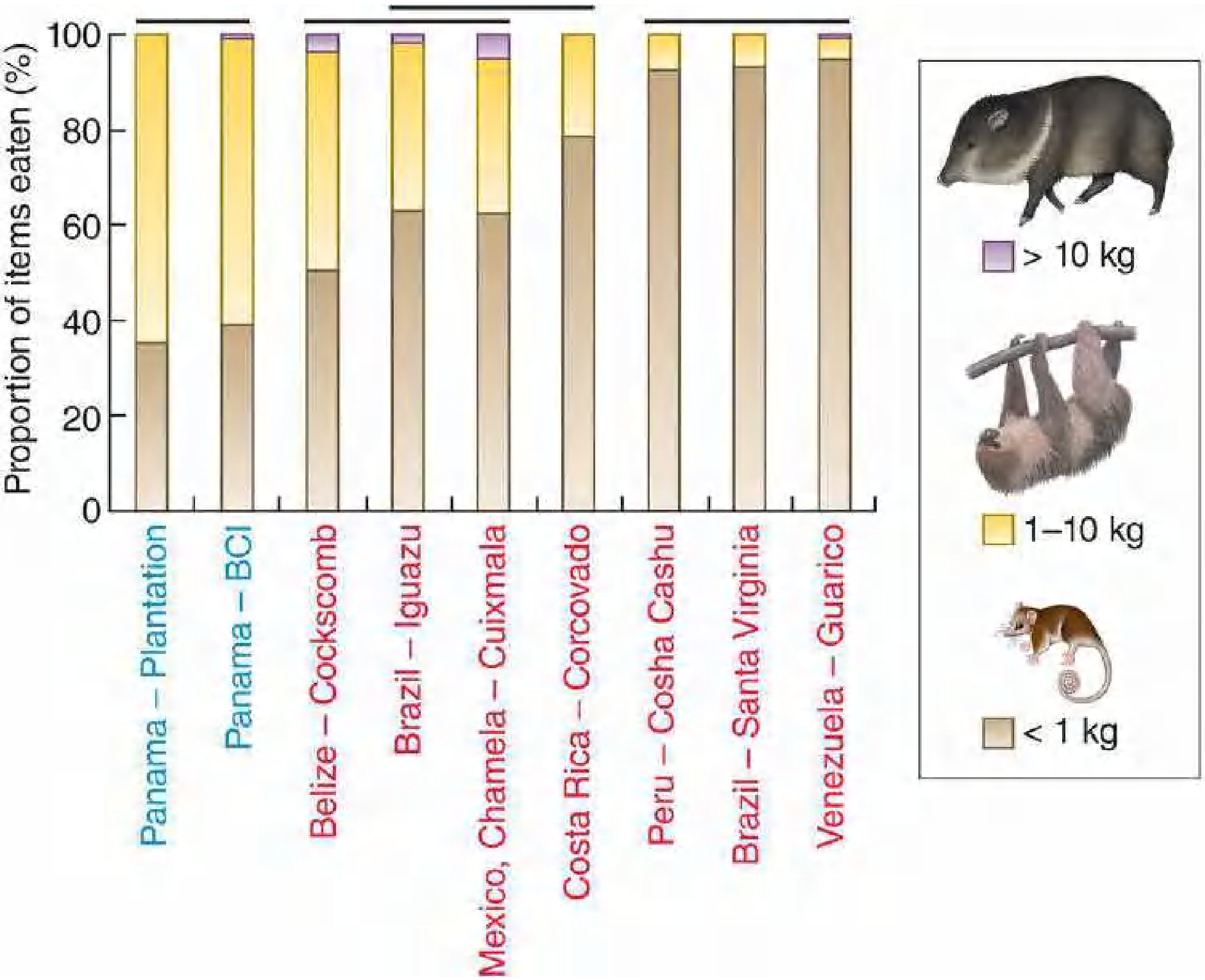
competitive release and character displacement

Evidence for competition from natural experiments usually comes either from niche expansion in the absence of a competitor (known as competitive release) or simply from a difference in the realised niche of a species between sympatric and allopatric populations. If this difference is accompanied by morphological changes, then the effect is referred to as character displacement, but physiological and behavioural traits are just as likely to be involved in competitive interactions and to be reflections of a species’ realised niche. One difference may be that morphological distinctions are most obviously the result of evolutionary change, but as we shall see, physiological and behavioural ‘characters’ are also liable to ‘competitive displacement’.

ocelots in Panama: competitive release

One example of natural competitive release is provided by work on ocelots in Panama. Ocelots (*Leopardus pardalis*) are wild cats distributed extensively across South and Central America. They are small (8–18 kg) compared with the pumas (*Puma concolor*) and jaguars (*Panthera onca*) with which they often coexist, which weigh in at 22–80 kg and 56–96 kg, respectively. It is not surprising, therefore, that when they coexist, they typically take much smaller prey than the other species. In some sites, though, they live in the absence of these other species, and Moreno *et al.* (2006) examined their scats (faeces) to analyse their diets in two of these sites: one on Barro Colorado Island in Panama where jaguar were absent, and another in central Panama where neither of the large cats was present.

Even comparing the two Panamanian sites, with and without pumas, there are notable differences. For example, the ocelots had a diet composed of 46% of the generally smaller rodents when they competed with pumas on Barro Colorado Island compared to only 25% at the mainland site. Most striking, however, is the comparison between the diets of ocelots in this study, in the absence of jaguars, and those from other studies throughout South and Central America where jaguars were present ([Figure 8.35]). The Panamanian sites stand out in having ocelot diets with 60% or more of prey items with weights in excess of 1 kg, whereas when jaguars are competing with the ocelots, the proportion was never more than 50% and often less than 10%. The fundamental niche of the ocelot clearly includes these larger prey, but in the presence of larger, superior competitors, that niche is substantially compressed.



[Figure 8.35] The diets of ocelots demonstrate competitive release in the absence of jaguars and pumas. The diets of ocelots, classified according to the size of prey items, across nine sites in South and Central America. At the sites in Panama, in blue, there were no jaguars present, but these were present at all the other sites, shown in red. Horizontal lines above the bars connect sites that are not significantly different in their proportion of prey sizes ($P < 0.05$). BCI, Barro Colorado Island.

Source: After Moreno *et al.* (2006).

morphological character displacement in Indian mongooses ...

A case of apparent morphological character displacement comes from work on Indian mongooses. In the western parts of its natural range, the small Indian mongoose (*Herpestes javanicus*) coexists with one or two slightly larger species in the same genus (*H. edwardsii* and *H. smithii*), but these species are absent in the eastern part of its range (Figure 8.36a). Simberloff *et al.* (2000) examined size variation in the upper canine tooth, the animal's principal prey-killing organ. In the east, where it occurs alone (area VII in Figure 8.36a), both males and females have larger canines than in the western areas (III, V, VI) where it coexists with the larger species (Figure 8.37b; note that female mongooses are smaller than males). This is consistent with the view that where similar but larger competitors are present, the prey-catching apparatus of *H. javanicus* has been selected for reduced size. This is likely to reduce the strength of competition with other species in the genus because smaller predators tend to take smaller prey than larger predators. Where *H. javanicus* occurs in isolation, its canine teeth are much larger.

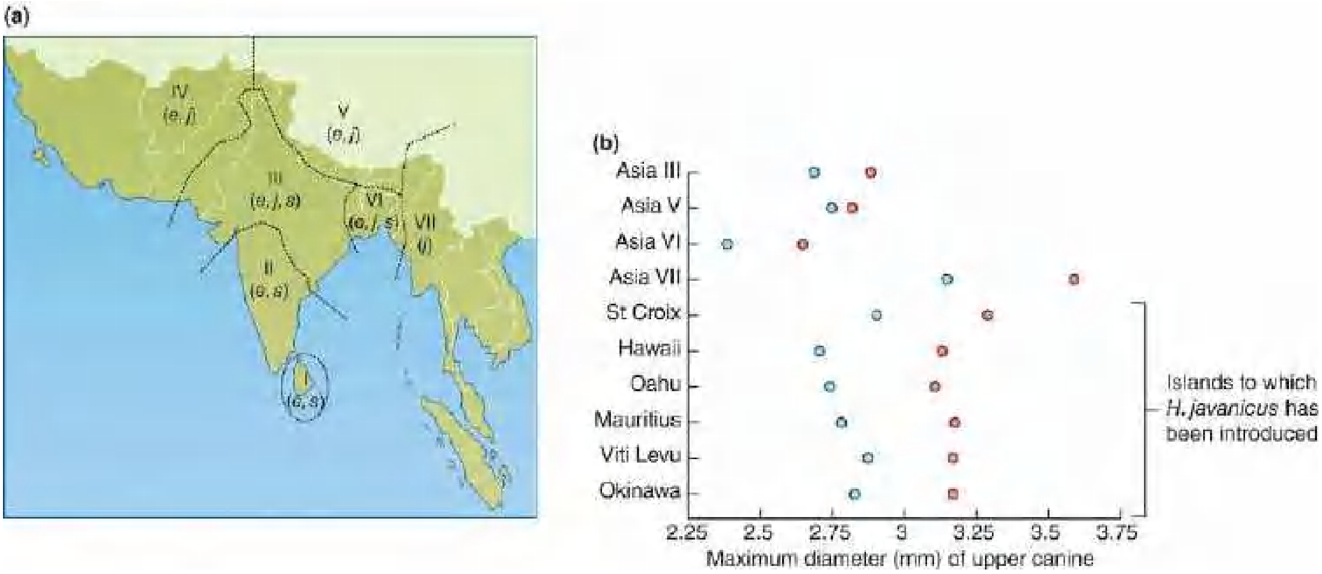


Figure 8.36 Character displacement in the canine teeth of Indian mongooses. (a) Native geographic ranges of *Herpestes javanicus* (*j*), *H. edwardsii* (*e*) and *H. smithii* (*s*). (b) Maximum diameter (mm) of the upper canine for *Herpestes javanicus* in its native range (data only for areas III, V, VI and VII from (a)) and islands on which it has been introduced. Symbols in blue represent mean female size and in red mean male size. Compared with area VII (*H. javanicus* alone), animals in areas III, V and VI, where they compete with the two larger species, are smaller. On the islands, they have increased in size since their introduction, but are still not as large as in area VII.

Source: After Simberloff *et al.* (2000).

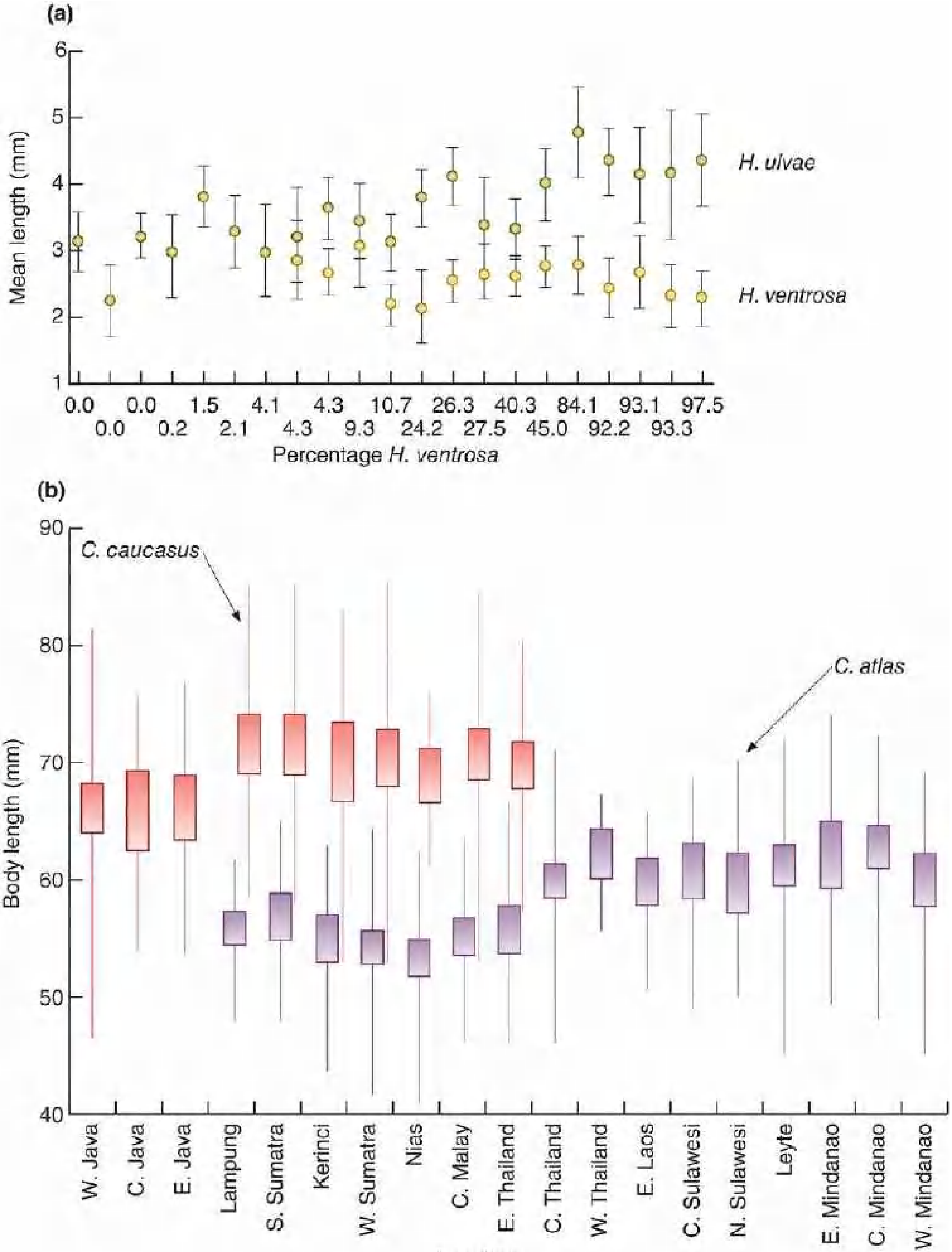

(a)
Mean length (mm)
1
2
3
4
5
6
H. ulvae
H. ventrosa
0.0
0.0
0.0
0.2
1.5
2.1
4.1
4.3
4.3
9.3
10.7
24.2
26.3
27.5
40.3
45.0
84.1
92.2
93.1
93.3
97.5
Percentage H. ventrosa
(b)
Body length (mm)
40
50
60
70
80
90
C. caucasus
C. atlas
W. Java
C. Java
E. Java
Lampung
S. Sumatra
Kerinci
W. Sumatra
Nias
C. Malay
E. Thailand
C. Thailand
W. Thailand
E. Laos
C. Sulawesi
N. Sulawesi
Leyte
E. Mindanao
C. Mindanao
W. Mindanao
Location

Figure 8.37 Apparent character displacement in the body size of mud snails and rhinoceros beetles. (a) Mud snails in Finland (average lengths of *Hydrobia ulvae* and *H. ventrosa*, arranged in order of increasing percentage of *H. ventrosa*). (b) Giant rhinoceros beetles in south-east Asia (average lengths of *Chalcosoma caucasus* and *C. atlas*). In each case in allopatry, body sizes overlap broadly, but in sympatry body sizes are significantly different.

Source: (a) After Saloniemi (1993). (b) After Kawano (2002).

It is of particular interest that the small Indian mongoose was introduced about a century ago to many islands outside its native range (often as part of a naive attempt to control introduced rodents). In these places, the larger competitor mongoose species are absent. Within 100–200 generations the small Indian mongoose has increased in size (Figure 8.36b), so that the sizes of island individuals are now intermediate between those in the region of origin (where they coexisted with other species and were small) and those in the east where they occur alone. On the islands they show variation consistent with 'ecological release' from competition with larger species.

... and in brook sticklebacks in Canada

If character displacement is an evolutionary response to interspecific competition, then the effects of competition should decline with the degree of displacement. Brook sticklebacks, *Culaea inconstans*, coexist in some Canadian lakes with nine-spine sticklebacks, *Pungitius pungitius*, whereas in other lakes, brook sticklebacks live alone. In sympatry, the brook sticklebacks possess significantly shorter gill rakers (more suited for capturing food from open waters), longer jaws and deeper bodies than they do when they live alone. It is therefore reasonable to suppose that brook sticklebacks living alone possess predisplacement characteristics, whereas those coexisting with nine-spine sticklebacks display character displacement. To test this, each type was placed separately in enclosures in the presence of nine-spine sticklebacks (Gray & Robinson, 2001). The predisplacement brook sticklebacks grew significantly less well than their sympatric postdisplacement counterparts. This is clearly consistent with the hypothesis that the postdisplacement phenotype has evolved to avoid competition, and hence enhance fitness, in the presence of nine-spine sticklebacks.

mud snails: a classic example of character displacement?

Two further, plausible examples of character displacement are provided by work on mud snails in Finland (*Hydrobia ulvae* and *H. ventrosa*) and giant rhinoceros beetles in Southeast Asia (*Chalcosoma caucasus* and *C. atlas*). When the two mud snail species live apart, their sizes are more or less identical, but when they coexist they are always different in size (Figure 8.37a; Saloniemi, 1993) and they tend to consume different food particle sizes (Fenchel 1975). The beetles display a similar morphological pattern (Figure 8.37b; Kawano, 2002). These data, therefore, strongly suggest character displacement, allowing coexistence. However, even an apparently exemplary example such as that of the mud snails is open to serious question (Saloniemi, 1993). In fact, the sympatric and allopatric habitats were not identical: where *H. ulvae* and *H. ventrosa* coexisted, they did so in sheltered water bodies rarely affected by tidal action, whereas *H. ulvae* was found alone in relatively exposed tidal mudflats and salt marshes, and *H. ventrosa* was found alone in non-tidal lagoons and pools. Moreover, *H. ulvae* naturally grows larger in less tidal habitats, and *H. ventrosa* may grow less well in this habitat. This alone could account for the size differences between sympatry and allopatry in these species, emphasising the major problem with such natural experiments: sympatric and allopatric populations can occur in different environmental conditions over which the observer has no control. Sometimes it will be these environmental differences, rather than competition, that have led to the character displacement.

8.10.2 Experimenting with natural experiments

niche divergence in clover–grass competition

Sometimes, natural experiments may themselves provide an opportunity for a further – and more informative – experimental manipulation. In one such case, niche divergence was sought in clover, *Trifolium repens*, that may have resulted from its having to compete with the grass *Lolium perenne* (Turkington & Mehrhoff, 1990). Clover was examined from two sites: (i) a 'two-species' site, in which clover achieved a ground coverage of 48% and the grass achieved a coverage of 96% (the two added together exceed 100% because their leaves can overlap); and (ii) a 'clover-alone' site in which clover achieved 40% coverage, but *L. perenne* covered only 4%. A total of three transplant experiments (into the other site) and three re-plant experiments (back into the home site) were carried out. Thus, *T. repens* from both sites was planted in: (i) plots at the two-species site cleared of *T. repens* only; (ii) plots at the two-species site cleared of both *T. repens* and *L. perenne*; and (iii) plots at the clover-alone site cleared of *T. repens* (Figure 8.38a). The extent of competitive suppression or release was assessed from the amount of growth achieved by the different plantings of *T. repens*. From this the extent of the evolution of niche divergence between 'clover-alone' and 'two-species' *T. repens* was deduced, as was that between *T. repens* and *L. perenne*.

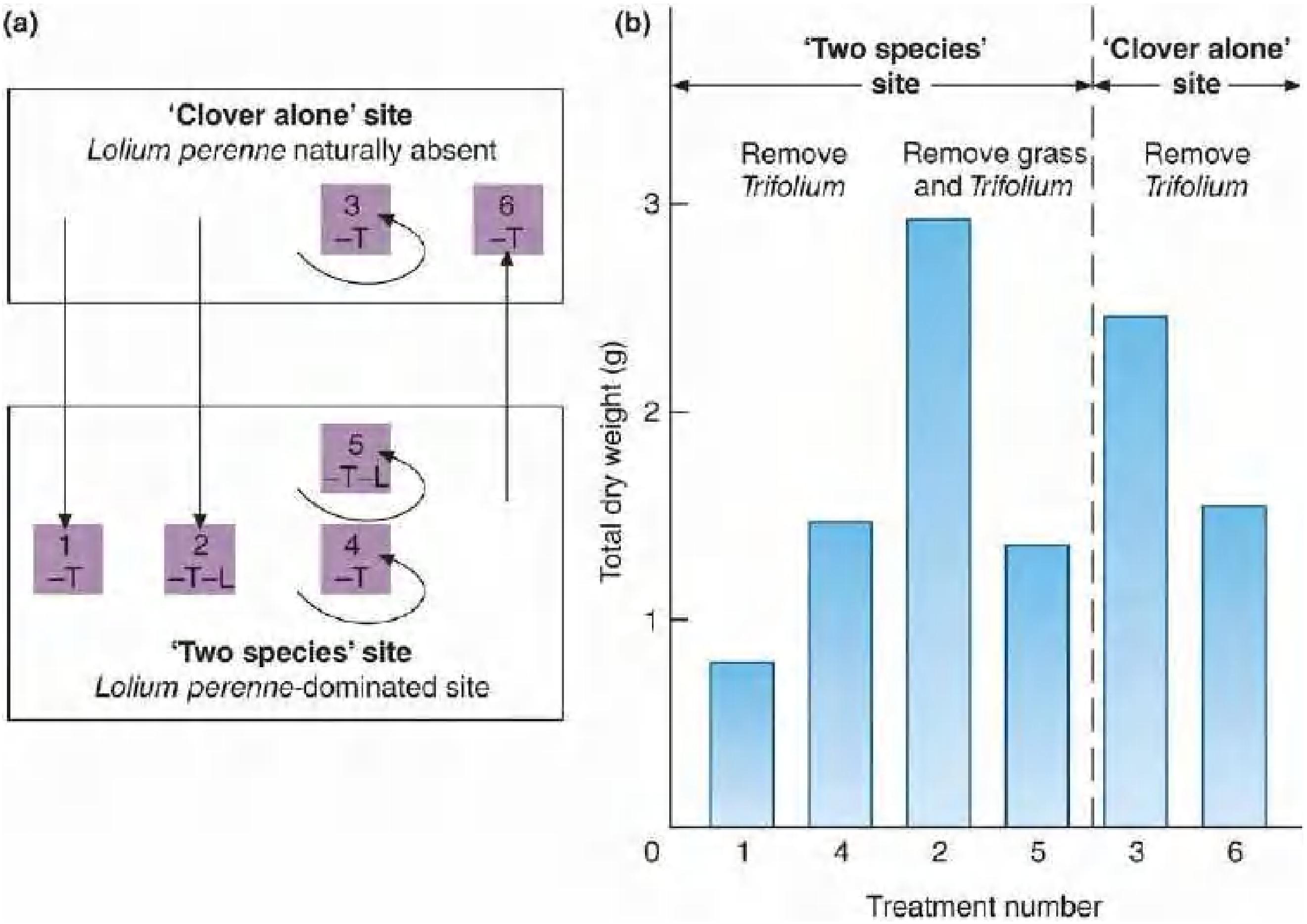


Figure 8.38 An experimental manipulation demonstrates evolution in clover in response to competition with a grass. (a) Experimental design to test for the evolution of *Trifolium repens* (T) in competition with *Lolium perenne* (L). Indigenous populations of *T. repens*, and sometimes also *L. perenne*, were removed. *Trifolium repens* was removed from the base of the arrow and transplanted, or replanted, at the head of the arrow. (b) The results of this experiment in terms of the total plot dry weight achieved by *T. repens* in the various treatments. Significance levels for comparisons between pairs of treatments are given in the text.

Source: After Turkington & Mehrhoff (1990).

The clover population from the two-species site had indeed apparently diverged from the *L. perenne* population with which it was coexisting (and with which it may otherwise have competed strongly), and had diverged too from the clover-alone population (Figure 8.38b). When the two-species site was cleared of clover only, the local, *re*-planted clover grew better than the *trans*planted clover-alone plants (treatments 1 and 4, respectively; $P = 0.086$, close to significance), suggesting that the transplanted plants were competing more with the resident *L. perenne*. Moreover, when *L. perenne* was also removed, this made no difference to the clover from the two-species site (treatments 4 and 5; $P > 0.9$), but led to a large increase in the growth of clover from the clover-alone site (treatments 1 and 2; $P < 0.005$). Also, when *L. perenne* was removed, the clover plants from the clover-alone site grew better than those from the two-species site (treatments 2 and 5; $P < 0.05$) – all of which suggests that only the plants at the clover-alone site were being released from competition by the absence of *L. perenne*. Finally, at the clover-alone site, the clover plants from the two-species sites grew no better than they had at their home site (treatments 4 and 6; $P > 0.7$), whereas the clover from the clover-alone site grew far better than it had at the two-species site in the presence of the grass (treatments 1 and 3; $P < 0.05$). Thus, the clover from the two-species population hardly competes with the *L. perenne* with which it coexists, whereas the clover-alone population would do – and does so if transplanted to the two-species site.

8.10.3 Selection experiments

direct demonstrations of evolutionary effects of competition have been rare – and focused on microorganisms

The most direct way of demonstrating the evolutionary effects of competition within a pair of competing species would be for the experimenter to induce these effects – impose the selection pressure (competition) and observe the outcome. In fact, there have been very few successful experiments of this type. In large part, this is due to the sheer practical difficulty of maintaining competing species together for long enough, and with sufficient replication, for any changes to be detected. It is no surprise, therefore, that most of these few examples are for bacteria, but these have suffered the disadvantage of using competing strains of the same bacterial species, which, while they do not interbreed, may fail to satisfy whole-organism biologists seeking evidence from quite separate species.

One now classic example examined competition between three types of the same bacterial species, *Pseudomonas fluorescens*, which behave as separate species because they reproduce asexually (Rainey & Travisano, 1998 ; see also Buckling *et al*., 2009). The three types are named 'smooth' (SM), 'wrinkly spreader' (WS) and 'fuzzy spreader' (FS) on the basis of the morphology of their colonies plated out on solid medium. In liquid medium ('broth'), they also occupy quite different parts of the culture vessel: SM in the broth, WS at the broth surface, and FS at the bottom of the vessel. In vessels that were continually shaken, so that no separate niches for the different types could be established, an initially pure culture of SM individuals retained its purity (Figure 8.39a). But in the absence of shaking, WS and FS types emerged through mutation, invaded and established (Figure 8.39b). Furthermore, it was possible to determine the competitive abilities of the different types, when rare, to invade pure cultures of the other types (Figure 8.39c). Five of six possible invasions are favoured. The exception – WS repels the invasion of FS – is unlikely to lead to the elimination of FS, because FS can invade cultures of SM, and SM can invade cultures of WS.

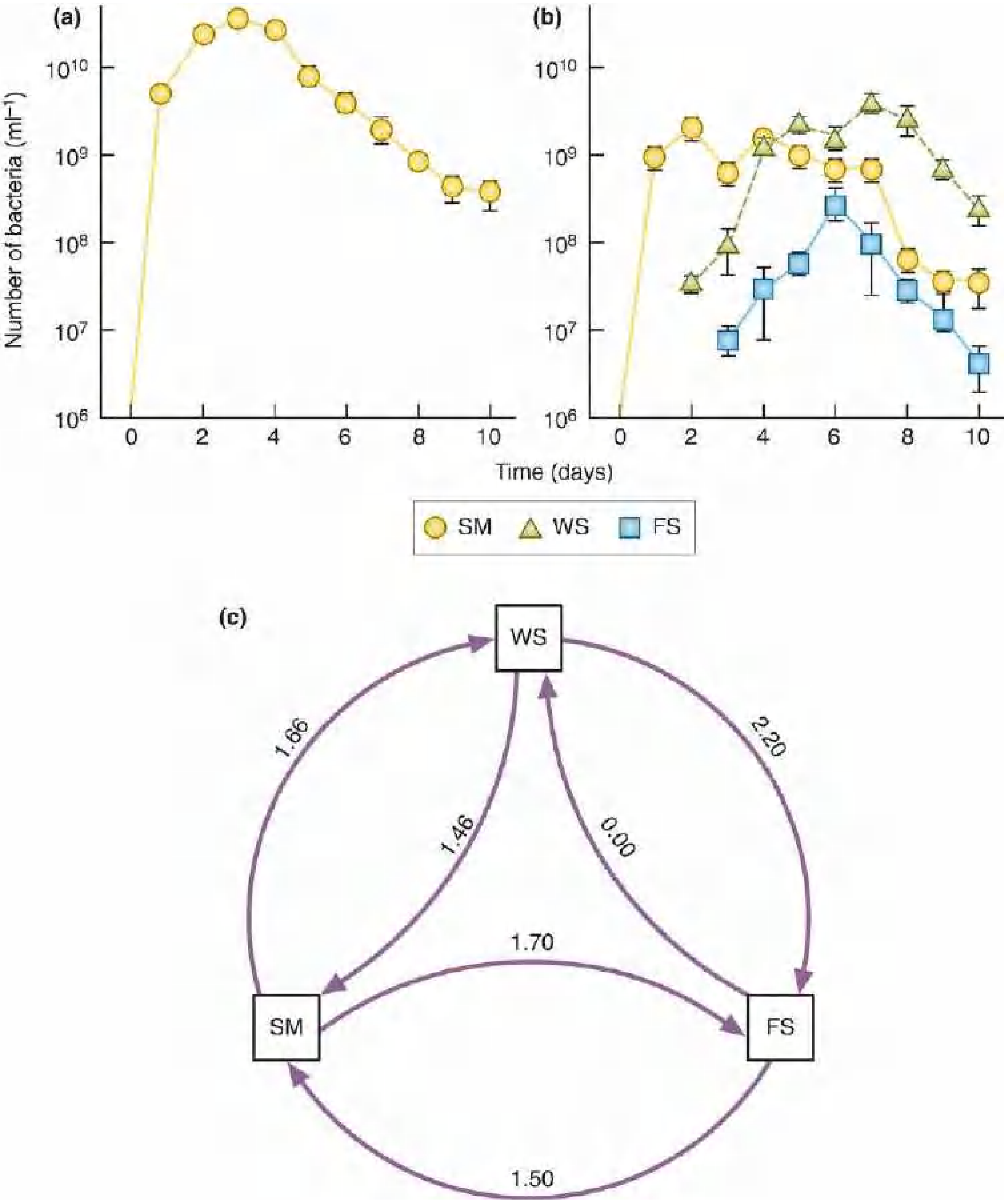

(a)
(b)
Number of bacteria (ml⁻¹)
10¹⁰
10⁹
10⁸
10⁷
10⁶
0
2
4
6
8
10
Time (days)
SM
WS
FS
(c)
WS
SM
FS
1.66
2.20
1.46
0.00
1.70
1.50

Figure 8.39 Experimental evolution of niche differentiation in *Pseudomonas*. Pure cultures of three types of the bacterium, *Pseudomonas fluorescens* (smooth, SM, wrinkly spreader, WS, and fuzzy spreader, FS) concentrate their growth in different parts of a liquid culture vessel, and hence compete less with one another than they would otherwise do. (a) In shaken vessels, there is only one niche, and only the smooth type survives, excluding its competitors. (b) But in unshaken culture vessels, an initial smooth-type culture gives rise to wrinkly spreader, and fuzzy spreader mutants and all three coexist (bars represent SEs, and cultures are identified as SM, WS and FS). (c) The competitive abilities (relative rates of increase) when an initially rare type (foot of the arrow) invades a pure colony of another type (head of the arrow). Hence, values >1 indicate an ability to invade (superior competitor when rare) and values <1 an inability to invade.

Source: After Rainey & Travisano (1998).

In another example, this time with truly separate species, a protozoan, *Colpoda* sp., that lives in the water-filled leaves of purple pitcher plants, *Sarracenia purpurea*, was competed for 20 days (60–120 generations) against another protozoan species, *Pseudocyrtolophis alpestris*, in replicate laboratory microcosms designed to mimic the pitcher plant environment. *Colpoda* was taken from natural, pitcher plant populations and maintained either in monoculture (intraspecific competition only) or with *P. alpestris* added, initially at about twice the starting density of *Colpoda*. The characteristics of *Colpoda* from the ancestral population, and after selection both in monoculture and with the competitor, were then compared in the monoculture environment. Clearly, *Colpoda* made an evolutionary response to competition – population growth rate was greater and cell size was smaller (Figure 8.40a, b) – and while this response was apparent following exposure to intraspecific competition only, it was significantly greater when *P. alpestris* had been present throughout the selection period. What is less clear though, unlike the bacterial example in Figure 8.39, is whether this response was associated with any niche differentiation. The competitor, *P. alpestris*, is mostly found in the water column rather than in a 'refuge' provided by glass beads in the laboratory tubes. Following its period of selection with *P. alpestris*, there was an increased tendency for *Colpoda* to be found in this refuge (Figure 8.40c), but this effect was not statistically significant.

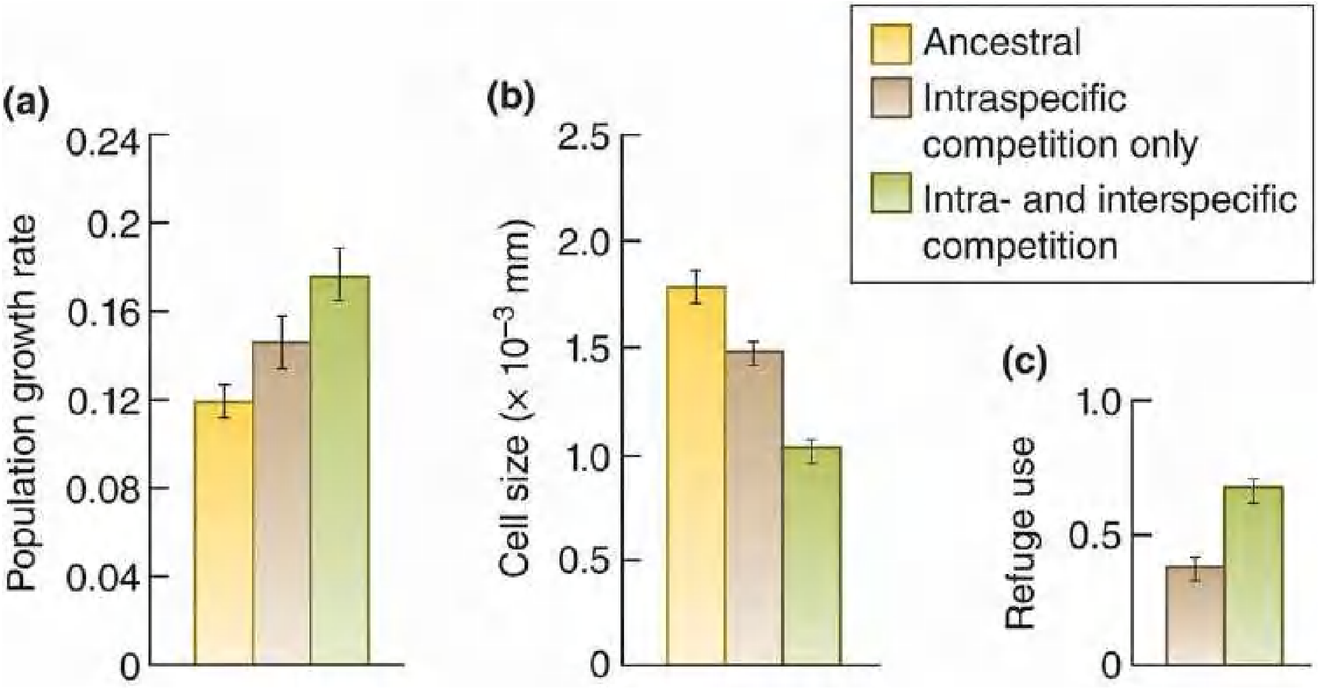


Figure 8.40 Experimental evolution of competitive ability in a protozoan. When the pitcher-plant protozoan, *Colpoda*, was subjected either to a period of intraspecific competition or of competition against another protozoan, *Pseudocyrtolophis alpestris*, it evolved a higher population growth rate (a) and a smaller cell size (b), especially when there had been interspecific competition, associated with an increased tendency to make greater use of a glass-bead refuge from the competitor (c). Bars are SEs.

Source: After terHorst (2011).

Overall, therefore, while the evolutionary effects of interspecific competition are generally considered to be profound, demonstrating these effects directly remains a considerable challenge for ecologists.

Chapter 9

The Nature of Predation

9.1 Introduction

9.1.1 The types of predators

Predation, put simply, is consumption of one organism (the prey) by another organism (the predator) where the prey is alive when the predator first attacks it. This excludes detritivory, the consumption of dead organic matter (see Chapter 11), but it is a definition that encompasses a wide variety of interactions and a wide variety of 'predators'.

taxonomic and functional classifications of predators

There are two main ways of classifying predators. One is 'taxonomic': carnivores consume animals, herbivores consume plants and omnivores consume prey from more than one trophic level. Our preferred alternative is a 'functional' classification of the type already outlined in Chapter 3. Here, there are four main types of predator: true predators, grazers, parasitoids and parasites. (See Lafferty *et al.* (2015) for a general model of consumer–resource interactions within which these four types, and others, are special cases.)

true predators

True predators kill their prey more or less immediately after attacking them; during their lifetime they kill several or many different prey individuals, often consuming prey in their entirety. Most of the more obvious carnivores like tigers, eagles, spiders and carnivorous plants are true predators, but so too are seed-eating rodents and ants, plankton-consuming whales, and so on.

grazers

Grazers also attack large numbers of prey during their lifetime, but they remove only part of each prey individual rather than the whole. Their effect on a prey individual is typically harmful but rarely lethal in the short term – and certainly never predictably lethal since that would make them true predators. Obvious examples are the large vertebrate herbivores like sheep and cattle, but the flies that bite a succession of vertebrate prey, and leeches that suck their blood, are also undoubtedly grazers by this definition.

parasites

Parasites, like grazers, consume parts of their prey (their ‘host’), rather than the whole, and are typically harmful but rarely lethal in the short term. Unlike grazers, however, their attacks are concentrated on one or a very few individuals during their life. There is, therefore, an intimacy of association between parasites and their hosts that is not seen in true predators and grazers. Tapeworms, liver flukes, the measles virus, the tuberculosis bacterium and the flies and wasps that form mines and galls on plants are all good examples of parasites. There are also many plants, fungi and microorganisms that are parasitic on plants (often called ‘plant pathogens’), including the tobacco mosaic virus, the rusts and smuts and the mistletoes. Moreover, many herbivores should strictly be thought of as parasites, though we would never usually do so. For example, aphids extract sap from one or a very few individual plants with which they enter into intimate contact. Even caterpillars often rely on a single plant for their development. Plant pathogens, and animals parasitic on animals – ‘conventional’ parasites – will be dealt with together in Chapter 12. ‘Parasitic’ herbivores, like aphids and caterpillars, are dealt with here.

parasitoids

The parasitoids are a group of insects that belong mainly to the order Hymenoptera, but also include many Diptera. They are free-living as adults, but lay their eggs in, on or near other insects (or, more rarely, in spiders or woodlice). The larval parasitoid then develops inside or on its host. Initially, it does little apparent harm, but eventually it almost totally consumes the host and therefore kills it. An adult parasitoid emerges from what is apparently a developing host. Often, just one parasitoid develops from each host, but in some cases several or many individuals share a host. Thus, parasitoids are intimately associated with a single host individual (like parasites), they do not cause immediate death of the host (like parasites and grazers), but their eventual lethality is inevitable (like predators). This might seem to make parasitoids an unusual group of limited general importance. However, they account for an estimated 10% or more of the world’s species (Godfray, 1994), which is not surprising given that there are so many species of insects, that most of these are attacked by at least one parasitoid, and that parasitoids may in turn be attacked by parasitoids. What’s more, a number of parasitoid species have been intensively studied by ecologists, and they have provided a wealth of information relevant to predation generally.

9.1.2 Patterns of abundance and the need for their explanation

Whatever the type of predator, we are bound as ecologists to ask: What are the effects of predation on the distribution and abundance of the predator and its prey? Even a limited survey of the data reveals a wide array of dynamics. Sometimes, certainly, predation has a profoundly detrimental effect on a prey population. For example, the ‘vedalia’ ladybird beetle (*Rodolia cardinalis*) is famous for having virtually eradicated the cottony cushion-scale insect (*Icerya purchasi*), a pest that threatened the California citrus industry in the late 1880s (see Section 15.2.4). On the other hand, there are many cases where predators and herbivores have no apparent effect on their prey’s dynamics or abundance. For example, another beetle, the weevil *Apion ulicis*, has been introduced into many parts of the world in an attempt to control the abundance of gorse bushes (*Ulex europaeus*), and it has often become well established. Yet in Chile – and this is fairly typical – despite eating up to 94% of the seeds produced, it has had no appreciable impact on gorse invasiveness (Norambuena & Piper, 2000). There are also examples that appear to show predator and prey populations linked together by coupled oscillations in abundance (Figure 9.1), but there are many more examples in which predator and prey populations fluctuate in abundance apparently independently of one another.

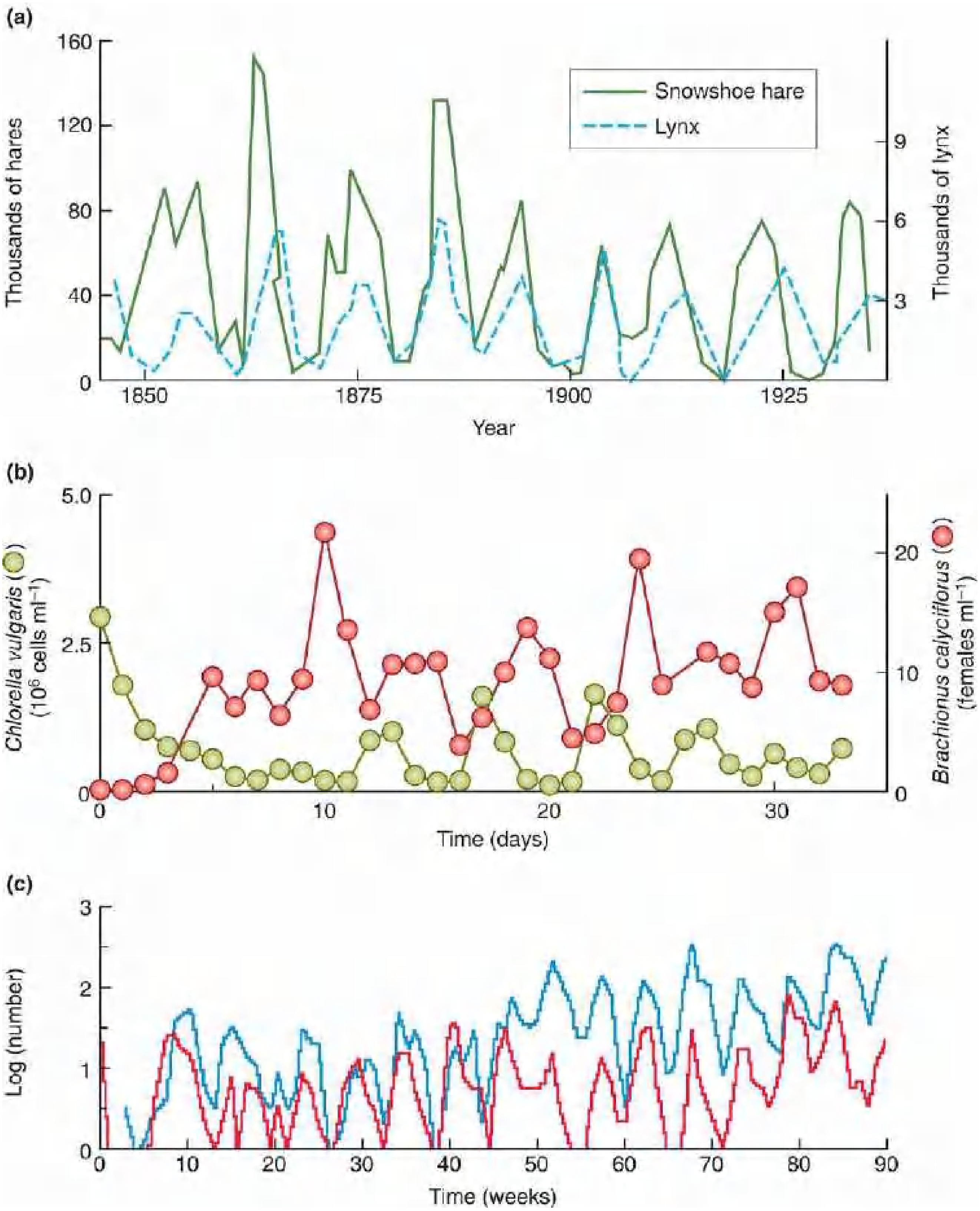


Figure 9.1 Coupled oscillations in the abundance of predators and prey. (a) The snowshoe hare (*Lepus americanus*) and the Canadian lynx (*Lynx canadensis*) as determined by the number of pelts lodged with the Hudson Bay Company. (b) Parthenogenetic female rotifers, *Brachionus calyciflorus* (predators), and unicellular green algae, *Chlorella vulgaris* (prey) in laboratory cultures. (c) The parasitoid *Venturia canescens* (red line) and its moth host *Plodia interpunctella* (blue line) in laboratory cultures.

Source: (a) After MacLulick (1937). (b) After Yoshida *et al.* (2003). (c) After Bjørnstad *et al.* (2001).

Clearly, we would like to understand these patterns of predator–prey abundance, and to account for the differences from one example to the next. But it is equally clear that none of these populations exist as isolated pairs. All are parts of multispecies systems, and all are affected by environmental conditions. These broader issues are taken up later: interactions within 'webs' of several species (see Chapter 17), and how a whole range of factors combines to determine a species' abundance (see Chapter 14). However, as with any complex process, we cannot understand the full complexity without a reasonable understanding of the components – in this case, populations of predators and prey.

Here, we begin by looking at the behaviour of predators and especially the factors that determine the composition of their diets (Section 9.2). We then turn in Section 9.3 to the responses of prey: how, as individuals, they may resist or tolerate those that seek to consume them (or part of them), and how there may be compensatory responses at the population level such that the effects on prey populations are not as profound as they might seem. We then turn, in Chapter 10, to establishing the fundamental patterns and drivers of predator–prey population dynamics, before going on, in subsequent sections of that chapter, to examine a series of features – the effects of consumption on consumers, the effects of conspecifics, responses to patchiness, and so on – and the ways in which these affect the population dynamics.

9.2 Foraging: widths and compositions of diets

range and classification of diet widths

Consumers can be classified along a specialist-generalist continuum as *monophagous* (feeding on a single prey type), *oligophagous* (few prey types) or *polyphagous* (many prey types). There are true predators with specialised diets (for instance the snail kite, *Rostrhamus sociabilis*, a bird that feeds almost entirely on snails of the genus *Pomacea*), but most have relatively broad diets. Parasitoids, on the other hand, are typically specialised and often monophagous. Among herbivores, grazing and 'predatory' herbivores typically have broad diets, but 'parasitic' herbivores are very often highly specialised. For instance, Novotny and Basset (2005) reviewed studies on the host specificity of herbivores in tropical forests and found that those feeding as larvae within the seeds of their host plants ('parasitising' them) were typically the most specialised (99% feeding within a single plant family and 76% feeding on a single species), followed by leaf-miners (parasites living on individual leaves), fruit-eaters, leaf-chewers and sap-suckers, wood-eaters and finally root-feeders as the least specialised. As they note, a similar hierarchy is seen among temperate herbivores.

9.2.1 Food preferences

preference is defined by comparing diet with 'availability'

Even among generalist predators, some degree of preference is almost always apparent. An animal is said to exhibit a preference for a particular type of food when the proportion of that type in the animal's diet is higher than its proportion in the animal's environment. To measure food preference in nature, therefore, it is necessary not only to determine the animal's diet, but also to assess the 'availability' of different food types, ideally not by simply sampling the environment, but also taking account of the animal's access to the food types.

preferences among perfectly substitutable and complementary food types

A food type may be preferred or chosen because it is the most valuable amongst those available – most commonly because it has the highest nutritional value, and especially when the resource types are ‘perfectly substitutable’ (see Section 3.8.1). Alternatively, it may be chosen because it provides an integral part of a mixed and balanced diet. That is, when the resource types are ‘complementary’. Figure 9.2, for instance, shows two examples in which carnivores actively selected prey items that were the most profitable in terms of energy intake per unit time spent dealing with (or ‘handling’) prey. Results such as these reflect the fact that a carnivore’s food often varies little in composition (see Section 3.7.1), but may vary in size or accessibility. This allows a single measure to be used to rank food items. Thus, Figure 9.2a shows consumers exhibiting an active preference for food where the ‘energy gained per unit handling time’ was greatest: in this case leeches, *Whitmania laevis*, that enter the shells of molluscs to attack them. On the other hand, as we shall see later when we examine the effects of predator-avoidance on consumer behaviour, consumers will often evolve a strategy that combines energy intake with other imperatives. Figure 9.2b, for example, shows how sea stars, *Asterias rubens*, prefer otherwise less profitable small blue mussels, *Mytilus edulis*, that lessen their chances of being damaged themselves in the act of predation.

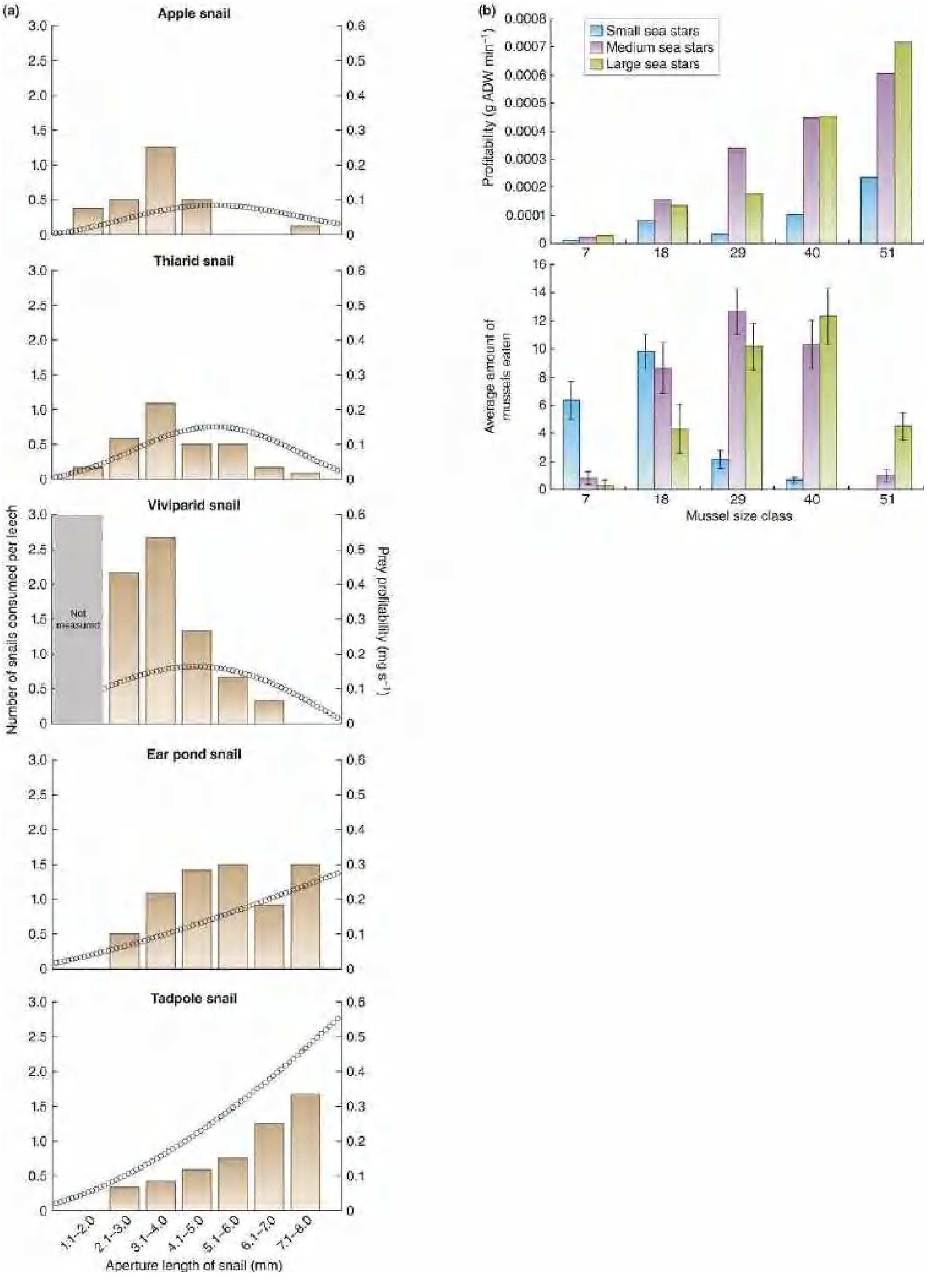

(a)
Apple snail
Thiarid snail
Viviparid snail
Ear pond snail
Tadpole snail
Number of snails consumed per leech
Prey profitability (mg s^{-1})
Not measured
0
0.5
1.0
1.5
2.0
2.5
3.0
0.1
0.2
0.3
0.4
0.5
0.6
1.1–2.0
2.1–3.0
3.1–4.0
4.1–5.0
5.1–6.0
6.1–7.0
7.1–8.0
Aperture length of snail (mm)
(b)
Small sea stars
Medium sea stars
Large sea stars
Profitability (g ADW min^{-1})
0.0001
0.0002
0.0003
0.0004
0.0005
0.0006
0.0007
0.0008
Average amount of mussels eaten
2
4
6
8
10
12
14
16
7
18
29
40
51
Mussel size class

Figure 9.2 Predators tend to prefer more profitable food types but may modify these preferences to account for other factors. (a) For five species of snail, as indicated, there is a good match between the profitabilities of different size classes (circles) and the average number of them consumed by the leech, *Whitmania laevis*, over a 72 h period (bars). Profitability represents the rate of energy return to the leech and combines the soft tissue weight of the snail, the proportion this is of the whole, the probability of a leech attack being successful and the time it takes a leech to invade the snail. (b) When sea stars, *Asterias rubens*, attack blue mussels, *Mytilus edulis*, the upper panel shows that larger mussels are more profitable (energy content (ash-free dry weight) per minute of handling time) for small, medium and large sea stars (blue, purple and green bars, respectively). Nonetheless, the lower panel shows that small sea stars especially prefer smaller mussels, where the chances of damaging their feeding apparatus are lower. Bars are SEs.

Source: (a) After Lai *et al*. (2011). (b) After Hummel *et al*. (2011).

For many consumers, however, especially herbivores and omnivores, even if we ignore other pressures like the avoidance of damage, no simple ranking is appropriate, since none of the available food items matches the nutritional requirements of the consumer. These requirements can therefore only be satisfied either by eating large quantities of food, and eliminating much of it in order to get a sufficient quantity of the nutrient in most limited supply (for example aphids and scale insects excrete vast amounts of carbon in honeydew to get sufficient nitrogen from plant sap), or by eating a combination of food items that between them match the consumer's requirements. Behmer (2009), for example, reviewed studies on the regulation of nutrient intake in insects and found numerous examples in which species took in a target ratio of protein to carbohydrate, irrespective of the ratio in the diet offered them; though that target ratio varied according to species requirements from 0.7 : 1 to 4 : 1, depending on their lifestyle, taxonomic group, and so on. We have only to look at ourselves to see an example in which 'performance' is far better on a mixed diet than on a pure diet of even the 'best' food.

mixed diets can be favoured for a variety of reasons

There are two other important reasons why a mixed diet may be favoured. Firstly, consumers may accept low-quality items simply because, having encountered them, they have more to gain by eating them (poor as they are) than by ignoring them and continuing to search. This is discussed in detail in Section 9.2.3. Secondly, consumers may benefit from a mixed diet because each food type may contain a different undesirable toxic chemical. A mixed diet would then keep the concentrations of all of these chemicals within acceptable limits. It is certainly the case that toxins can play an important role in food preference. For instance, dry matter intake by Australian ringtail possums (*Pseudocheirus peregrinus*) feeding on *Eucalyptus* tree leaves was strongly negatively correlated with the concentration of sideroxylonal, a toxin found in *Eucalyptus* leaves, but was not related to nutritional characteristics such as nitrogen or cellulose (Lawler *et al*., 2000).

Of course, it would be wrong to suggest that we understand the underlying mechanisms for all the consumer preferences that have been observed, or that those preferences are always reflected in improved performance by the consumers satisfying those preferences. Indeed, much thought has been given (e.g. Thompson, 1988) to why there may be a mismatch between preference and performance. Helpfully, therefore, Gripenberg *et al*. (2010) carried out a meta-analysis of one group of relevant studies: those on oviposition preferences of phytophagous insects among host plants on which their larval offspring are consumers – testing the 'mother knows best' hypothesis. (A meta-analysis combines the results of all previous studies, but typically does so in a way that selects only valid ones and aims to control for factors that may have led to discrepant results between those studies.) The results were strongly supportive of the hypothesis: eggs were typically laid most where larval performance was strongest.

9.2.2 Switching

switching involves a preference for food types that are common

The preferences of many consumers are fixed; that is, they are maintained irrespective of the relative availabilities of alternative food types. But many others switch their preference, such that food items are eaten disproportionately often when they are common and are disproportionately ignored when they are rare. We see this in Figure 9.3a, where guppies (fish), when offered a choice between fruit-flies and tubificid worms, switched their preference and consumed a disproportionate number of the type that was more abundant.

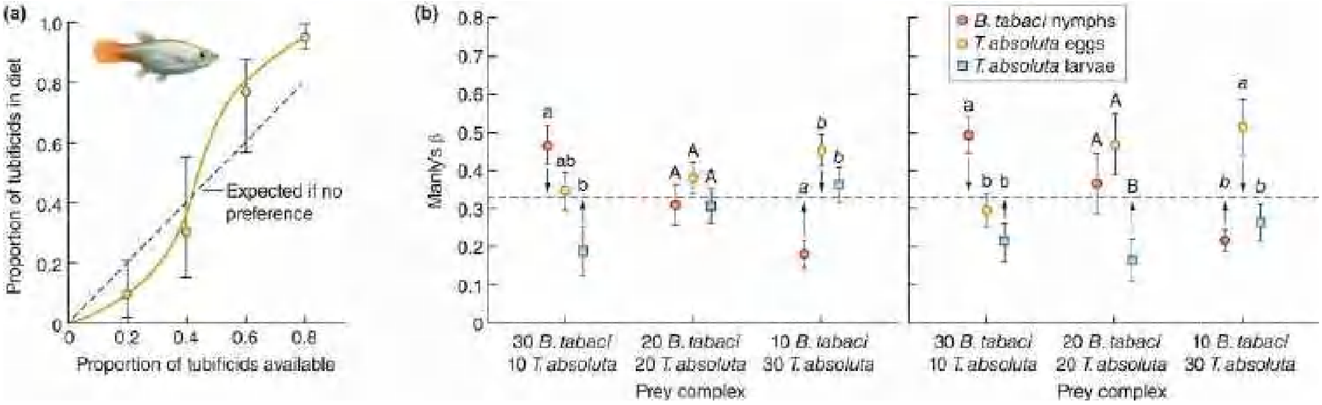


Figure 9.3 Switching of preferences by predators depends on the relative abundance of prey types. (a) Switching by guppies fed on tubificids and fruit-flies: they take a disproportionate amount of whichever prey type is the more available (means and total ranges). (b) Switching by mirid bugs, *Macrolophus pigmaeus*, feeding on three prey types belonging to two pests of tomato plants, *Bamisia tabaci* and *Tuta absoluta*, as indicated. Results for adult bugs are shown in the left panel; juvenile bugs in the right. In both cases, Manly's β measures prey preference, with the dashed line indicating the value to be expected with no preference. Bars are SEs and letters above them differ when preferences for the respective prey types are significantly different. Both prey species showed a clear tendency to be more preferred when they were common and less preferred when they were rare.

Source: (a) After Murdoch *et al.* (1975). (b) After Jaworski *et al.* (2013).

APPLICATION 9.1 Switching can have economic importance in horticulture

A second example of prey switching is shown in Figure 9.3b, where the switching has potential economic significance since it was observed in a generalist insect predator, the mirid bug *Macrolophus pigmaeus*, feeding on two pests of tomato plants in Europe, the native whitefly *Bamisia tabaci*, and the invasive South American tomato pinworm, *Tuta absoluta*. As we shall discuss much more fully in Chapter 15, natural enemies are increasingly used as biological control agents against pests. Here, the switching behaviour of *M. pigmaeus* therefore offers the prospect of particular effectiveness in the simultaneous control of the two pests, which often co-occur on tomato plants, since it will focus its attacks on whichever species is most common (and so most threatening).

when might switching arise?

There are a number of situations in which switching can arise. Probably the most common is where different types of prey are found in different microhabitats, and the consumers concentrate on the most profitable microhabitat. This was the case for the guppies in Figure 9.3a: the fruit-flies floated at the water surface while the tubificids were found at the bottom. Switching can also occur when consumers develop a 'search image' for abundant food (Tinbergen, 1960) and concentrate on their 'image' prey to the relative exclusion of non-image prey, or when there is an increased probability of pursuing, capturing or handling a common prey type (Bergelson, 1985). In all these cases, increasingly common prey generate increased interest and/or success on the part of the predator, and hence an increased rate of consumption.

9.2.3 The optimal foraging approach to diet width

assumptions inherent in optimal foraging theory

Even within their obvious constraints of, for example, size and morphological and physiological specialisation, most predators consume a narrower range of food types than they are capable of consuming. *Optimal foraging theory* helps us understand what determines a consumer's diet within its wide potential range by predicting foraging strategies based on a number of assumptions. The first is that the foraging behaviour we observe is the one that has been favoured by natural selection in the past but also most enhances an animal's fitness at present. Secondly, high fitness is achieved by a high net rate of energy intake (i.e. gross energy intake minus the energetic costs of obtaining that energy). And last, predators are observed in an environment to which their foraging behaviour is suited, i.e. it is a natural environment very similar to that in which they evolved, or an experimental arena similar in essential respects to the natural environment.

These assumptions will not always be justified. Firstly, other aspects of an organism's behaviour, such as predator avoidance (Section 9.2.4), may influence fitness more than optimal foraging does. Secondly, and just as important, for many consumers (particularly herbivores and omnivores) it may be of prime importance for the forager to consume a mixed and balanced diet. However, in circumstances where the energy maximisation premise can be expected to apply, optimal foraging theory offers a powerful insight into the significance of the foraging 'decisions' that predators make (for reviews see Stephens and Krebs (1986), Krebs and Kacelnik (1991) and Sih and Christensen (2001)).

theoreticians are omniscient mathematicians – the foragers need not be

Typically, optimal foraging theory makes predictions about foraging behaviour based on mathematical models constructed by modellers who are omniscient ('all knowing') as far as their model systems are concerned. We might therefore ask: is it necessary for a real forager to be equally omniscient and mathematical, if it is to adopt the appropriate, optimal strategy? The answer is 'no'. The theory simply says that if there is a forager that in some way (in any way) manages to do the right thing in the right circumstances, then this forager will be favoured by natural selection; and if its abilities are inherited, these should spread, in evolutionary time, throughout the population. Optimal foraging theory does not specify precisely how the forager should make the right decisions, and it does not require the forager to carry out the same calculations as the modeller. It simply predicts the nature of the strategy that should be favoured by natural selection.

The first paper on optimal foraging theory (MacArthur & Pianka, 1966), subsequently given sharper algebraic teeth, notably by Charnov (1976a), sought to understand the range of food types eaten by an animal within different types of habitat. It still helps us understand diet widths today. MacArthur and Pianka argued that to obtain food, any predator must expend time and energy, first in searching for its prey, and then in handling it (i.e. pursuing, subduing and consuming it). While searching, a predator is likely to encounter a wide variety of food items. MacArthur and Pianka therefore saw diet width as depending on the responses of predators once they had encountered prey. Generalists pursue (and may then subdue and consume) a large proportion of the prey types they encounter; specialists continue searching except when they encounter prey of their specifically preferred type.

to pursue or not pursue?

The ‘problem’ for any forager is this: if it is a specialist, then it will only pursue profitable prey items, but it may expend a great deal of time and energy searching for them. Whereas if it is a generalist, it will spend relatively little time searching, but it will pursue both more and less profitable types of prey. An optimal forager should balance the pros and cons so as to maximise its overall rate of energy intake. MacArthur and Pianka expressed the problem as follows: given that a predator already includes a certain number of profitable items in its diet, should it expand its diet (and thereby decrease its search time) by including the next most profitable item as well?

We can refer to this ‘next most profitable’ item as the *i*th item. E_i/h_i is then the profitability of the item, where E_i is its energy content, and h_i its handling time. In addition, $\overline{E}/\overline{h}$ is the average profitability of the ‘present’ diet (i.e. one that includes all prey types that are more profitable than *i*, but does not include prey type *i* itself), and $\overline{s}$ is the average search time for the present diet. If a predator does pursue a prey item of type *i*, then its expected rate of energy intake is E_i/h_i. But if it ignores this prey item, while pursuing all those that are more profitable, then it can expect to search for a further $\overline{s}$, following which its expected rate of energy intake is $\overline{E}/\overline{h}$. The total time spent in this latter case is $\overline{s}+\overline{h}$, and so the overall expected rate of energy intake is $\overline{E}/\left(\overline{s}+\overline{h}\right)$. The most profitable, optimal strategy for a predator will be to pursue the *i*th item if, and only if:

$$E_i/h_i \geq \overline{E}/\left(\overline{s}+\overline{h}\right). \qquad (9.1)$$

In other words, a predator should continue to add increasingly less profitable items to its diet as long as Equation 9.1 is satisfied (i.e. as long as this increases its overall rate of energy intake). This will serve to maximise its overall rate of energy intake, $\overline{E}/\left(\overline{s}+\overline{h}\right)$.

This optimal diet model leads to a number of predictions.

searchers should be generalists

1. Predators with handling times that are typically short compared with their search times should be generalists, because in the short time it takes them to handle a prey item that has already been found, they can barely begin to search for another prey item. (In terms of Equation 9.1: E_i/h_i is large (h_i is small) for a wide range of prey types, whereas $\overline{E}/\left(\overline{s}+\overline{h}\right)$ is small ($\overline{s}$ is large) even for broad diets.) This prediction seems to be supported by the broad diets of many insectivorous birds feeding in trees and shrubs. Searching is always moderately time consuming, but handling the minute insects takes negligible time and is almost always successful. A bird, therefore, has something to gain and virtually nothing to lose by consuming an item once found, and overall profitability is maximised by a broad diet.

handlers should be specialists

2. By contrast, predators with handling times that are long relative to their search times should be specialists. That is, if $\bar{s}$ is always small, then $\bar{E}/(\bar{s}+\bar{h})$ is similar to $\bar{E}/\bar{h}$. Thus, maximising $\bar{E}/(\bar{s}+\bar{h})$ is much the same as maximising $\bar{E}/\bar{h}$, which is achieved, clearly, by including only the most profitable items in the diet. For instance, lions live more or less constantly in sight of their prey so that search time is negligible; handling time, on the other hand, and particularly pursuit time, can be long (and very energy consuming). Lions consequently specialise on prey that can be pursued most profitably: the immature, the lame and the old.

specialisation should be greater in productive environments

3. Other things being equal, a predator should have a broader diet in an unproductive environment (where prey items are relatively rare and $\bar{s}$ is relatively large) than in a productive environment (where $\bar{s}$ is smaller). This prediction is broadly supported by the two examples shown in Figure 9.4: in experimental arenas, both bluegill sunfish (*Lepomis macrochirus*) and great tits (*Parus major*) had more specialised diets when prey density was higher. A related result has been reported from predators in their natural setting – brown and black bears (*Ursus arctos* and *U. americanus*) feeding on salmon in Bristol Bay in Alaska. When salmon availability was high, bears consumed less biomass per captured fish, targeting energy-rich fish (those that had not spawned) or energy-rich body parts (eggs in females, brain in males). In essence their diet became more specialised when prey were abundant (Gende *et al*., 2001).

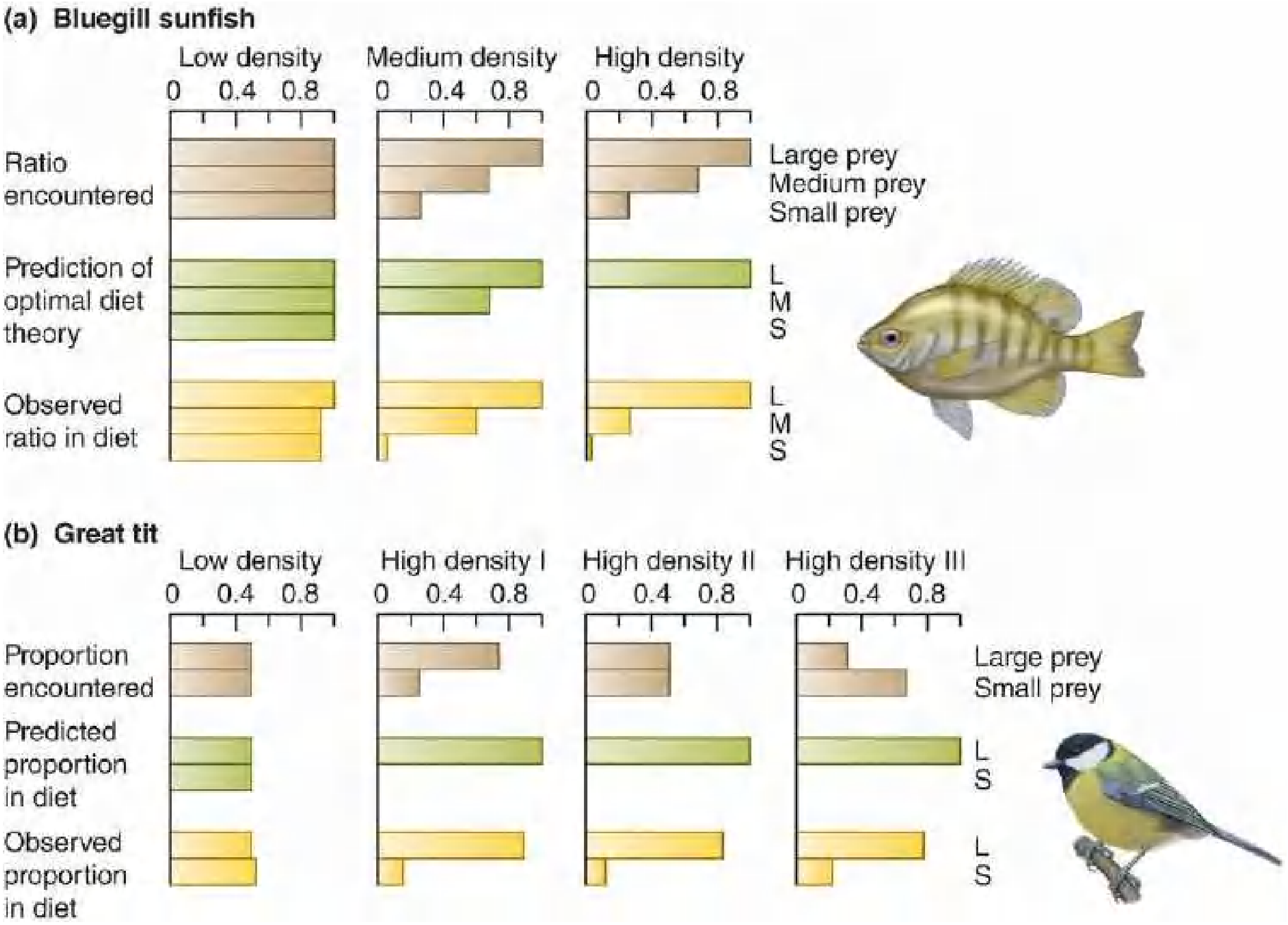


Figure 9.4 Studies of optimal diet choice showing a clear but limited correspondence with the predictions of Charnov's (1976a) optimal diet model. Diets are more specialised at high prey densities; but more low profitability items are included than predicted by the theory. (a) Bluegill sunfish preying on different size classes of *Daphnia*: the histograms show ratios of encounter rates with each size class at three different densities, together with the predicted and observed ratios in the diet. (b) Great tits preying on large and small pieces of mealworm at low and high (three replicates) density. The histograms in this case refer to the proportions of the two types of item taken.

Source: (a) After Werner & Hall (1974). (b) After Krebs *et al.* (1977).

the abundance of unprofitable prey types is irrelevant

4. Equation 9.1 depends on the profitability of the ith item (E_i/h_i), depends on the profitabilities of the items already in the diet ($\overline{E}/\overline{h}$) and depends on the search times for items already in the diet ($\overline{s}$) and thus on their abundance. But it does not depend on the search time for the ith item, s_i. In other words, predators should ignore insufficiently profitable food types irrespective of their abundance. In the examples in Figure 9.4 the optimal diet model does indeed predict that the least profitable items should be ignored completely. The foraging behaviour was very similar to this prediction, but in both cases the animals consistently took slightly more than expected of the less profitable food types. In fact, this sort of discrepancy has been uncovered repeatedly, and there are a number of reasons why it may occur, which can be summarised crudely by noting that the animals are not omniscient. The optimal diet model, however, does not predict a perfect correspondence between observation and expectation. It predicts the sort of strategy that will be favoured by natural selection, and says that the animals that come closest to this strategy will be most

favoured. From this point of view, the correspondence between data and theory in Figure 9.4 seems much more satisfactory. Sih and Christensen (2001) reviewed 134 studies of optimal diet theory, focusing on the question of what factors might explain the ability of the theory to correctly predict diets. Their major conclusion was that while optimal diet theory generally works well for foragers that feed on immobile prey (leaves, seeds, mealworms, zooplankton relative to fish), it often fails to predict diets of foragers that attack mobile prey (small mammals, fish, zooplankton relative to insect predators). This may be because variations among mobile prey in vulnerability (encounter rate and capture success) are often more important in determining predator diets than are variations in the active choices of predators (Sih & Christensen, 2001).

5. Equation 9.1 also provides a context for understanding the narrow specialisation of predators that live in intimate association with their prey, especially where an individual predator is linked to an individual prey (e.g. many parasitoids and parasitic herbivores – and many parasites (see Chapter 12)). Since their whole lifestyle and life cycle are finely tuned to those of their prey (or host), handling time $(\bar{h})$ is low; but this precludes their being finely tuned to other prey species, for which, therefore, handling time is very high. Equation 9.1 will thus only apply within the specialist group, but not to any food item outside it.

   On the other hand, polyphagy has definite advantages. Search costs ($\bar{s}$) are typically low – food is easy to find – and an individual is unlikely to starve because of fluctuations in the abundance of one type of food. In addition, polyphagous consumers can, as we've seen previously, construct a balanced diet, and maintain this balance by varying preferences to suit altered circumstances, and can avoid consuming large quantities of a toxin produced by one of its food types. These are considerations ignored by Equation 9.1.

switching? or optimal diets?

coevolution: predator–prey arms races

There may seem, at first sight, to be a contradiction between the predictions of the optimal diet model and switching. In the latter, a consumer switches from one prey type to another as their relative densities change. But the optimal diet model suggests that the more profitable prey type should always be taken, irrespective of its density or the density of any alternative. Switching is presumed to occur, however, in circumstances to which the optimal diet model does not strictly apply. Specifically, switching often occurs when the different prey types occupy different microhabitats, whereas the optimal diet model predicts behaviour within a microhabitat. Moreover, most other cases of switching involve a change in the profitabilities of items of prey as their density changes, whereas in the optimal diet model these are constants. Indeed, in cases of switching, the more abundant prey type is the more profitable, and in such a case the optimal diet model does indeed predict switching: specialisation on whichever prey type is more profitable/abundant.

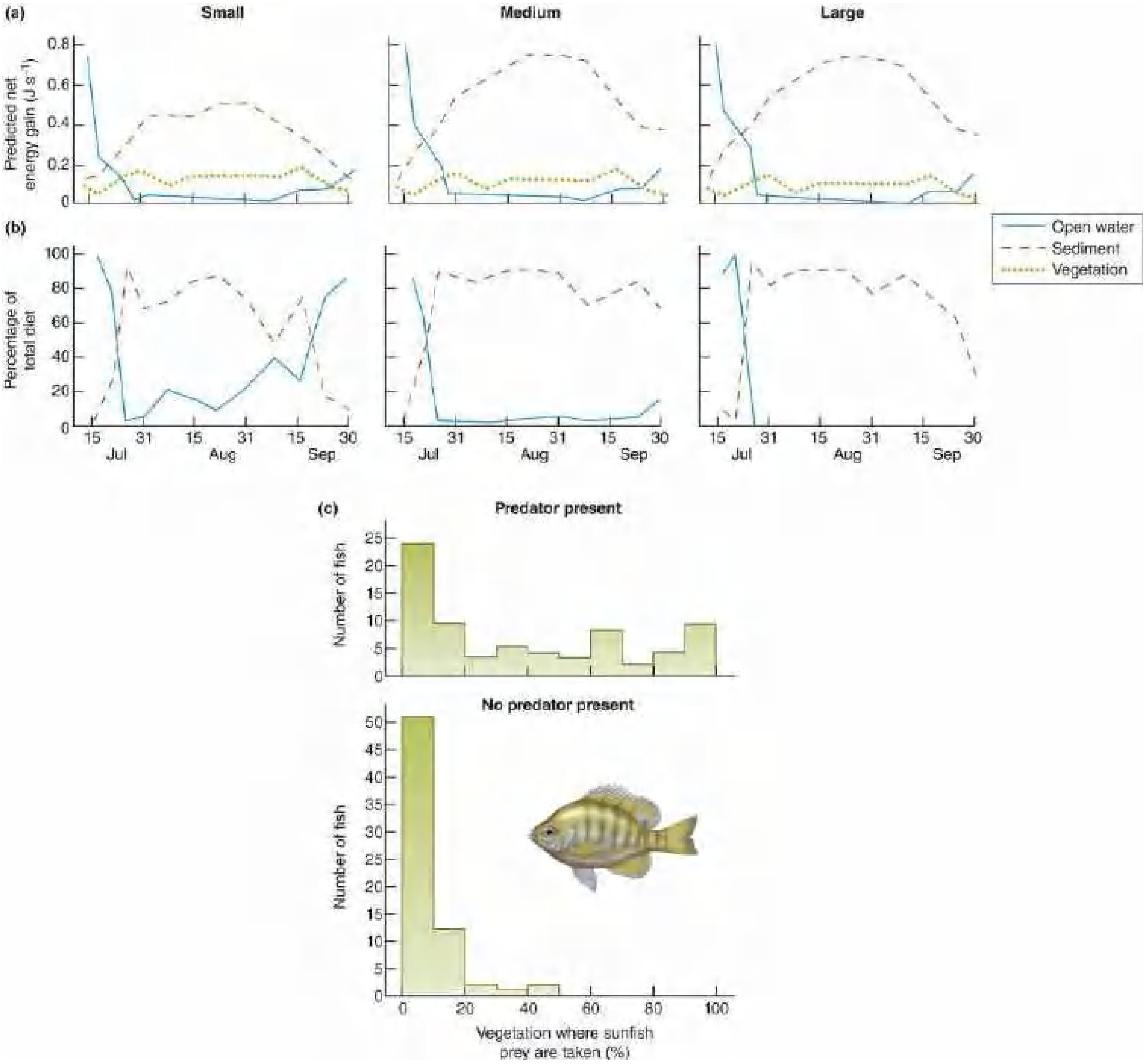
Overall, then, evolution may broaden or restrict diets. Where prey exert evolutionary pressures demanding specialised morphological or physiological responses from the consumer, restriction is often taken to extremes. But where consumers feed on items that are individually inaccessible or unpredictable or lacking in certain nutrients, the diet often remains broad. We discussed in Chapter 3 the idea that particular pairs of predator and prey species have not only evolved but have coevolved – that there has been an evolutionary 'arms race', whereby each improvement in predatory ability has been followed by an improvement in the prey's ability to avoid or resist the predator, which has been followed by a further improvement in predatory ability, and so on. This can lead consumers to some of the narrowest, and most exclusive diets, where they alone have evolved a counter-adaptation to the protective physical or chemical defences that have evolved in their prey.

9.2.4 Foraging in the presence of predators

sunfish forage suboptimally but avoid being preyed on ...

We have seen already that foraging strategies will not always be strategies for simply maximising feeding efficiency. On the contrary, natural selection will favour foragers that maximise their net benefits, and strategies will therefore often be modified by other, conflicting demands on the individuals concerned. In particular, the need to avoid predators will frequently affect an animal's foraging behaviour.

We can see this in studies carried out by Werner *et al.* (1983a ,b) on bluegill sunfish, *Lepomis macrochirus*. They estimated the net energy returns from foraging in three contrasting laboratory habitats – in open water, amongst water weeds and on bare sediment – and they examined how prey densities varied in comparable natural habitats in a lake through the seasons. They were then able to predict the time at which the sunfish should switch between different lake habitats so as to maximise their overall net energy returns. In the absence of predators, three sizes of sunfish behaved as predicted (Figure 9.5a, b). But in a further field experiment, this time in the presence of predatory largemouth bass (*Micropterus salmoides*), the small sunfish restricted their foraging to the water weed habitat (Figure 9.5c). Here, they were relatively safe from predation, although they could only achieve a markedly submaximal rate of energy intake. By contrast, the larger sunfish are more or less immune from predation by bass, and they continued to forage according to the optimal foraging predictions. In a similar vein, the nymphs of several species of algivorous mayflies largely restrict their feeding to the hours of darkness in streams that contain brown trout, reducing their overall feeding rates but also reducing the risk of predation (Townsend, 2003).



[**Figure 9.5**](#) **The foraging behaviour of bluegill sunfish changes in the presence of predators.** Seasonal patterns in (a) the predicted habitat profitabilities (net rate of energy gain) and (b) the actual percentage of the diet originating from each habitat, for three size classes of bluegill sunfish (*Lepomis macrochirus*). Predators were absent. (The 'vegetation' habitat is omitted from (b) for the sake of clarity – only 8–13% of the diet originated from this habitat for all size classes of fish.) There was good correspondence between the patterns in (a) and (b). (c) When largemouth bass (*Micropterus salmoides*, which prey on small bluegill sunfish) were present (upper panel), in contrast to (b), and to the lower panel, many sunfish took prey from areas where the percentage of vegetation was high and where they were relatively protected from predation.

Source: (a, b) After Werner *et al*. (1983b). (c) After Werner *et al*. (1983a).

predation and the realised niche

The places where animals occur, where they are maximally abundant and where they choose to feed are all key components of their 'realised niches'. We saw in Chapter 8 that realised niches can be defined by competitors. Now we see that they can also be defined by predators. For example, predation by the barn owl (*Tyto alba*) affects the foraging behaviour of three heteromyid rodents, the Arizona pocket mouse (*Perognathus amplus*), Bailey's pocket mouse (*P. baileyi*) and Merriam's kangaroo rat (*Dipodomys merriami*) (Brown *et al.*, 1988). In the presence of owls, all three species moved to microhabitats where they were less at risk from owl predation and where they reduced their foraging activity. However, they did so to varying extents, such that the way in which the microhabitat was partitioned between them was quite different in the presence and absence of owls. Fine *et al.* (2004) describe similar results for the effects of herbivores on habitat specialisation among trees in the Peruvian Amazon.

If predators affect the behaviour of predators, then this 'intimidation' is likely, in turn, to have effects on the foraging of those prey, in addition to the direct, consumptive effects that their predators have on them. We return to these effects in Section 9.6.1.

9.3 Plants' defensive responses to herbivory

In the face of predation, prey are unlikely to remain passive either on ecological or evolutionary timescales. As a result, the effects of predation on prey are often not as serious as they first appear to be: a comment equally applicable to individual prey and to prey populations. This is especially true of the effects of herbivory on plants, in part because so many herbivores are grazers (leaving part of their plant prey alive), and in part because of the plants' own modular nature, which makes the loss of body parts relatively tolerable. Hence, we focus in this and the next section mostly on plant defences against herbivory.

The effects of herbivory depend on which herbivores are involved, which plant parts are affected, and the timing of attack relative to the plant's development. The consequences of defoliating a germinating seedling are unlikely to be the same as those of defoliating a plant that is setting its own seed. Leaf biting, sap sucking, mining, flower and fruit damage, and root pruning are all likely to differ in the effect they have on the plant, and of course, some herbivores will have a greater impact than others. The effects of herbivory are also crucially dependent on the response of the plant.

deterrence, resistance and tolerance

As a framework for understanding prey defence, whether the interaction is plant–herbivore, prey–carnivore or parasite–host, we can distinguish three broad strategies: *deterrence*, where the attack rate of the consumer is slowed down or stopped altogether by diverting the consumer away from the protected prey, *resistance*, where the consumer is injured or killed, and *tolerance*, where the prey compensates for any harm or losses that the consumer may cause. All three, including tolerance, involve active investment by the prey in substances, or structures, or metabolic processes that would not be required in the absence of consumers. But as we shall see, it is often not possible to classify such investments into just one of these three categories. For example, a poisonous chemical produced by a prey may harm a consumer that attacks it but also deter that consumer from attacking subsequent potential prey that look similar. We begin here with plant resistance to attack, reducing the effectiveness of the herbivore (Sections 9.3.1–9.3.5) before turning to plants' tolerance to herbivore damage (Sections 9.4). When we turn lastly to the defences used by animal prey (Section 9.5) we will look more closely at deterrence.

9.3.1 Plant defences

constitutive and inducible defences

Defence itself may be either *constitutive* (produced independently of any actual or 'anticipated' attack by consumers) or *inducible* (produced only in response to such attack). It is easy to see the pros and cons of both. Constitutive defence is always available, but because it is produced, at least sometimes, when not immediately required, it is costly. Inducible defences are cheaper in that they are only produced when required, but any delay in their production may allow consumers to feed from and seriously harm the not-yet-protected prey. Cutting across this distinction, chemical defence may be either *direct*, where the chemical harms or deters the consumer, or *indirect*, where the effect is to attract natural enemies of the consumer. Constitutive direct defences of plants include many types of spines and thorns; constitutive *indirect* defences include the extrafloral nectaries that may attract ants to a plant, which then attack its consumers (this mutualism between ants and plants is discussed further in Chapter 12). On the other hand, as we discuss in Section 9.3.3, herbivore attack may induce a plant to increase its production of both directly harmful chemicals and those that attract consumers' natural enemies. The production of protective spines, for example, may also be increased by herbivore attack.

coevolutionary arms races

It is no surprise that most organisms have evolved these physical, chemical, morphological and/or behavioural defences (in the case of animal prey) that reduce the chance of an encounter with a consumer (deterrence) and/or increase the chance of surviving such an encounter (resistance and tolerance). However, a better defended food resource itself exerts a selection pressure on consumers to overcome that defence, and in overcoming that defence, rather than the defences of other species, the consumer is likely to become relatively specialised on that resource – which is then under particular pressure to defend itself against that particular consumer, and so on. This suggests a continuing interaction in which the consumer and the prey drive one another's evolution: a coevolutionary 'arms race'. The idea is usually traced back to Ehrlich and Raven (1964). They, like much work since, were particularly focused on plant–herbivore interactions, perhaps because, unlike animal prey, plants cannot simply run or fly away from their predators and so must rely on physical or chemical defences to resist or deter them. They also emphasised the potential of this process to generate a diversity of species amongst both the plants and the herbivores (see Marquis *et al*., 2016), since interbreeding between emerging specialists and either generalists or different specialists could readily hinder the evolution of the required degree of specialisation (see Section 1.3).

In its most extreme form, the arms race has a co-adapted pair of species locked together in perpetual struggle, but 'diffuse coevolution' is also possible, in which a plant, for example, evolves a new defence, which allows it to proliferate and diversify into new niches and hence new species, with herbivore speciation following at a later evolutionary stage (see Stamp (2003) for a review). However, demonstrable coevolution between predators and prey is by no means a general rule. For example, in a study carried out on species from the Amazon region of Peru, phylogenies were constructed of trees of the species-rich genus *Inga* (*c*.300 species described in total) and of their lepidopteran herbivores (Endara *et al*., 2017). The trees vary in their defences – chemical, physical and biotic (for example, protection by mutualistic ants – see Chapter 13) – and these too were characterised, allowing the construction of a 'defensogram': a branched structure similar to a phylogeny, in which the *Inga* species most similar defensively were the most closely linked, just as the most taxonomically similar species are closest in a phylogeny. For the *Inga* species, there was little congruence between the defensogram and the phylogeny – taxonomically-distant species often shared similar defences. Most significantly, though, the phylogenies of the herbivores mapped badly onto the *Inga* phylogeny but mapped well onto the defensogram (Figure 9.6). Thus, the radiations of herbivores have not followed the radiations of the plants in a process of diffuse coevolution, but appear to have followed them only in the sense of attacking those species to whose defences they are already pre-adapted.

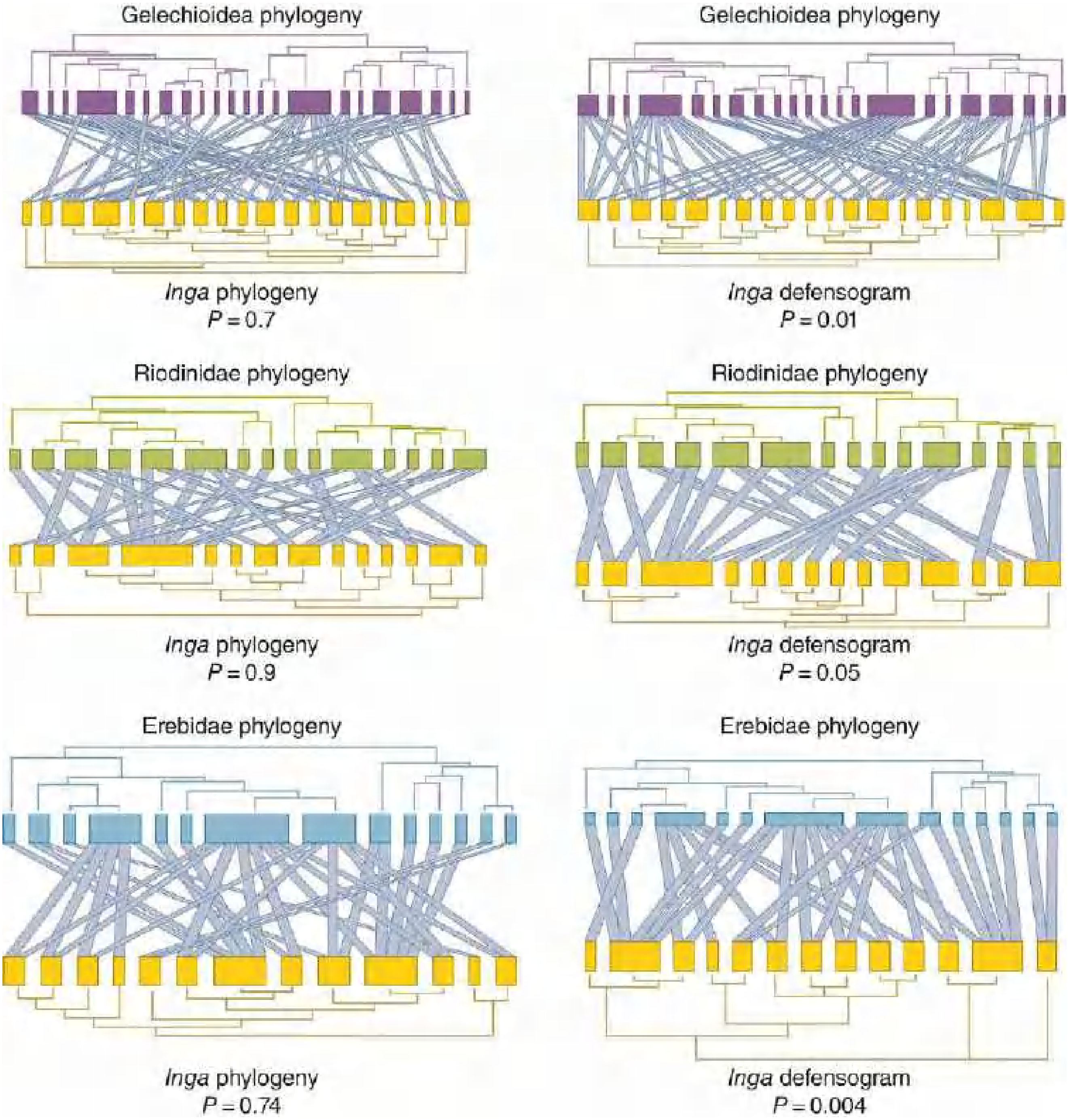


Figure 9.6 The phylogenies of Peruvian Lepidoptera map badly onto the phylogeny of the trees on which they feed, but well onto a defensogram of those trees. For three families/superfamilies of Lepidoptera that feed on species of *Inga* trees in Peru (Gelechioidea (purple), Riodinidae (green) and Erebidae (blue)), the panels on the left show the association between their phylogeny and that of the *Inga* species on which they feed (yellow). None is significant. Panels on the right show equivalent associations but with *Inga* 'defensograms' (species grouped with those most similar to them in terms of defence, not evolutionarily). All are significant. For each network, the widths of bars represent the abundance of the species concerned.

Source: After Endara *et al*. (2017).

physical and chemical defences

Simple spines are an effective deterrent to the consumption of plants (such as holly) and animals (such as hedgehogs), and at a smaller scale, many small planktonic invertebrates develop spines, crests and other appendages that reduce their vulnerability to predation in the presence of predators. At that smaller scale, many plant surfaces are clothed in epidermal hairs (trichomes), and in some species these develop thick secondary walls to form strong hooks or points that may trap or impale insects (though they are also often covered in toxic chemicals). Protection can also be provided by the thick shell of a nut, or the fibrous cone on a pine, if, as a consequence, the consumer eats less of it.

plant secondary chemicals

Green plants use none of their energetic resources in running away. Hence, they have more available to invest in such defence structures. They may, in particular, be relatively overprovided with energy resources, making it cheap to build shells around seeds, and woody spines on stems, mainly out of cellulose and lignin, protecting the real riches: the scarce resources of nitrogen, phosphorus, potassium, etc. in the embryos and meristems. Without doubt, though, most studies of plant defence have focused on chemicals. Plant defence chemicals have traditionally been considered to play no role in the normal (‘primary’) metabolic pathways of plant biochemistry that support growth development and reproduction. Hence, they are referred to as ‘secondary metabolites’ and range from simple molecules like oxalic acid and cyanide to the more complex glucosinolates, alkaloids, terpenoids, saponins, flavonoids and tannins (Hartmann, 2007). As such, it has often been assumed that there is a trade-off between the energy and resources that can be invested in secondary chemicals and those invested in growth, development and reproduction – and hence a cost associated with the production of secondary chemicals in terms of reduced growth and so on (Strauss *et al.*, 2002). Certainly, there are some examples where such costs have been demonstrated. One study involved ‘knock-out’ mutants of the model plant, *Arabidopsis thaliana*, in which genes coding for the production of herbivore-defensive compounds, notably glucosinolates, were silenced (‘knocked out’). These were grown alongside normal, wild-type plants, and their growth rates noted after five days (Figure 9.7a), following which aphids were allowed to attack the plants and their growth rates noted again after 29 days (Figure 9.7a). Prior to the introduction of the aphids, the mutants, relieved of the burden of glucosinolate production, mostly grew much more strongly than the wild-type plants. But by 29 days, following aphid attack, they had paid the price, and their growth rates were lower than the wild-type plants. A subsequent analysis of the metabolic economics of glusosinolate biosynthesis put a figure on this cost, estimating that their production would increase plants’ photosynthetic requirements by around 15% (Bekaert *et al.*, 2012).

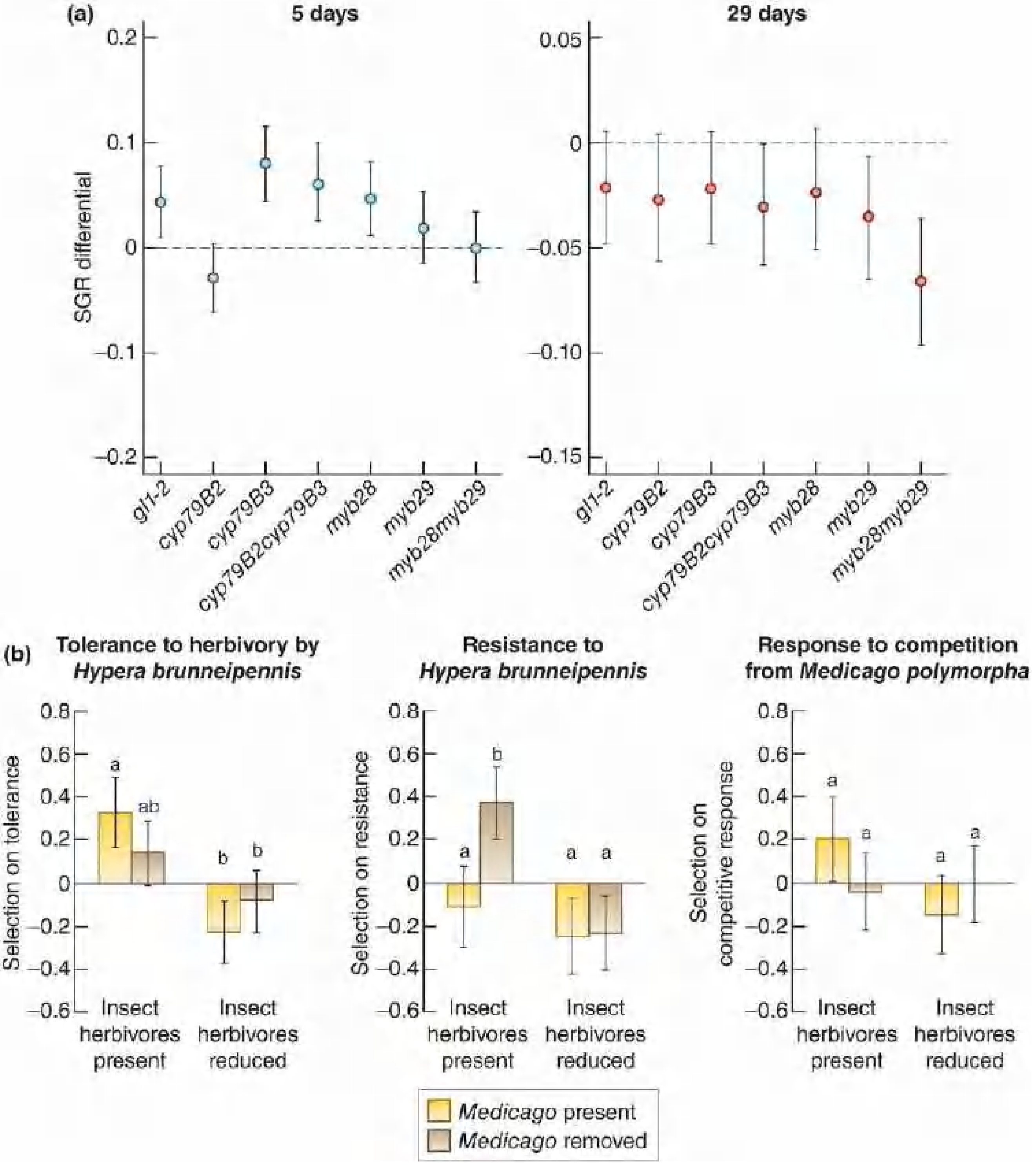

(a)
5 days
29 days
SGR differential
0.2
0.1
0
−0.1
−0.2
0.05
0
−0.05
−0.10
−0.15
gl1-2
cyp79B2
cyp79B3
cyp79B2cyp79B3
myb28
myb29
myb28myb29
(b)
Tolerance to herbivory by Hypera brunneipennis
Resistance to Hypera brunneipennis
Response to competition from Medicago polymorpha
Selection on tolerance
Selection on resistance
Selection on competitive response
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
a
ab
b
b
a
b
a
a
a
a
a
a
Insect herbivores present
Insect herbivores reduced
Medicago present
Medicago removed

Figure 9.7 **Demonstrable costs of plant defence against herbivores.** (a) Difference in the relative growth rate (weight gain per unit weight) standardised for size (SGR) between a range of *Arabidopsis thaliana* genotypes deficient in defensive glucosinolate production (as indicated) and the wild type, after five days, where the mutants were growing faster (left), suggesting a cost of glucosinolate production, and after 29 days (right), with aphids having been introduced after five days, where they were now growing more slowly. Bars are 95% CIs. (b) Selection differentials of *Lotus wrangelianus* for tolerance to herbivory (plants' ability to regrow and reproduce following attack by the beetle *Hypera brunneipennis*; left panel), resistance to the beetle (the proportion of leaflets surviving attack; middle) and the degree to which fitness is reduced by competition from *Medicago polymorpha* (right). Positive differentials indicate that genotypes with this trait are more favoured, and *vice versa*. In the presence of herbivores (left of each panel), tolerant genotypes were favoured but resistant genotypes only favoured when there were no competing plants (brown bar, middle panel). However, when herbivory was reduced (right of each panel), those tolerant and resistant genotypes were at a disadvantage, suggesting a cost associated with these traits. Bars are SEs. Values within a panel bearing the same letter are not significantly different.

Source: (a) After Züst *et al*. (2011). (b) After Lau *et al*. (2008).

Defence costs have also been demonstrated in more natural settings. For example, when the native plant *Lotus wrangelianus* was grown in Napa County, California, USA in the presence of the introduced herbivorous beetle *Hypera brunneipennis*, those *Lotus* genotypes that were most resistant to attack fared best (Figure 9.7b, middle panel). But when the beetles were suppressed with insecticide, the advantage of the resistant genotypes was lost. In other words, genotypes bearing the costs of defence for no benefit (since there were few beetles) performed poorly. Moreover, when the beetles were present, but there were also high densities of the introduced plant, *Medicago polymorpha*, with which *Lotus* competes, the advantages of defence were negated, presumably because the costs paid for defence were thereby unavailable for competition (no significant differences in the bottom right-hand column in Figure 9.7b).

However, these examples of demonstrable defence costs remain relatively few. What's more, many 'defence' chemicals play other roles too, and hence the idea that defence necessarily carries a cost in terms of reduced growth and reproduction has itself been called into question, at least as a general rule. For example, anthocyanins are also important for osmoregulation and in regulating seed longevity and dormancy, cyanogenic glucosides also play a role in nitrogen storage and transport, and monoterpenes can regulate mitochondrial respiration and aid root growth (Nielsen *et al*., 2013). Certainly, the distinction between primary and secondary plant metabolism seems to be more blurred than has often been assumed.

patterns in plant chemical production: optimal defence?

Since all plants are attacked by herbivores, and are therefore in need of protection, the question naturally arises as to how it is that plants vary in the types and levels of their defences – and related to this, how it is that there are as many generalist herbivores as there are, consuming many plant species in spite of their varied defences. A number of hypotheses have been proposed and tested (Stamp, 2003), but the most influential has been the optimal defence hypothesis, which itself can be broken down into four linked hypotheses, as follows. (1) Organisms evolve defences in direct proportion to their risk of predation and in inverse proportion to their costs – often referred to as 'apparency theory', as explained in Section 9.3.2. (2) Contemporary investment in defence is increased in direct response to a perceived increase in the risk of predation (Section 9.3.3). (3) Within organisms, investment in defence is greatest in organs and tissues that are most important for the organism's fitness (Section 9.3.4). (4) Investment in defence is traded off against investments in other vital functions (see previously), and so organisms with lower resource levels at their disposal invest less in defence (Section 9.3.5).

9.3.2 Apparency theory

Apparency theory (Feeny, 1976 ; Rhoades & Cates, 1976 ; see also Hartley & Jones (1997) for a review) proposes, firstly, that noxious plant chemicals can be classified broadly into two types. Firstly, toxic, or 'qualitative' chemicals are poisonous even in small quantities and tend to be small molecules, and so tend, too, to be relatively cheap to produce. Examples include alkaloids, glucosinolates and cyanogenic glycosides. By contrast, digestion-inhibiting, 'quantitative' chemicals tend to be larger molecules and act in proportion to their concentration, such that they are made in larger amounts and tend, therefore, to be relatively expensive to produce. Examples include tannins and phenolics. These quantitative chemicals, because of their non-specific mode of action, should be effective against both specialist and generalist herbivores, whereas noxious, qualitative chemicals should be effective against generalists but susceptible to coevolutionary counter-adaptations by specialists.

The theory then proposes that plants themselves can be classified broadly into two types. 'Apparent' plants tend to be large, long-lived and slow-growing, like oak trees, whereas 'unapparent' plants tend to be small, ephemeral and fast-growing, like many weeds. Drawing this together, apparent plants, precisely because they are apparent for long, predictable periods to a large number of herbivores, should invest in digestion-reducing chemicals that, while costly, will afford them broad protection; whereas unapparent plants should produce toxins, because these plants will be most often found by generalists (against which the toxins will be effective), while co-adapted specialists may often fail to find them.

There are certainly studies that support apparency theory, including those that originally inspired it. For example, in the bracken fern (*Pteridium aquilinum*), the young leaves that push up through the soil in spring are less apparent to potential herbivores than the luxuriant foliage of late summer. The young leaves are rich in cyanogenic glucosinolates, while the tannin content steadily increases in concentration to its maximum in mature leaves (Rhoades & Cates, 1976). More generally, a meta-analysis of 158 studies found, as predicted, a preponderance of quantitative chemicals produced by woody (apparent) plants and of qualitative chemicals produced by herbaceous (unapparent) plants (Figure 9.8). However, this preponderance was far from complete: only 68 out of 94 studies supported apparency theory in the case of woody plants, and only 82 out of 114 studies supported it for herbaceous ones. Further, again as predicted, the woody, apparent plants seemed equally well protected against specialist and generalist herbivores, but contrary to apparency theory, this was true, too, of herbaceous plants (Smilanich *et al*., 2016).

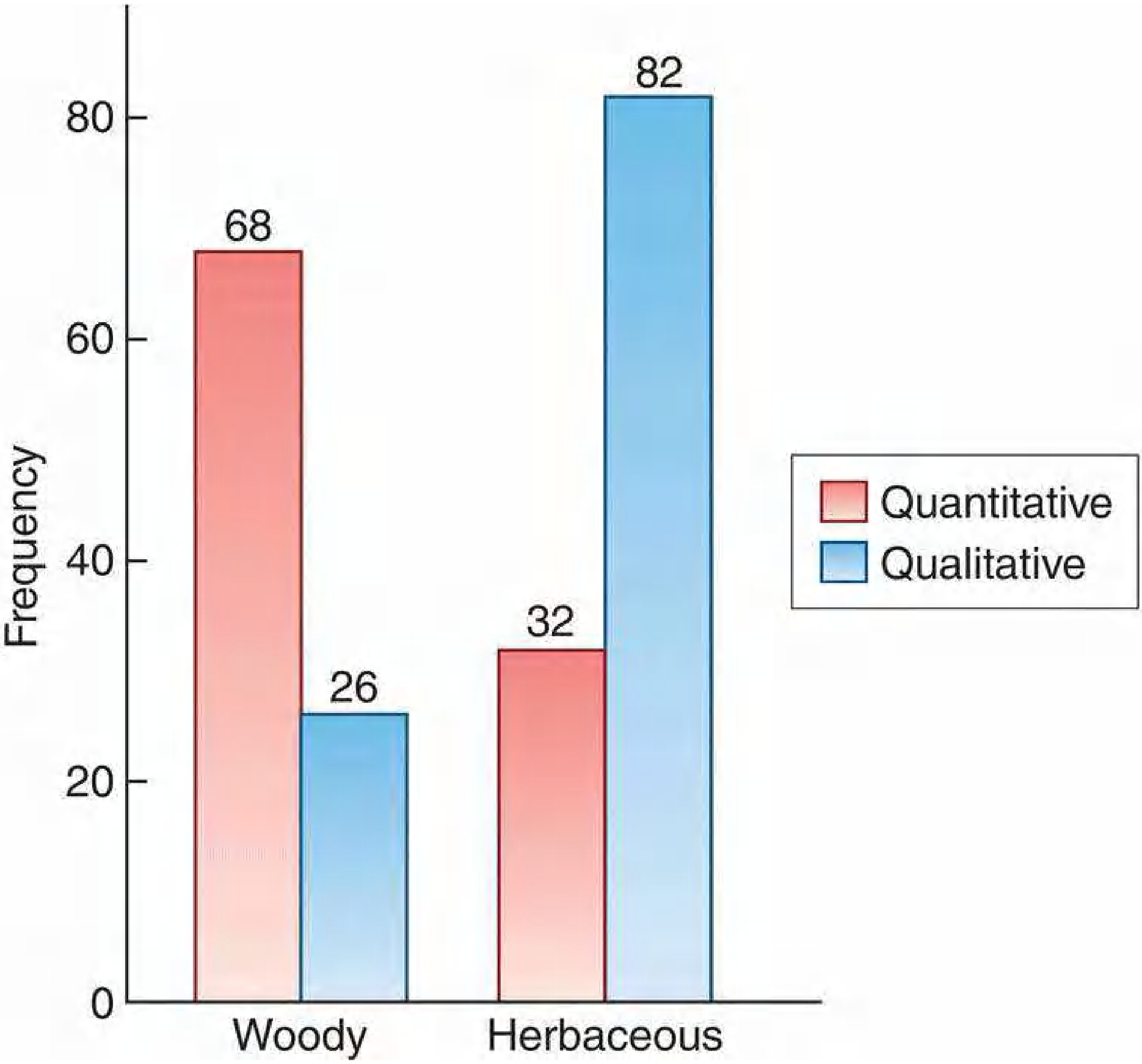


Figure 9.8 A meta-analysis of studies of the frequency of production of qualitative and quantitative plant defensive chemicals reported in woody and herbaceous plants.

Source: After Smilanich *et al.* (2016).

Such mixed results are typical of tests of apparency theory, and given this, it would be hard to claim that the theory has been fully validated. This is not surprising. The theory itself, like theories generally, was based, for clarity, on a series of simplifications, whereas it is now clear that apparency can be hard to define, and that simple proxies for variations in apparency like 'woody or herbaceous' are bound not to be watertight. Similarly, the distinction between quantitative and qualitative chemicals can be blurred (terpenes act as quantitative chemicals against some herbivores but as qualitative chemicals against others), and some qualitative chemicals are also costly to produce (Stamp, 2003). Hence, the range of determinants of the patterns of plant chemical production, and of herbivore attack, are undoubtedly broader than a simple apparent/unapparent distinction (Smilanich *et al.*, 2016). Nonetheless, apparency theory remains a useful starting point in our attempts to understand these patterns, as long as we do not expect too much of it.

9.3.3 The timing of defence: induced chemicals

The second optimal defence hypothesis predicts that investment in defences should be highest at times and in places where the threat from enemies is greatest. For plants, given that defences may be either induced or produced constitutively, this has most often been addressed by studying the pattern and timing of induced defences relative to these patterns of risk (see Karban (2011) and Kant *et al*. (2013) for reviews).

jasmonates

Induction itself is a process with (at least) two stages: first the presence of herbivores (or pathogens) activates the production of plant hormones, and then second, these hormones promote (actually, de-repress) the production of plant defensive chemicals. Probably the most important group of these hormones are jasmonates, especially jasmonic acid, studies of which emphasise how further stages in the process arise because other signals modulate this pathway (Ballaré, 2010). For example, cytokinins are a class of plant hormones promoting cell division that are therefore present at high concentrations in young tissues. Higher levels of these promote higher levels of jasmonates, and induced defences therefore tend to be greater in such tissues. Also, many plants respond to herbivory by producing various volatile organic compounds, which also increase the synthesis of jasmonates. Crucially, though, being volatile, they can induce jasmonate synthesis in branches distant from the ones being attacked and even in neighbouring plants – plants ‘talking’ to one another.

An example in which an indirect defence is induced by herbivory is shown in Figure 9.9. Passionfruit plants, *Passiflora edulis*, from Argentina, were reared and either damaged (small punctures) to simulate herbivory, sprayed with methyl jasmonate to simulate the plants’ immediate response to herbivory, or left as unmanipulated controls. These plants protect themselves from herbivores by producing nectar in extrafloral nectaries, which attracts ants to the plants, which then attack herbivores that would otherwise feed there. Nectar volume was massively increased by both simulated herbivory and the simulated jasmonate response (Figure 9.9).

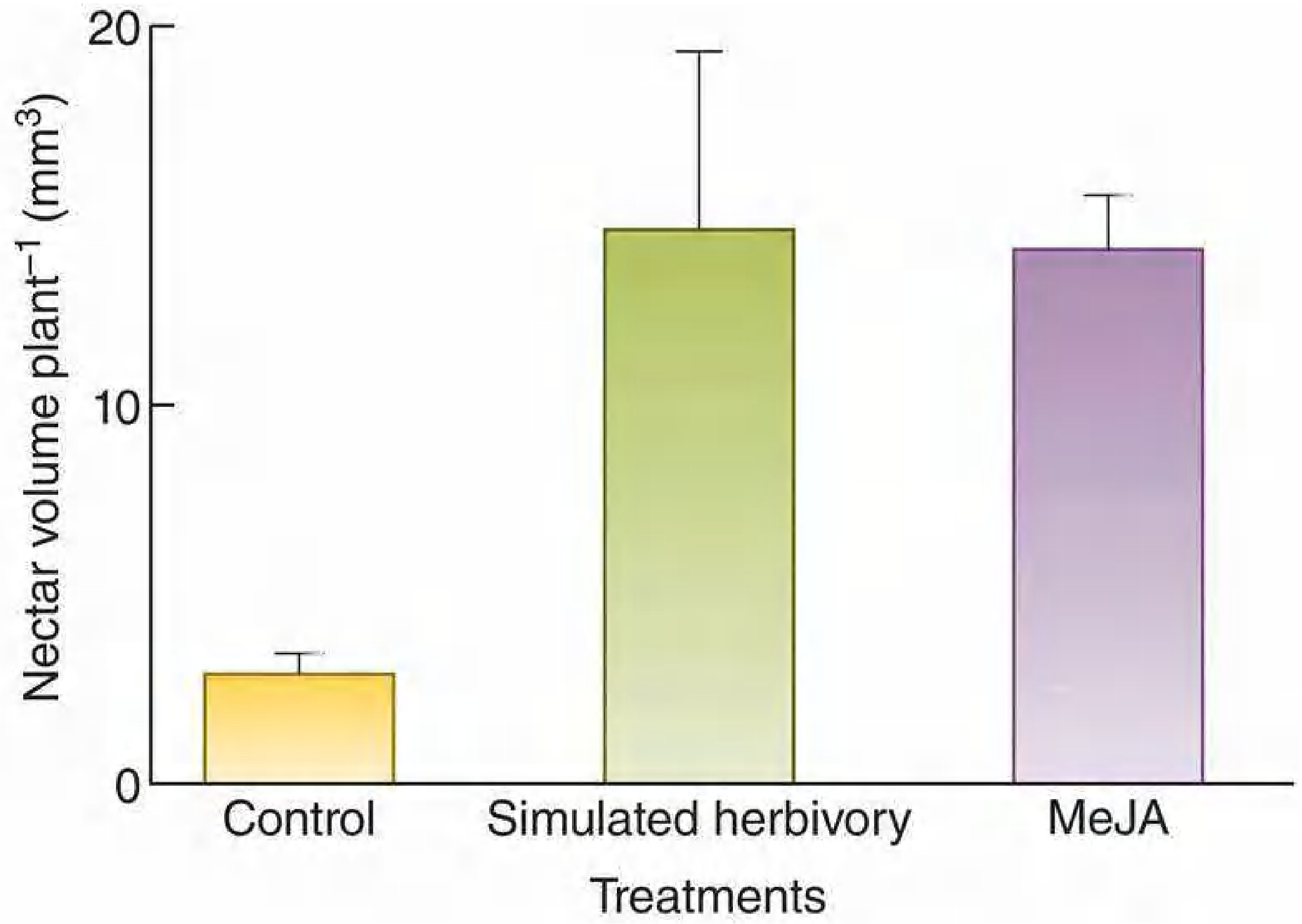
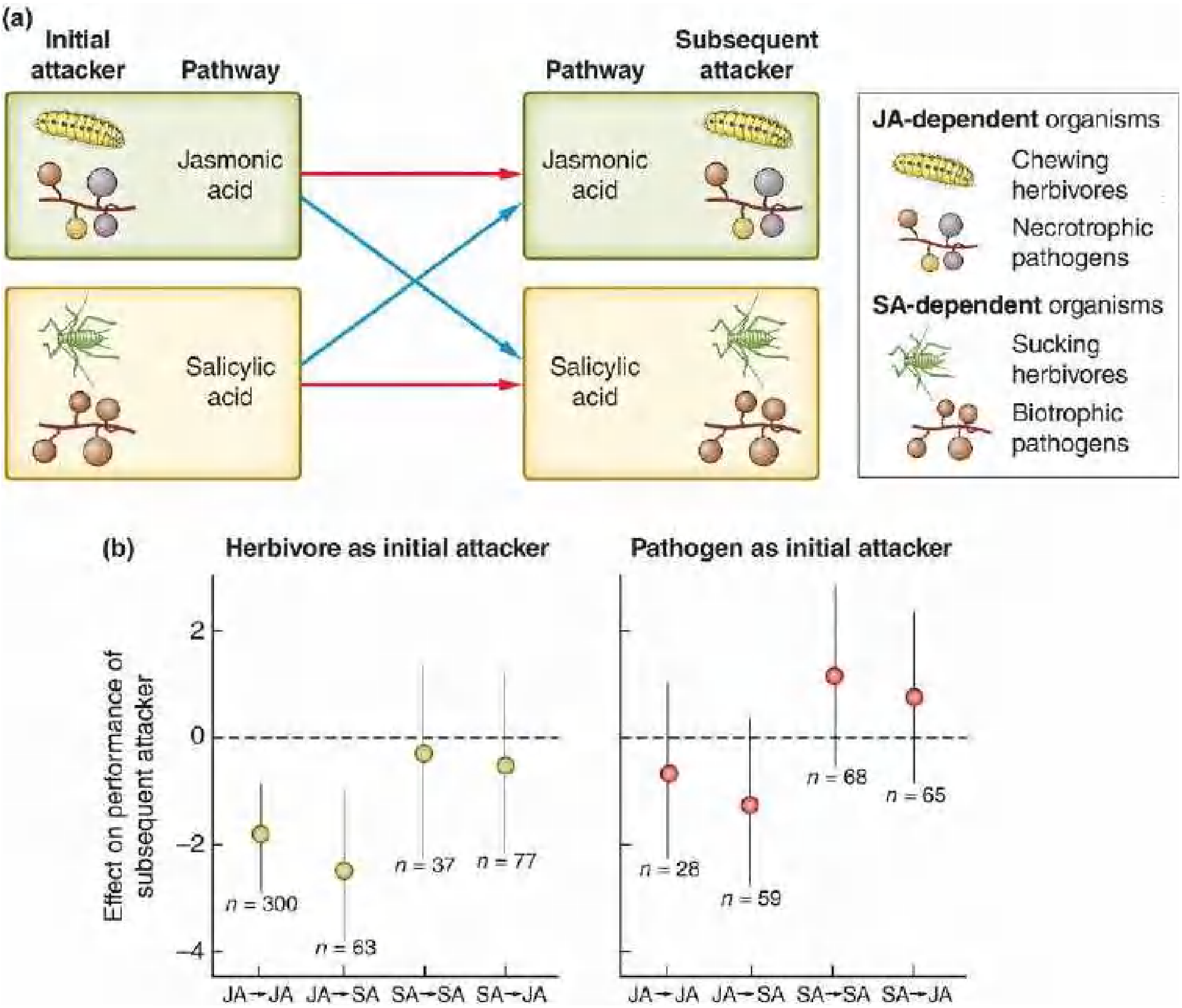


Figure 9.9 Passionfruit investment in extrafloral nectaries is increased by herbivory. The quantity of nectar produced from extrafloral nectaries in passionfruit plants, *Passiflora edulis*, subjected to simulated herbivory ($P = 0.002$, compared with control) and treated with methyl jasmonate, MeJA ($P < 0.0001$, compared with control). Bars are SEs.

Source: After Izaguirre *et al*. (2013).

jasmonates and salicylates

Another group of plant hormones, salicylates, and in particular salicylic acid, induce their own defence-signalling pathway, and have been said to depress jasmonate-induced defences (Thaler *et al*., 2012). This is essentially a trade-off-based argument – investment in one pathway leaving less resource available for investment in the other – which may be adaptive since each tends to be stimulated by different types of herbivore (Figure 9.10a). The commonly hypothesised effects are also set out in Figure 9.10a: initial attackers that induce defences by stimulating the jasmonic acid (JA) pathway thereby harm subsequent attackers stimulating that pathway, but they benefit subsequent attackers stimulating the salicylic acid (SA) pathway; there are reciprocal effects from initial attackers inducing the SA pathway. A meta-analysis of 774 case studies, however, showed limited support for such cross-talk between the JA and SA pathways (Figure 9.10b). JA-inducing attackers did indeed harm subsequent JA-inducers (though the effect was only significant when the initial attackers were herbivores, not pathogens), but this effect was extended, too, to subsequent SA-inducers, contrary to the hypothesised pattern. For SA-inducing attackers, none of the effects was significant. Of course, coming from a meta-analysis, these results report overall patterns: hidden within them are examples that support the idea of a JA–SA antagonism. Nonetheless, there are clearly no universal rules.

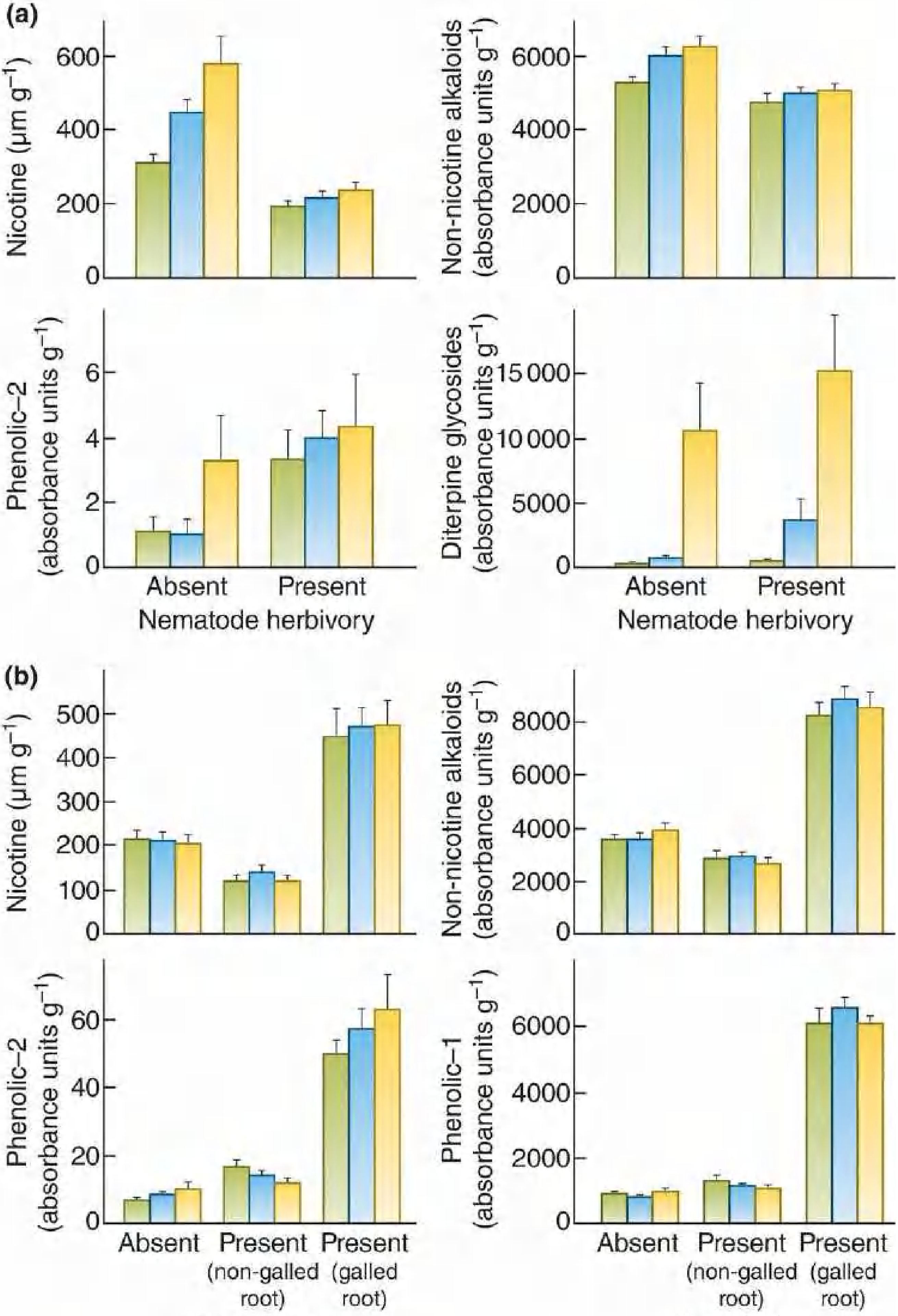


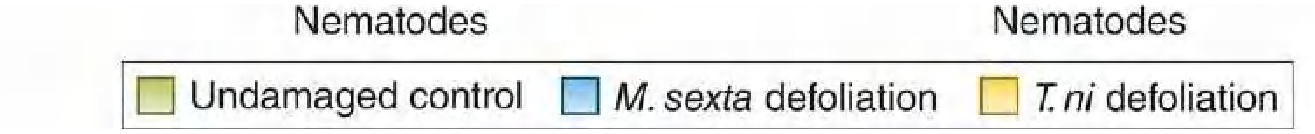
[Figure 9.10](#) Limited support for the cross-talk hypothesis between JA- and SA-inducing attacks. (a) The cross-talk hypothesis between JA- and SA-inducing herbivores and plant pathogens. Red arrows indicate negative effects, blue arrows positive effects. The organism types are typical rather than invariable. (b) Mean effect sizes (with 95% CIs) in a meta-analysis for the various combinations of pathways, with either herbivores (left) or pathogens (right) as initial attackers, as indicated (JA→JA means both the initial and subsequent attackers were JA-inducers, etc.). *n* is the number of studies in each case. Effect sizes were measured by Hedges' *d* statistic, the standardised mean difference between induced and control plants.

Source: After Moreira *et al.* (2018).

shoots and roots

We get some further sense of the complexities of induced defences from the data in Figure 9.11. Tobacco plants, *Nicotiana tabacum*, were grown with and without herbivores attacking their leaves (larvae of the tobacco hornworm, *Manduca sexta*, and the cabbage looper, *Trichoplusia ni*) and their roots (the gall forming root-knot nematode, *Meloidogyne incognita*). Concentrations of a range of inducible defence chemicals were then measured, again in both the leaves and the roots. Overall, the leaf herbivores, especially the generalist, *T. ni*, had strong effects on the concentrations of chemicals in the leaves (Figure 9.11a, nematodes absent), but induced no significant effects in the roots (Figure 9.11b, nematodes absent). By contrast, the root-knot nematode induced changes in both roots and leaves, but whereas in the roots themselves the concentrations of virtually all chemicals were elevated (Figure 9.11b, contrasting nematodes absent and present, especially where they formed galls), in the leaves effects were much more variable (Figure 9.11a). For chemicals synthesised in the roots, the nematodes, in increasing the chemicals' concentrations there, reduced their concentrations in the leaves (for example, nicotine and non-nicotine alkaloids). However, for chemicals synthesised in the leaves, nematodes in the roots increased their concentrations in both the roots and the leaves (for example, 'phenolic-2'). With variations in sites of attack, sites of synthesis, costs of synthesis, and in the ability of herbivores to derail or circumvent production, it is no surprise that definitive support for the proposal that patterns of induction are demonstrably evolutionarily optimal is hard to come by.


(a)
Nicotine (µm g^{-1})
0
200
400
600
Non-nicotine alkaloids (absorbance units g^{-1})
0
2000
4000
6000
Phenolic–2 (absorbance units g^{-1})
0
2
4
6
Absent
Present
Nematode herbivory
Diterpine glycosides (absorbance units g^{-1})
0
5000
10 000
15 000
Absent
Present
Nematode herbivory
(b)
Nicotine (µm g^{-1})
0
100
200
300
400
500
Non-nicotine alkaloids (absorbance units g^{-1})
0
2000
4000
6000
8000
Phenolic–2 (absorbance units g^{-1})
0
20
40
60
Absent
Present (non-galled root)
Present (galled root)
Phenolic–1 (absorbance units g^{-1})
0
2000
4000
6000
Absent
Present (non-galled root)
Present (galled root)



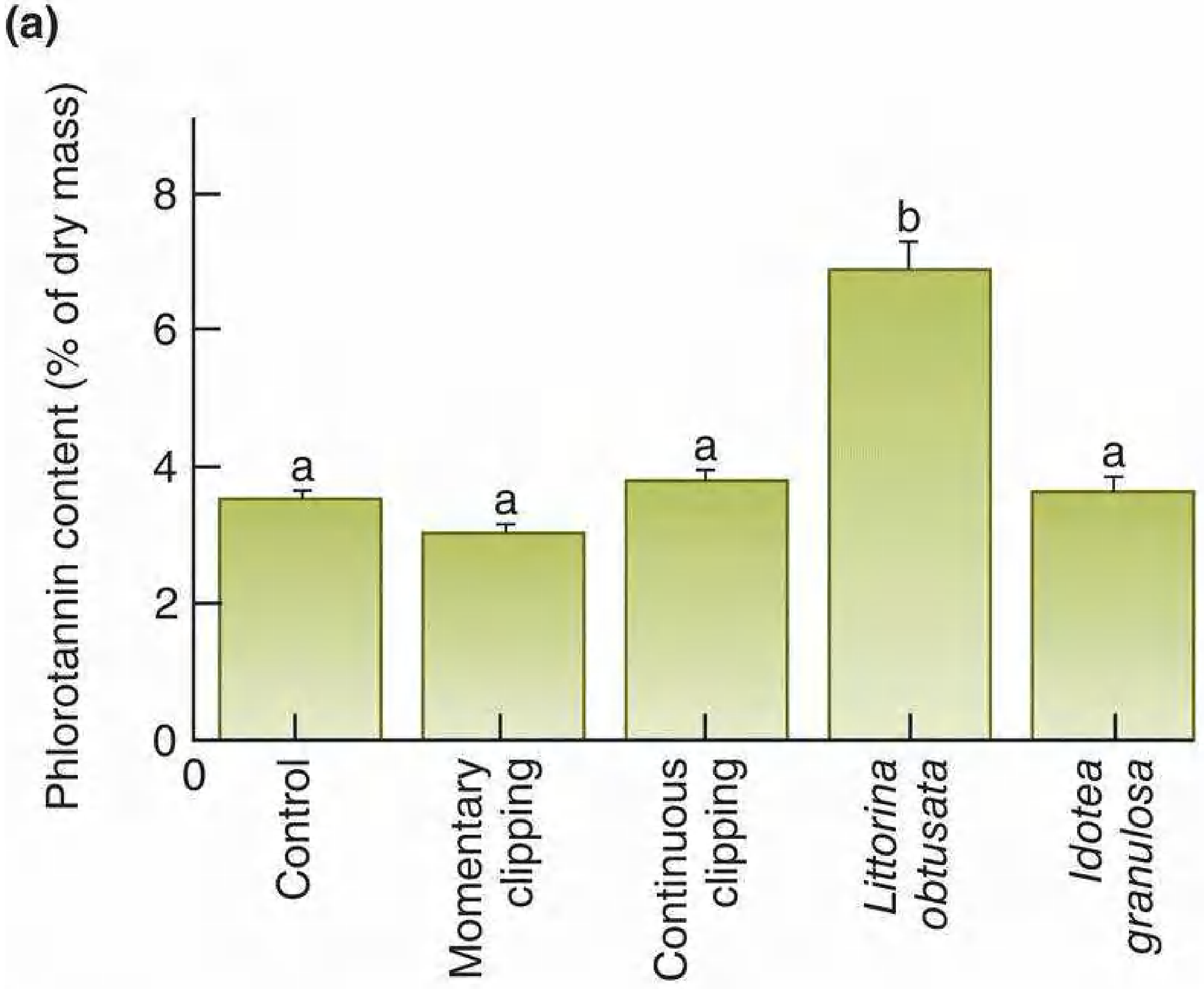
[Figure 9.11](#) Root and shoot herbivory inducing different patterns of root and shoot production of defensive chemicals. The effects of above-ground herbivory (the caterpillars *Manduca sexta* and *Trichoplusia ni*) and below-ground herbivory (the root-knot nematode, *Meloidogyne incognita*) on the production of defensive chemicals, as indicated, in (a) the leaves and (b) the roots of tobacco plants, *Nicotiana tabacum*. Bars are SEs. Nicotine and the non-nicotine alkaloids are synthesised in the roots. The phenolics and glycosides are synthesised in the leaves.

Source: After Kaplan *et al.* (2008).

but are herbivores really adversely affected? ...

Moreover, studies of these responses of individual plants to herbivore attack leave unanswered some important questions: do induced chemicals have an ecologically significant negative effect on the herbivores that induce them?; and do they have a measurable, positive impact on the plant making them? In an early review, Fowler and Lawton (1985) found little clear-cut evidence that rapidly inducible plant defences are effective against insect herbivores, despite a widespread belief that they were. More recently, too, Moore *et al.* (2013) noted that, because of their diversity and number, many modes of action remain undiscovered and their adaptive advantages await rigorous demonstration. However, there are now plenty of cases in which the plant's responses do seem to be genuinely harmful to the herbivores. When larch trees were defoliated by the larch budmoth, *Zeiraphera diniana*, the survival and adult fecundity of the moths were reduced throughout the succeeding 4–5 years as a combined result of delayed leaf production, tougher leaves, higher fibre and resin concentration and lower nitrogen levels (Baltensweiler *et al.*, 1977). Another common response to leaf damage is early abscission ('dropping off') of mined leaves. In the case of the leaf-mining insect *Phyllonorycter* spp. on willow trees (*Salix lasiolepis*), for example, early abscission of mined leaves was an important mortality factor for the moths (Preszler & Price, 1993).

As a further example, a few weeks of grazing on the brown seaweed *Ascophyllum nodosum* by snails (*Littorina obtusata*) induces substantially increased concentrations of phlorotannins (Figure 9.12a), which reduce snail grazing subsequently (Figure 9.12b). In this case, simple clipping of the plants did not have the same effect as the herbivore. Indeed, grazing by another herbivore, the isopod *Idotea granulosa*, also failed to induce the chemical defence. The reason seems to be that the snails can stay and feed on the same plant for long time periods, so that induced responses that take time to develop can still be effective. The isopods, by contrast, are much more mobile and are likely to have moved on by the time the production of any chemicals has been induced. In terrestrial environments, the induction of silicon defences in grasses shows similar patterns to this. Simple clipping is generally much less effective than genuine herbivory, and repeated damage and damage exceeding a threshold level seem to be necessary for induction to occur at all (Hartley & DeGabriel, 2016).


(a)
Phlorotannin content (% of dry mass)
0
2
4
6
8
0
a
a
a
b
a
Control
Momentary clipping
Continuous clipping
Littorina obtusata
Idotea granulosa

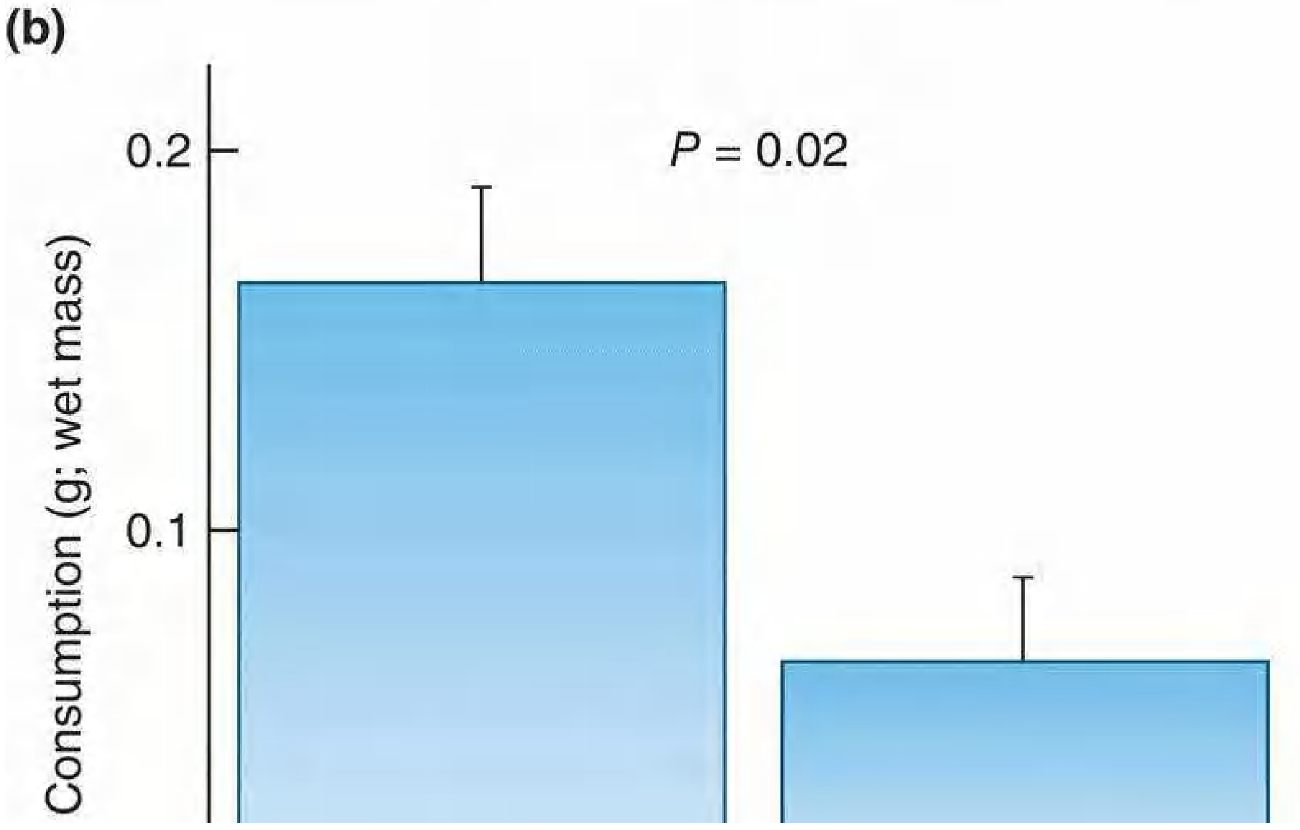

(b)
Consumption (g; wet mass)
0.1
0.2
P = 0.02

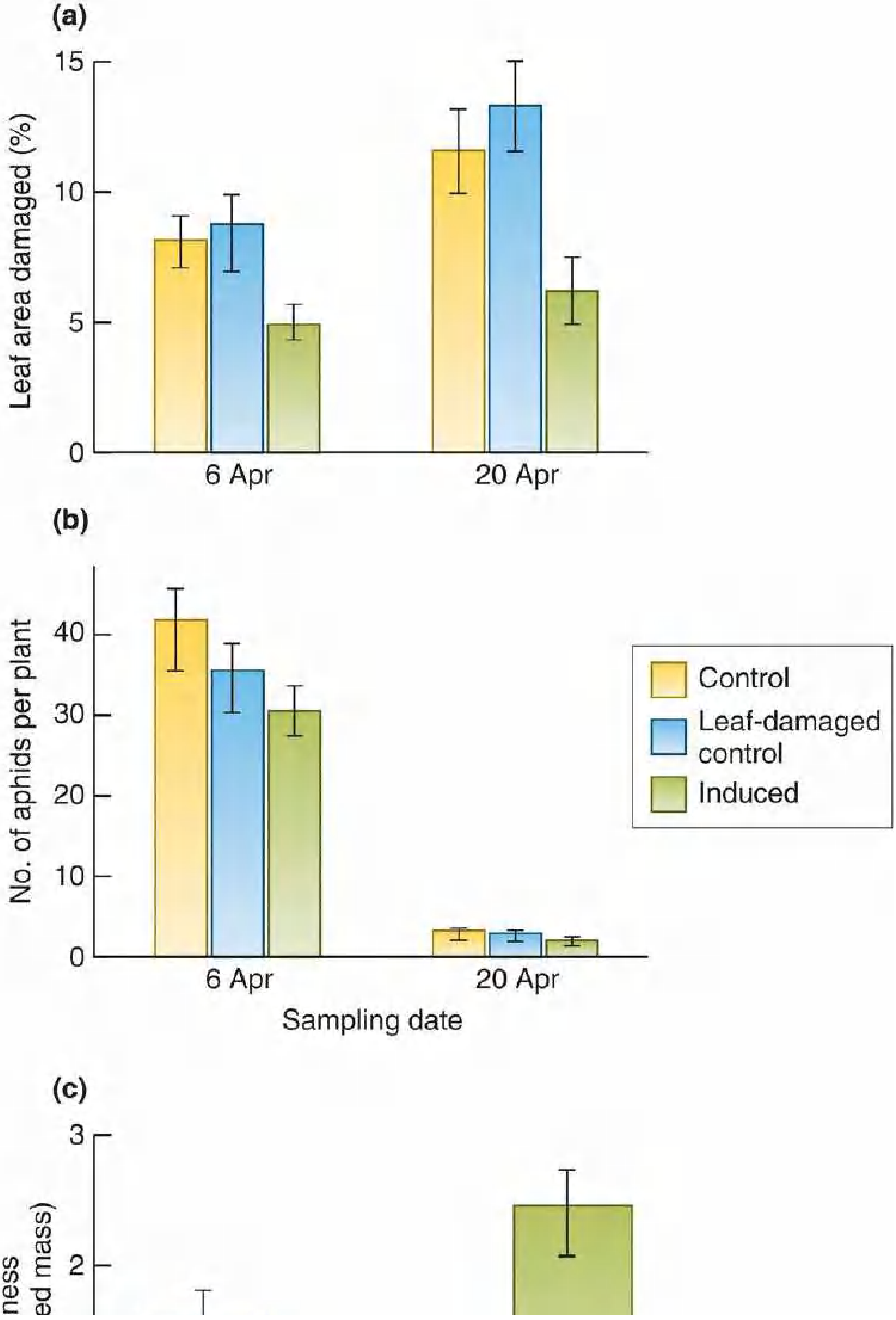
Figure 9.12 Snails induce a defensive response in seaweeds that protects them from snail grazing. (a) Phlorotannin content of *Ascophyllum nodosum* plants after exposure to simulated herbivory (removing tissue with a hole punch) or grazing by real herbivores of two species. Means and SE are shown. Only the snail *Littorina obtusata* had the effect of inducing increased concentrations of the defensive chemical in the seaweed. Different letters indicate that means are significantly different ($P < 0.05$). (b) In a subsequent experiment, the snails were presented with algal shoots from the control and snail-grazed treatments in (a); the snails ate significantly less of plants with a high phlorotannin content.

Source: After Pavia & Toth (2000).

Remember finally that we saw in Section 8.3.3 that allelochemicals (acting against competitors rather than herbivores) can also be induced in plants, and indeed that in the case discussed there – the seaweed *Galaxaura filimentosa* competing with the coral *Porites cylindrica* but being preyed upon by surgeonfish – the induction of the allelochemicals was accompanied by a reduction in investment in antiherbivore defences and hence an increase in herbivory.

... and do plants really benefit?

The second question – ‘do plants benefit from their induced defensive responses?’ – has proved even more difficult to answer, particularly because suggested benefits need to have taken account of the costs to the plants of making the responses, discussed above (Strauss et al., 2002), and relatively few well-designed field studies have been performed (Karban *et al*., 1999). Agrawal (1998) assigned wild radish plants (*Raphanus sativus*) to one of three treatments: grazed by caterpillars of *Pieris rapae*, leaf-damaged controls (equivalent amount of biomass removed using scissors) and overall controls (undamaged). Damage-induced responses included increased concentrations of defensive glucosinolates and increased densities of trichomes (hair-like structures). He then compared the effects of other herbivores on the three treatments. Earwigs (*Forficula* spp.) and other chewing herbivores caused 100% more leaf damage on the overall control and leaf-damaged control plants than on grazed plants, and there were 30% more sucking aphids (*Myzus persicae*) on both types of control plants (Figure 9.13a, b). Resistance induced by *P. rapae* grazing significantly increased a lifetime index of fitness of the radish plants (number of seeds multiplied by seed mass) by more than 60% compared with the overall control plants. However, leaf-damaged control plants had 38% lower fitness than the overall controls, indicating the negative effect of tissue loss without the benefits of induction (Figure 9.13c).


(a)
Leaf area damaged (%)
15
10
5
0
6 Apr
20 Apr
(b)
No. of aphids per plant
40
30
20
10
0
6 Apr
20 Apr
Sampling date
Control
Leaf-damaged control
Induced
(c)
3
2
ness
ed mass)

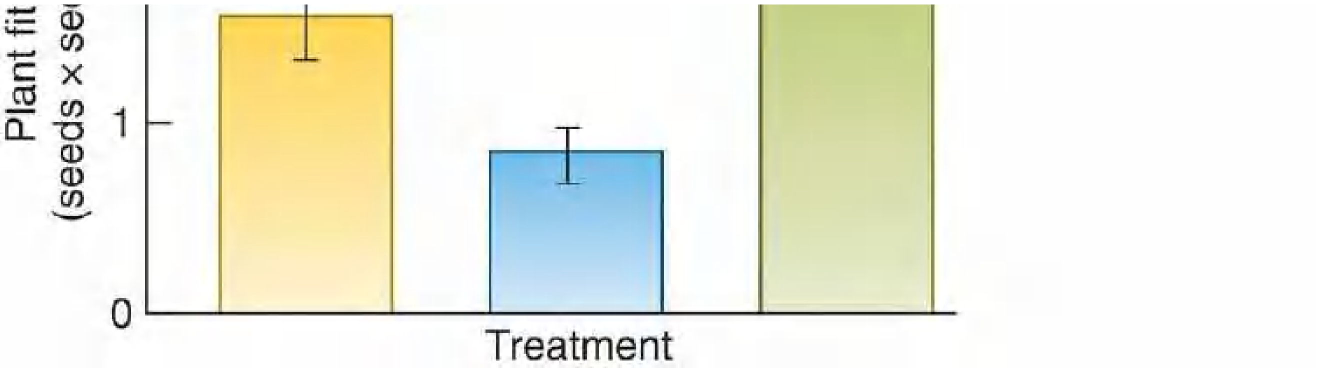


Figure 9.13 Responses to herbivory (but not simulated herbivory) reduced subsequent attacks by other herbivores, ultimately increasing fitness in the presence of herbivores. (a) Percentage of leaf area consumed by chewing herbivores and (b) number of aphids per plant, measured on two dates (6 April and 20 April) in three field treatments: overall control, leaf-damaged control (tissue removed by scissors) and induced (caused by grazing of caterpillars of *Pieris rapae*). (c) The fitness of plants in the three treatments calculated by multiplying the number of seeds produced by the mean seed mass (in mg). Bars are 95% CIs.

Source: After Agrawal (1998).

However, this fitness benefit to wild radish occurred only in environments containing herbivores. In their absence, an induced defensive response brought no benefit and the plants suffered reduced fitness (Karban *et al*., 1999). A similar result has been shown in a study of caterpillars of *Manduca sexta*, which are specialist consumers of wild tobacco (*Nicotiana attenuata*). The caterpillars not only induce an accumulation of secondary metabolites and proteinase inhibitors when they feed, they also induce the plants to release volatile organic compounds that attract the generalist predatory bug *Geocoris pallens*, which feeds on the slow-moving caterpillars (Kessler & Baldwin, 2004). In a laboratory experiment, naturally occurring genotypes that lacked the ability to produce proteinase inhibitors were damaged more, and sustained greater caterpillar growth, than inhibitor-producing genotypes (Glawe *et al*., 2003). In the absence of herbivory, however, plant genotypes that produced little or no proteinase inhibitor grew faster and taller and produced more seed capsules than inhibitor-producing genotypes (Zavala *et al*., 2004).

It is clear from the radish and tobacco examples that the evolution of inducible (plastic) responses can bring significant costs to the plant. Inducible responses are likely to be favoured by selection, therefore, only when past herbivory is a reliable predictor of future risk *and* if the likelihood of herbivory is not constant: constant herbivory should select for a fixed defensive phenotype that is best for that set of conditions (Karban *et al*., 1999). Of course, it is not only the costs of inducible defences that can be set against fitness benefits. Constitutive defences, such as spines, trichomes or defensive chemicals, also have costs, measured in phenotypes or genotypes lacking the defence, that can be seen in reductions in growth or the production of flowers, fruits or seeds (see the review by Strauss *et al*., 2002).

APPLICATION 9.2 Using maize ‘landraces’ to improve inducible indirect defence

The breeding of crop plants from their natural progenitors, and their continued ‘improvement’, has most often been focused on improvements in yield. However, natural defences against herbivores can be a casualty of this process, and the loss of such defences is a growing concern at a time when attempts are being made to reduce the use of (unnatural) chemical pesticides (see Chapter 15). For example, teosintes are the wild grasses generally considered to be one of the parent plants of modern maize, *Zea mays*. In most species of teosinte, eggs laid by stem-borers induce the production of volatile chemicals that attract parasitoid wasps that will attack larvae of the stem-borers once they hatch (Tamiru *et al.*, 2015). However, a study of the responses of commercially available hybrid and inbred lines of maize found that when eggs of the stem-borer, *Chilo partellus*, were laid on them and the response of the parasitoid, *Cotesia sesamiae*, noted, an overwhelming majority of the lines lacked this trait (Figure 9.14). On the other hand, 13 out of 25 landraces of maize tested did make this inducible response (Figure 9.14). (Landraces are domesticated but locally adapted varieties, usually associated with traditional farming systems.) Interbreeding between these landraces and high-yield hybrids therefore offers the prospect of improved crop defence for maize and a reduced need for spraying potentially toxic chemicals (Tamiru *et al.*, 2015).

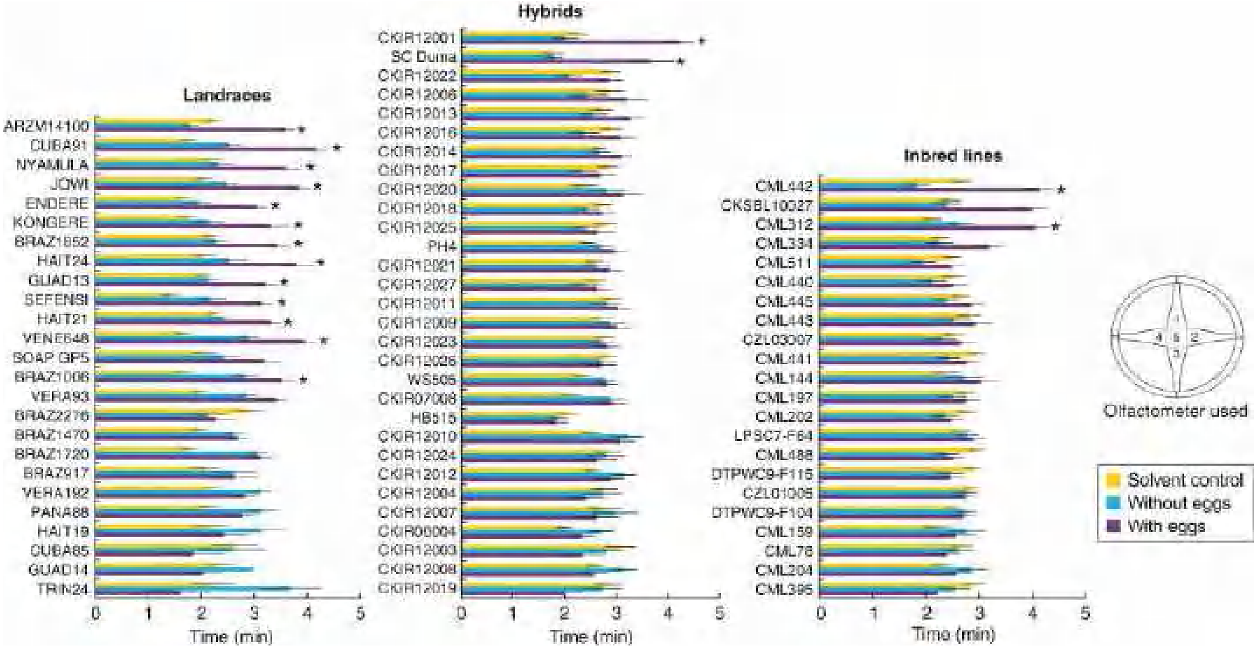


Figure 9.14 Landraces of maize frequently make an inducible defence response to stem-borer attack largely lacking in commercial varieties. The response of the parasitoid, *Cotesia sesamiae*, to odours from maize plants, *Zea mays*, with and without eggs of the stem-borer, *Chilo partellus*, and ‘solvent’ controls. A range of landraces, hybrids and inbred maize lines were used, as indicated. Adult parasitoids were placed at the centre (‘5’) of an olfactometer (bottom-right) with volatiles solvent-extracted from plants with eggs in limb ‘1’, volatiles from plants without eggs in limb ‘3’, and solvent only in limbs ‘2’ and ‘4’. Bars are SEs. Lines where a significantly longer time was spent in limb ‘1’ are indicated by *.

Source: After Tamiru *et al.* (2015).

9.3.4 Defending what’s most valuable

The third optimal defence hypothesis predicts that the more important an organ or tissue is for an organism's fitness, the better protected it will be. One aspect of this is that more important plant parts should be protected by constitutive chemicals produced all the time, whereas less important parts should rely on inducible chemicals, only produced in response to damage itself, and hence with far lower fixed costs to the plants. This is confirmed, for example, by another study of wild radish, *Raphanus sativus*, in which plants were either subjected to herbivory by caterpillars of the butterfly, *Pieris rapae*, or left as unmanipulated controls (Strauss *et al.*, 2004). The plant is insect-pollinated, and so petals, and all parts of the flower, are highly important to fitness. Matching this, concentrations of protective glucosinolates in petals were high (twice as high as in undamaged leaves) and were maintained constitutively, irrespective of whether the petals were damaged by the caterpillars (Figure 9.15). Leaves, on the other hand, have a much less direct influence on fitness, and as already noted, constitutive levels of glucosinolates were low. But if the leaves were damaged, the (induced) concentrations were even higher than in the petals.

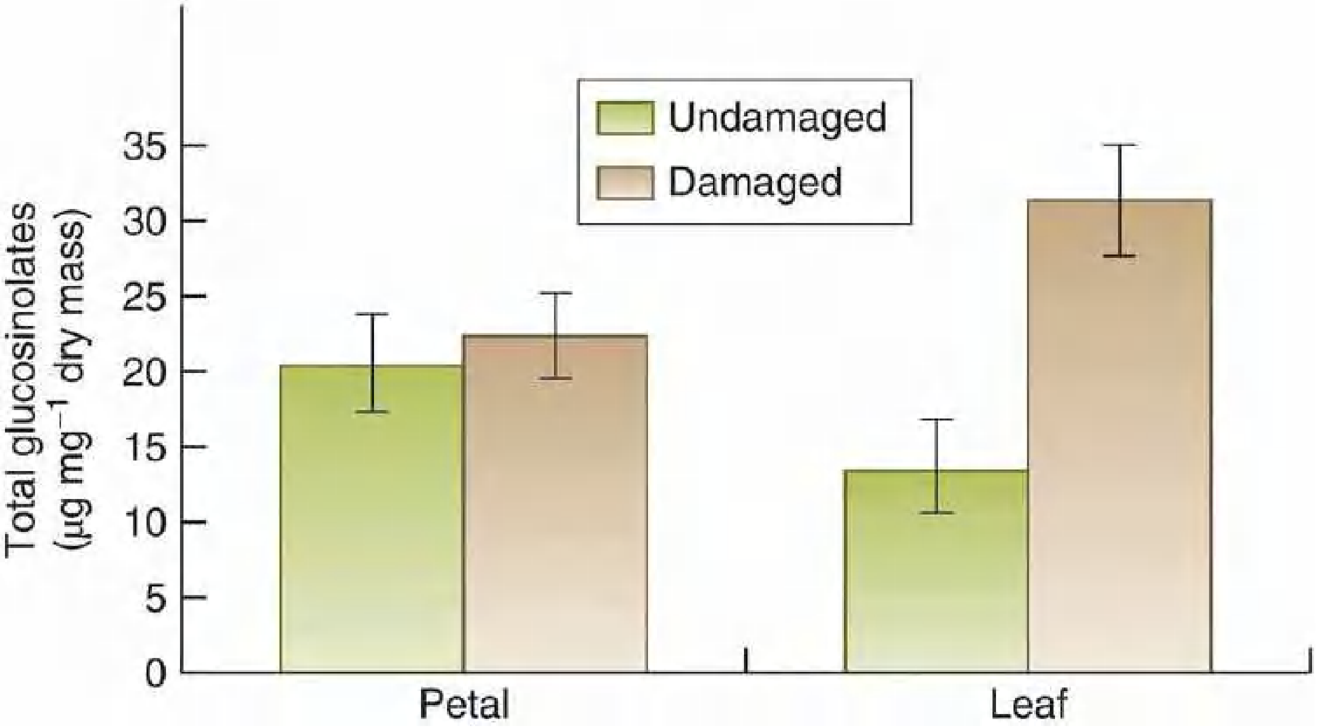


Figure 9.15 Constitutive levels of defence were high in valuable wild radish petals. Concentrations of glucosinolates in the petals and leaves of wild radish, *Raphanus sativus*, either undamaged or damaged by caterpillars of *Pieris rapae*. Bars are SE.

Source: After Strauss *et al.* (2004).

Similar results were found for the brown seaweed, *Sargassum filipendula*, where the holdfast at its base is the most valuable tissue, since without it, the plant would be cast adrift in the water (Taylor *et al.*, 2002). This was protected by costly constitutive, quantitative chemicals. The much less valuable youngest stipes (effectively stems) near the tip of the plant were protected only by toxic chemicals induced by grazing.

9.3.5 Defence when times are hard

resource-allocation based hypotheses of plant defence

We have noted several times that the benefits of both constitutive and induced defences need to be set against the costs of producing them. The fourth optimal defence hypothesis takes this further by proposing that investments in defence will themselves be modulated by the availability of resources and hence the intensity of competing demands for those resources. This hypothesis is in turn related to a number of others that have often been discussed outside the optimal defence framework (Stamp, 2003 ; see also Hartley & Jones, 1997), but can equally be seen as variants of it. Of these, we briefly examine three. Firstly, the *carbon : nutrient balance hypothesis* (Bryant *et al*., 1983) proposes that plants make their allocations of available resources to defensive structures and chemicals as a direct (phenotypic) response to whether they have excess carbon (for example in poor soils, where carbon (non-nitrogenous) based defences should be increased and nitrogenous defences reduced), or excess nutrients, especially nitrogen (for example in shade, or following fertilisation, where carbon-based defences should be reduced and nitrogenous defences increased). Secondly, the *resource availability* or *growth rate hypothesis* (Coley *et al*., 1985) proposes that for plants adapted to resource-rich environments, the intense competition there favours rapid growth, and plants, because of their allocation of resources to growth, have relatively little to allocate to defence; whereas for those adapted to resource-poor environments, plant growth is inevitably slow, and plants, predictably vulnerable to attack for an extended period, invest a relatively large amount in defence.

Finally, the *growth-differentiation balance hypothesis* (Herms & Mattson, 1992) proposes that limited carbon is available for non-nitrogenous based defences when nitrogen availability and growth rates are high (since the carbon is needed for growth), but also when nitrogen availability and growth rates are low, because net assimilation rates are low and production rates of all compounds are limited. However, production of these defence chemicals should be high at intermediate levels of nitrogen availability, because net assimilation rates are high enough to make excess carbon available for defence. (‘Differentiation’ in the hypothesis title refers, confusingly, to anything that is not growth.) This third hypothesis, therefore, incorporates important elements of the other two, associating rapid growth with low defence investments as in the growth rate hypothesis, but acknowledging the broader importance of the balance between carbon and plant nutrients as in the carbon : nutrient balance hypothesis.

We see in Figure 9.16 an example in which nitrogen-free flavonoid defences were produced at maximum rates in marula (*Sclerocarya birrea*) orchards in South Africa at intermediate rates of fertiliser application, in support of the growth-differentiation balance hypothesis, and in partial support of the other two. And in the case of carbon-rich extrafloral nectar production in passionfruit, discussed in Section 9.3.3 (Figure 9.9), rates of production were depressed by low-red : high-red light ratios characteristic of shade, as predicted by all three hypotheses. More generally, however, meta-analyses and other reviews of evidence paint a picture for all three hypotheses of encouraging but by no means universal support (Stamp, 2003). Overall, therefore, it is clear that many of the drivers of variation in plant defence have been identified, but none of them tells the whole story.

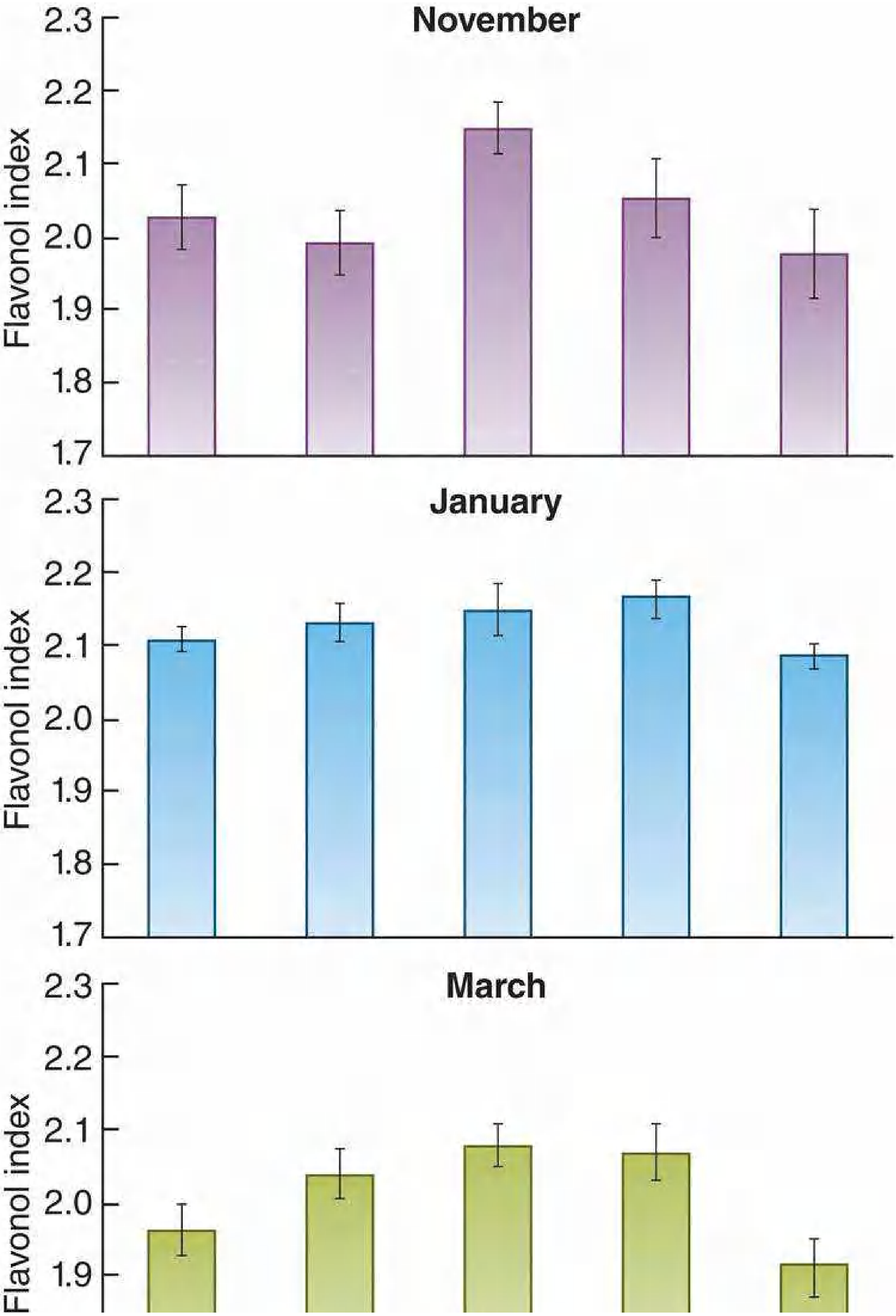

November
Flavonol index
2.3
2.2
2.1
2.0
1.9
1.8
1.7
January
Flavonol index
2.3
2.2
2.1
2.0
1.9
1.8
1.7
March
Flavonol index
2.3
2.2
2.1
2.0
1.9

Figure 9.16 Maximum levels of defence at intermediate rates of fertiliser application. Production of antiherbivore flavinols in leaves marula saplings, *Sclerocarula birrea*, with fertiliser applications equivalent to 0–120 g N per m^2, as indicated, applied in September and November 2013 and January 2014 and production monitored two months later. The flavinol index is a measure of investment in non-nitrogenous flavonoid defences determined by the ratio of fluorescence of plant material under red and UV light; it was highest at intermediate fertilisation rates ($P = 0.059$, $P = 0.091$ and $P = 0.009$, respectively). Bars are SEs.

Source: After Scogings (2018).

APPLICATION 9.3 Reducing stress to resist pine bark beetle attack

The general proposition that plant ‘stress’ (a shortage of necessary resources) may compromise their ability to mount defensive responses against herbivore attack is especially relevant in the context of plant nurseries, where each individual seedling is an investment whose value needs to be protected. The interaction between water shortage (‘drought stress’) and herbivore attack was studied for pine seedlings, *Pinus radiata*, in New Zealand being attacked by the invasive pine bark beetle, *Hylastes ater*, which can cause seedling mortality rates as high as 30% (Sopow *et al*., 2015). In a 2 × 2 factorial design (with and without drought stress prior to being planted out; and with and without insecticide treatment), the beetles themselves showed a preference for the healthier seedlings, not previously deprived of water (Figure 9.17a). However, the stressed seedlings were twice as likely to suffer sustained attack, resulting in ‘girdling’, where bark is removed from the whole circumference of the stem, markedly reducing the plants’ rate of growth during the course of the experiment and no doubt threatening their survival subsequently. This more favourable outcome for the control plants, despite their being preferred by the beetles, seems clearly to have been a result of their defensive responses: much greater production rates of defensive resins (Figure 9.17b). As the authors note, nursery managers should focus on providing a stress-free transition from the nursery to the field to allow the seedlings in their charge to maximise their chances of survival by producing sufficient quantities of defensive chemicals.

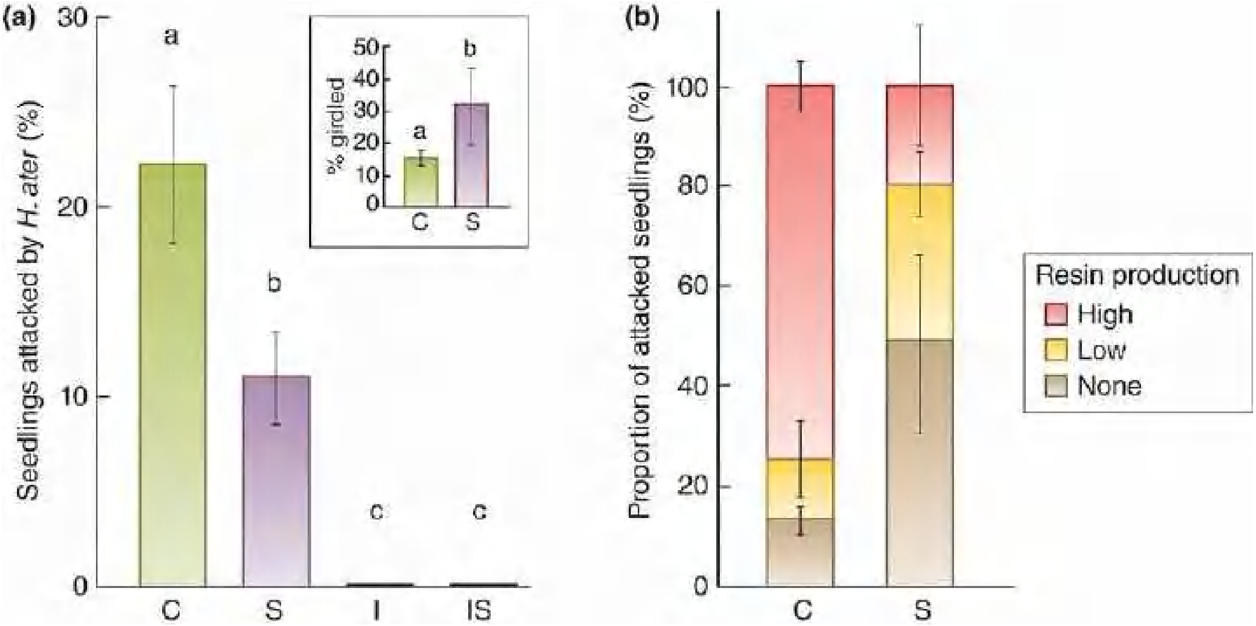


Figure 9.17 Bark beetles prefer non-stressed pine seedlings but the stressed seedlings suffer more from beetle attack. (a) The percentage of pine seedlings, *Pinus radiata*, attacked by the bark beetle, *Hylastes ater*, and (insert) the percentage of these that are girdled by the attacks (bark removed from the whole circumference of a stem). C, control; S, drought stressed; I, treated with insecticide; IS, stress + insecticide. Bars are SE. Means with different letters above are significantly different ($P < 0.05$). (b) Defensive resin production by these seedlings, classified as high, low or none. The difference between control and stressed plants was highly significant ($P < 0.001$).

Source: After Sopow *et al* (2015).

APPLICATION 9.4 Elevated CO_2 and plant defence in sugar cane

The issue of the effect on plant defence of the balance of availability of carbon and nitrogen is likely to become increasingly important as levels of atmospheric CO_2 continue to rise – to say nothing of the effect of this as a greenhouse gas (see Section 22.2). We noted before how complex the effects of these shifting balances can be, and in the case of elevated atmospheric CO_2, there are reports of suppression of the jasmonate pathway, and hence of the inducibility of defences, but also of increased ratios of carbon to nitrogen in plant tissues and hence increased consumption rates by herbivores as they seek to compensate for the difficulties of obtaining sufficient nitrogen (see Frew *et al.*, 2017). Sugar cane (*Saccharum* spp. hybrid) is a key crop in many parts of the world, including Australia, where the cane beetle, *Dermolepida albohirtum*, attacks the roots and causes an estimated AU$40 million loss in production per year (Frew *et al.*, 2017). An important weapon in the armoury of sugarcane defence, and that of many other grasses as noted above (Hartley & DeGabriel, 2016), is silicon, typically deposited as small, essentially glass, ‘phytoliths’. Figure 9.18 shows results when sugarcane in Australia was grown with and without elevated levels of CO_2, with and without a silicate supplement added to the soil, and in the presence and absence of cane beetles.

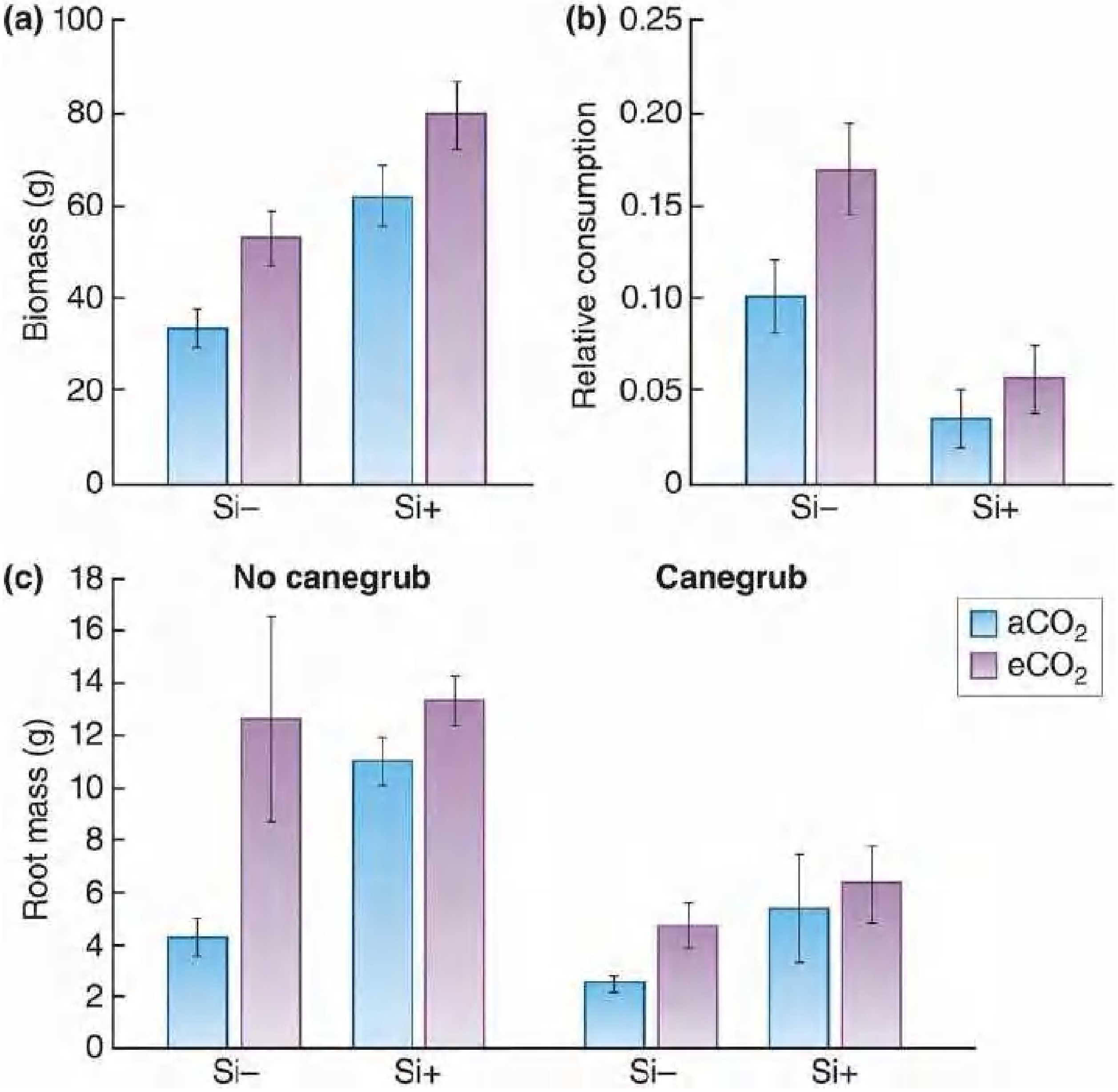


Figure 9.18 Silicate supplementation can help ameliorate the harmful effects of elevated CO_2on beetle damage of sugar cane. (a) Effects of silicate supplementation to the soil (Si+) and elevated atmospheric CO_2 (eCO_2 compared with aCO_2) on the biomass of sugar cane plants, *Saccharum* spp. hybrid, grown for 18 weeks. The effects of both were significant ($P < 0.001$ and $P < 0.01$, respectively). (b) The effects of these treatments on root consumption by the cane beetle, *Dermolepida albohirtum*. The effect of silicate was significant ($P < 0.01$), the effect of CO_2 nearly so ($P < 0.1$). (c) The effects of these treatments, and of canegrub herbivory, on root biomass. The effects of all three were significant ($P < 0.05$, $P < 0.01$ and $P < 0.001$, respectively) such that silicate could ameliorate the otherwise very harmful effects of elevated CO_2 on plant loss to herbivory. Bars are SEs.

Source: After Frew *et al.* (2017).

Elevated CO_2 itself increased the production rates of the sugarcane (as did supplementary silicate; Figure 9.18a). However, in the absence of supplementary silicate, this increased the consumption rates when cane beetles fed on the plants (Figure 9.18b), leading to no net change in root mass overall (Figure 9.18c). This, though, was reversed when the silicon supplement was added: consumption rates went down (Figure 9.18b) and root mass overall was raised by the elevated CO_2 (Figure 9.18c). Thus, anticipated rises in CO_2 levels could lead to pest-induced reductions in crop productivity as a result of compromised defences. But an understanding of the defence mechanisms, and targeted addition of silicate (see Debona *et al*., 2017) and perhaps other supplements could at least mitigate some of these negative effects.

9.4 Effects of herbivory and plants’ tolerance of those effects

tolerance and plant compensation

Despite a plethora of defensive structures and chemicals, herbivores still eat plants. Plant compensation refers to the degree to which any tolerance exhibited by plants is effective. If damaged plants, despite some tolerance, have lower fitness than their undamaged counterparts then they have undercompensated for herbivory; if they have higher fitness then they have overcompensated (Strauss & Agrawal, 1999).

9.4.1 Herbivory, defoliation and plant growth

Individual plants can compensate for the effects of herbivory in a variety of ways. The removal of shaded leaves (with their normal rates of respiration but low rates of photosynthesis; see Chapter 3) may improve the balance between photosynthesis and respiration in the plant as a whole. Or, in the immediate aftermath of an attack from a herbivore, plants may compensate by utilising stored reserves or by altering the distribution of photosynthate within the plant. Herbivore damage may also lead to an increase in the rate of photosynthesis per unit area of surviving leaf. Often, there is compensatory regrowth of defoliated plants when buds that would otherwise remain dormant are stimulated to develop. There is also, commonly, a reduced death rate of surviving plant parts.

APPLICATION 9.5 Invasion of a tolerant seaweed

An example is shown in Figure 9.19. *Caulerpa cylindracea*, originally from Western Australia, is one of the most successful seaweed species in invading Mediterranean shores, despite having been found there for the first time in the 1990s, and in the face of heavy attack from native herbivores. The data in the figure help us understand why and may help ultimately in at least slowing the progress of this invasion. Subjected to moderate or even intense simulated herbivory (up to 75% of its cover removed), compensatory regrowth was able to restore original coverage levels in all treatments in just two months.

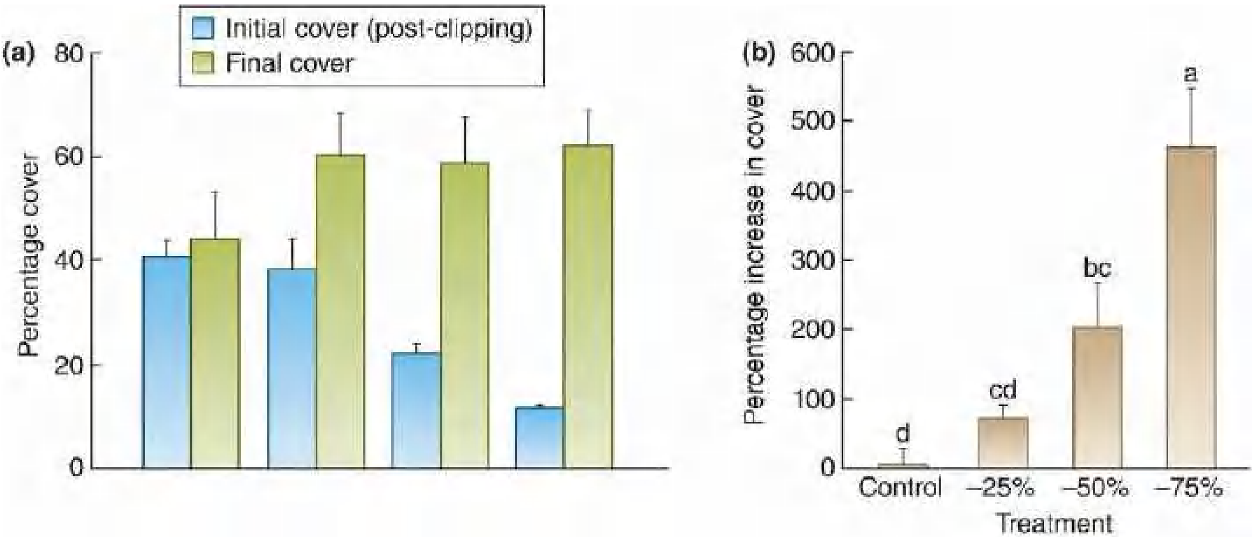


Figure 9.19 Rapid compensatory regrowth of an invasive seaweed following herbivory. (a) The initial cover of the seaweed *Caulerpa cylindracea* after having been artificially clipped to the level indicated, and the cover in the same treatments 57 days later. (b) The corresponding percentage increases in cover over the period. Bars are SEs. Different letters above bars indicate treatments that were significantly different from one another.

Source: After Bulleri & Malquori (2015).

herbivory and competition, and the particular tolerance of grasses

The plants that seem most tolerant of grazing, especially vertebrate grazing, are the grasses. In most species, the meristem is almost at ground level amongst the basal leaf sheaths, and this major point of growth (and regrowth) is therefore usually protected from grazers' bites. Following defoliation, new leaves are produced using either stored carbohydrates or the photosynthate of surviving leaves, and new tillers are also often produced. Grasses do not benefit directly from their grazers' attentions. But it is likely that they are helped by grazers in their competitive interactions with other plants (which are more strongly affected by the grazers), accounting for the predominance of grasses in so many natural habitats that suffer intense vertebrate grazing. This is an example of the most widespread reason for herbivory having a more drastic effect on grazing-intolerant species than is initially apparent – the interaction between herbivory and plant competition (the range of possible consequences has been discussed by Pacala and Crawley, 1992).

9.4.2 Herbivory and plant survival

mortality: the result of an interaction with another factor?

Indeed, generally it is more usual for herbivores to increase a plant's susceptibility to mortality than to kill it outright, with competition from other species being a particularly common second factor in promoting a plant's ultimate demise. Two species of perennial herbaceous plants native to the eastern USA, *Solanum carolinense* and *Solidago altissima*, were grown together at a range of densities both in the presence and absence of insect herbivores. Figure 9.20 was then constructed by combining the results of the experiments into a model of the dynamics of the two species, and using the model as a basis for a series of simulations of what the outcome would be if competition between the species was allowed to run on for 50 years. *S. altissima* was the superior competitor, but it was only in combination with herbivory that its superiority, and the suppression of *S. carolinense*, was substantially expressed. The study is noteworthy in that the effects of herbivory were determined separately for the growth rates of both species, for their intensities of self-regulation (intraspecific competition), and for the intensity of interspecific competition between them. Herbivory affected all three.

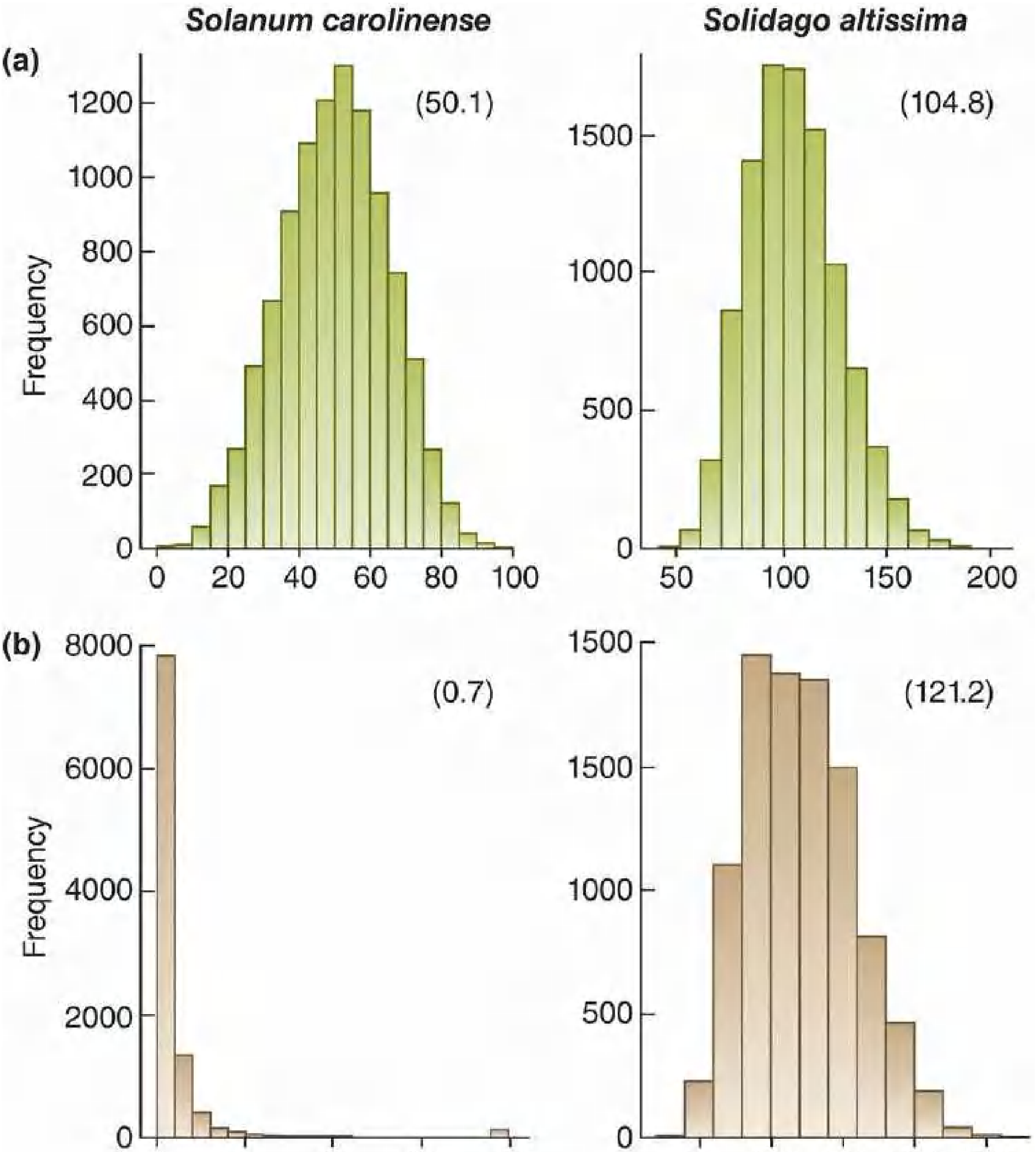


Figure 9.20 Herbivory affects the outcome of competition between two plant species. The frequency distribution of outcomes for *Solanum carolinense* and *Solidago altissima* when they are competed against one another, (a) with and (b) without herbivory, in simulation models (10 000 times) using parameters based on experiments in which their competitive strengths and the effects of herbivory were estimated. Numbers in parentheses are average densities. Note the varying scales on the horizontal axes.

Source: After Kim *et al*. (2013).

repeated defoliation or ring-barking can kill

Repeated defoliation can have an especially drastic effect. In a glasshouse experiment in which eight grassland species were clipped between zero (control) and eight times (at intensities of 25%, 50% or 100% of leaf area), clipping once or twice had very little effect on plant survival, but clipping four and especially eight times had a much more profound effect (Figure 9.21a). This non-linearity was, if anything, even more marked in the effects of defoliation on the growth of the surviving plants (Figure 9.21b).

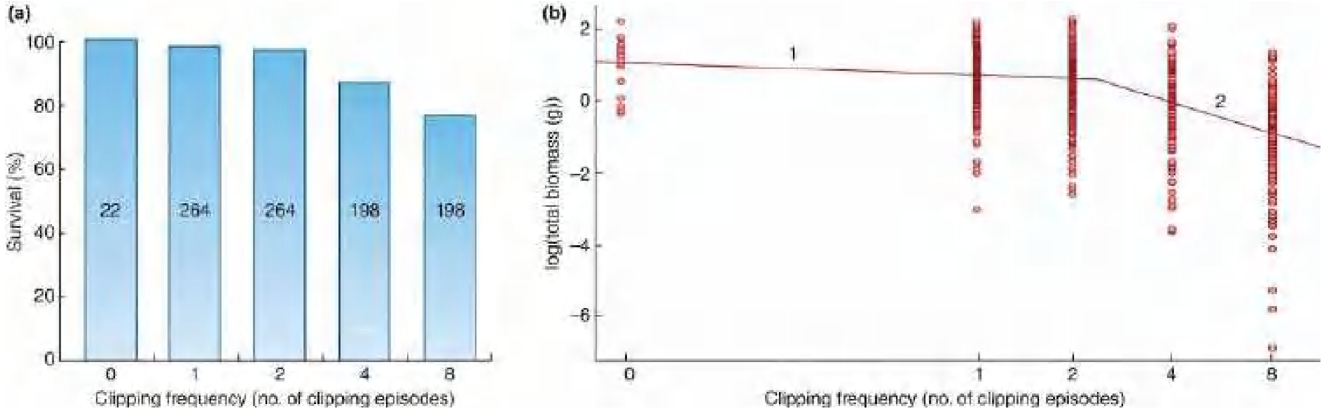


Figure 9.21 **Drastic effects of repeated defoliation.** The effects of the frequency of artificial herbivory (clipping) on a set of eight grassland species (*Senecio jacobaea*, *Trifolium repens*, *Rumex acetosella*, *Holcus lanatus*, *Vicia sativa*, *Lathyrus pratensis*, *Festuca rubra* and *Achillea millefolium*) in a glasshouse experiment. (a) Average effects on survival (difference between species were non-significant). Numbers in bars represent sample sizes. (b) Overall effects on growth. In both cases, the effects of repeated defoliation (four and especially eight clippings) were disproportionately great. In the case of growth, piecewise regression was used to show the existence of a threshold for the increased effect between two and four clippings. (Line 1: $y = 0.76 - 0.04x$; Line 2: $y = 1.23 - 1.06x$.)

Source: After del-Val & Crawley (2005).

The mortality of established plants, however, is not necessarily associated with massive amounts of defoliation. Ring-barking of trees, for example, by squirrels or porcupines, can have a disproportionately profound effect. The cambial tissues and the phloem are torn away from the woody xylem, and the carbohydrate supply link between the leaves and the roots is broken. Thus, these pests of forestry plantations often kill young trees while removing very little tissue. Surface-feeding slugs can also do more damage to newly established grass populations than might be expected from the quantity of material they consume. The slugs chew through the young shoots at ground level, leaving the felled leaves uneaten on the soil surface but consuming the meristematic region at the base of shoots from which regrowth would occur. They therefore effectively destroy the plant.

Predation of seeds, not surprisingly, has a predictably harmful effect on individual plants (i.e. the seeds themselves). Azcárate and Peco (2006), for example, showed that ants foraging in grasslands in central Spain took disproportionate numbers of larger and heavier seeds, shifting the balance of species in the seed bank, and thus being instrumental, at least in part, in maintaining the predominance of small-seeded annuals in this and other Mediterranean grasslands.

Herbivores can also have severe effects on plants when they act as vectors for plant pathogens (bacteria, fungi and especially viruses) – what the herbivores take from the plant may be far less important than what they give it. For instance, scolytid beetles feeding on the growing twigs of elm trees act as vectors for the fungus that causes Dutch elm disease. This killed vast numbers of elms in north-eastern USA in the 1960s (see Figure 12.31), and virtually eradicated them in southern England in the 1970s and early 1980s.

APPLICATION 9.6 The effects of aphids and the viruses they carry in response to climate change

Looking forward, it is important to ask what effects rising temperatures and rising levels of CO_2 will have on key pest groups such as aphids, since they harm many crops not only directly but also through the viruses they carry. Any or all of the processes determining their pest status are likely to be affected by climate change (Canto *et al.*, 2009). A good example is the bird cherry-oat aphid, *Rhopalosiphum padi*, an important pest in its own right of wheat and other cereal crops, especially in northern Europe, but also the vector of several dangerous plant viruses, including barley yellow dwarf virus (BYDV). Figure 9.22a shows how rapidly the aphid's own rate of development responds to rising temperatures, while Figure 9.22b shows the drastic effect, over a similar temperature range, on the efficiency with which it transmits BYDV. Several other key processes in the wheat-aphid-virus system are equally profoundly affected by temperature, and probably, too, by elevated levels of CO_2 (Finlay & Luck, 2011). Global food production struggles already to meet the needs of the growing human population (Chapter 5, Application 5.2). Climate change, as well as its other harmful effects (Section 22.2), clearly has the potential to threaten further our already challenged food supply.

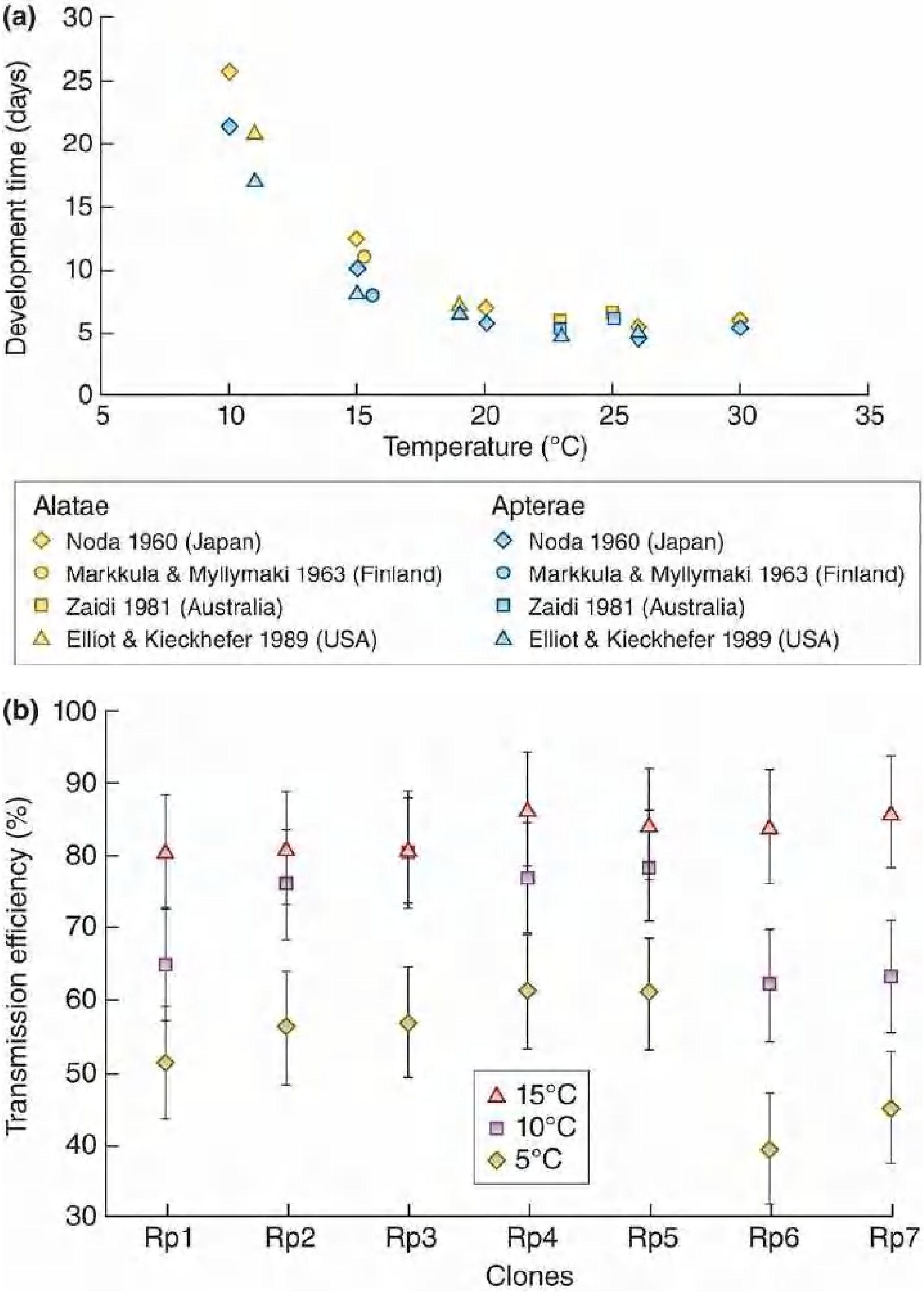

(a)
Development time (days)
0
5
10
15
20
25
30
Temperature (°C)
5
10
15
20
25
30
35
Alatae
◇ Noda 1960 (Japan)
○ Markkula & Myllymaki 1963 (Finland)
□ Zaidi 1981 (Australia)
△ Elliot & Kieckhefer 1989 (USA)
Apterae
◇ Noda 1960 (Japan)
○ Markkula & Myllymaki 1963 (Finland)
□ Zaidi 1981 (Australia)
△ Elliot & Kieckhefer 1989 (USA)
(b)
Transmission efficiency (%)
30
40
50
60
70
80
90
100
Clones
Rp1
Rp2
Rp3
Rp4
Rp5
Rp6
Rp7
△ 15°C
□ 10°C
◇ 5°C

Figure 9.22 An aphid pest's development is speeded up, and its efficiency in transmitting a harmful virus increased, by higher temperatures. (a) Data from a range of studies, as indicated (see the reference for details) on the effects of fixed temperatures on the development time of nymphs of the bird cherry-oat aphid, *Rhopalosiphum padi*. Alatae are winged forms; apterae are wingless. (b) Mean percentages of effective transmission of barley yellow dwarf virus, at three temperatures, as indicated, from infected to target wheat plants by *R. padi* infected as nymphs and transmitted after they had moulted into adults. Results are shown for seven clones of *R.* padi (Rp 1–7). Bars are 95% CIs.

Source: (a) After Finlay & Luck (2011). (b) After Smyrnioudis *et al.* (2001).

9.4.3 Herbivory and plant fecundity

herbivores affect plant fecundity both indirectly and directly

The effects of herbivory on plant fecundity are, to a considerable extent, reflections of the effects on plant growth: smaller plants bear fewer seeds. However, even when growth appears to be fully compensated, seed production may nevertheless be reduced because of a shift of resources from reproductive output to shoots and roots. This was the case for a study in which the biomass of the cruciferous plant *Brassica napus* was monitored in response to 0, 25 and 75% defoliation of seedling plants by three herbivore species with biting and chewing mouthparts – adult flea beetles *Phyllotreta cruciferae*, and larvae of the moths *Plutella xylostella* and *Mamestra configurata*. Compensation in growth was complete after 21 days (Figure 9.23a), but subsequent seed production was still significantly lower in the herbivore-damaged plants (Figure 9.23b).

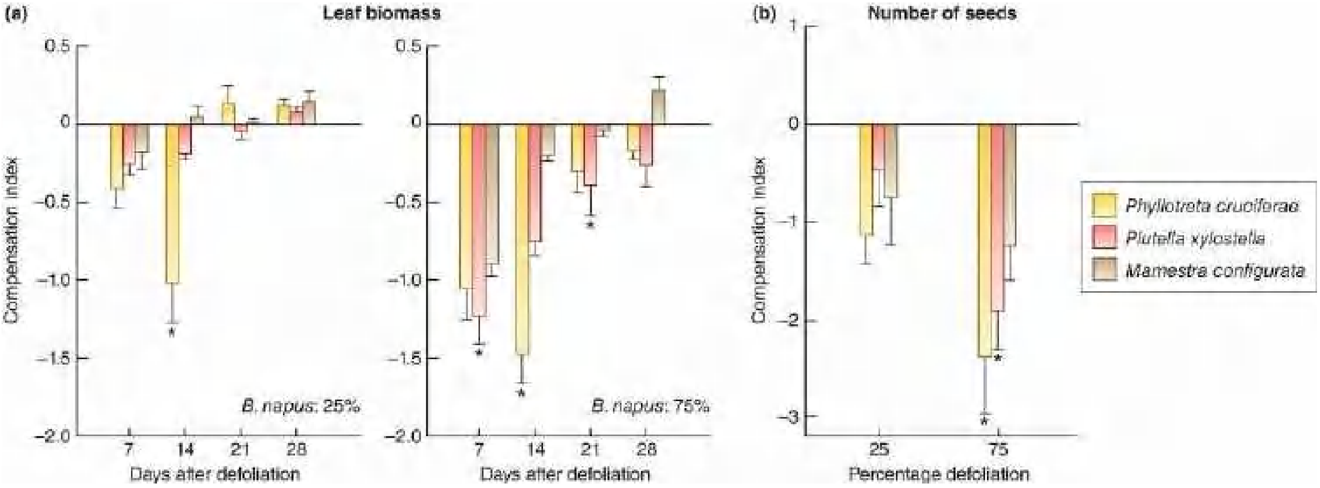


Figure 9.23 Plant fecundity can be affected by herbivory even when there is compensatory re-growth. (a) Compensation of leaf biomass (mean ± SE: ($\log_e$ biomass defoliated plant)–($\log_e$ of mean for control plants)) of *Brassica napus* seedlings with 25% or 75% defoliation by three species of insect (see key) in a controlled environment. On the vertical axis, zero equates to perfect compensation, negative values to undercompensation and positive values to overcompensation. Mean biomasses of defoliated plants that differ significantly from corresponding controls are indicated by an asterisk. (b) The effects of this same defoliation on the number of seeds set by the plants.

Source: After Gavloski & Lamb (2000).

Plants may also be affected more directly, by the removal or destruction of flowers, flower buds or seeds. Thus, caterpillars of the large blue butterfly *Maculinea rebeli* feed only in the flowers and on the fruits of the rare plant *Gentiana cruciata*, and the number of seeds per fruit (70 compared to 120) is reduced where this specialist herbivore occurs (Kery *et al*., 2001).

much pollen and fruit herbivory benefits the plant

It is important to remember, however, that many cases of 'herbivory' of reproductive tissues are actually mutualistic, benefitting both the herbivore and the plant (see Chapter 13). Animals that 'consume' pollen and nectar usually transfer pollen inadvertently from plant to plant in the process; and there are many fruit-eating animals that also confer a net benefit on both the parent plant and the individual seed within the fruit. Most vertebrate fruit-eaters, in particular, either eat the fruit but discard the seed, or eat the fruit but expel the seed in the faeces. This disperses the seed, rarely harms it and frequently enhances its ability to germinate. Insects that attack fruit, on the other hand, are very unlikely to have a beneficial effect on the plant. They do nothing to enhance dispersal, and they may even make the fruit less palatable to vertebrates.

Some large animals that normally kill seeds can also play a part in dispersing them, and they may therefore have at least a partially beneficial effect. There are some 'scatter-hoarding' species, like certain squirrels, that take nuts and bury them at scattered locations; and there are other 'seed-caching' species, like some mice and voles, that collect scattered seeds into a number of hidden caches. In both cases, although many seeds are eaten, the seeds are dispersed, they are hidden from other seed predators and a number are never relocated by the hoarder or cacher. Indeed, the plants appear to manipulate the behaviour of the scatter-hoarders in a variety of ways: stimulating hoarding by producing seed and nuts that are large and nutritious; discouraging immediate consumption by imposing handling costs on the hoarders; masting (producing vast crops in some years and little in others – see Section 10.2.4) so that the hoarders are frequently satiated; and producing seeds without strong odours so that the buried cache is not easily discovered (Vander Wall, 2010).

Herbivores also influence fecundity in a number of other ways. One of the most common responses to herbivore attack is a delay in flowering. For instance, in longer lived semelparous species, herbivory frequently delays flowering for one year or more, and this typically increases the longevity of such plants since death almost invariably follows their single burst of reproduction (see Chapter 4). *Poa annua* on a lawn can be made almost immortal by mowing it at weekly intervals, whereas in natural habitats, where it is allowed to flower, it is commonly an annual – as its name implies.

the timing of herbivory is critical

Generally, the timing of defoliation is critical in determining the effect on plant fecundity. If leaves are removed before inflorescences are formed, then the extent to which fecundity is depressed clearly depends on how much the plant is able to compensate. Early defoliation of a plant with sequential leaf production may have a negligible effect on fecundity; but where defoliation takes place later, or where leaf production is synchronous, flowering may be reduced or even inhibited completely. If leaves are removed after the inflorescence has been formed, the effect is usually to increase seed abortion or to reduce the size of individual seeds.

An example where timing is important is provided by field gentians (*Gentianella campestris*). When herbivory on this biennial plant is simulated by clipping to remove half its biomass (Figure 9.24a), the outcome depends on the timing of the clipping (Figure 9.24b). Fruit production was much increased over controls if clipping occurred between 1 and 20 July, but if clipping occurred later than this, fruit production was less in the clipped plants than in the unclipped controls. The period when the plants show compensation coincides with the time when damage by herbivores normally occurs.

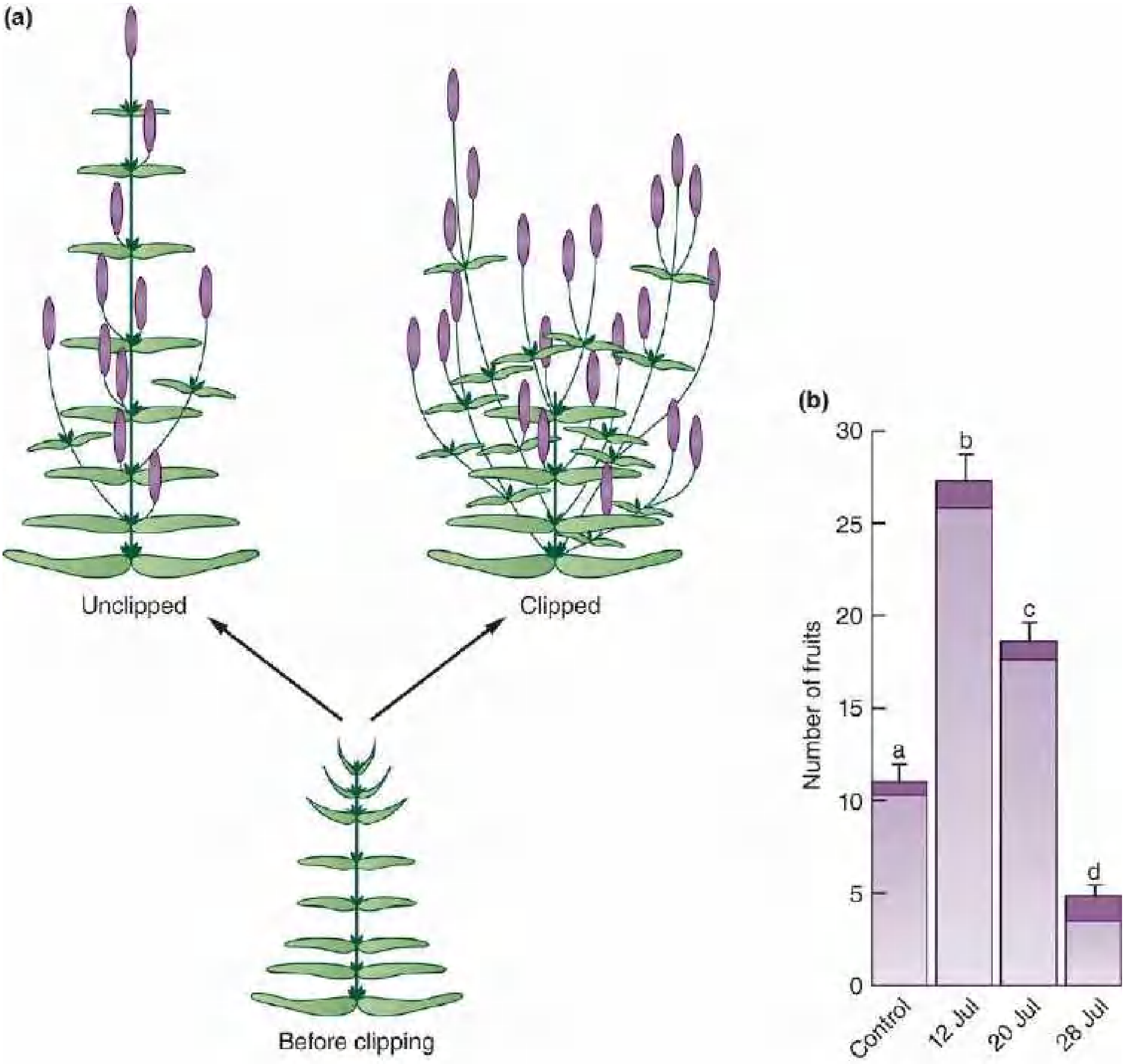


Figure 9.24 The importance of the timing of herbivory. (a) Clipping of field gentians to simulate herbivory causes changes in the architecture and numbers of flowers produced. (b) Production of mature (pale histograms) and immature fruits (dark histograms) of unclipped control plants and plants clipped on different occasions from 12 to 28 July 1992. Means and SEs are shown and all means are significantly different from each other ($P < 0.05$). Plants clipped on 12 and 20 July developed significantly more fruits than unclipped controls. Plants clipped on 28 July developed significantly fewer fruits than controls.

Source: After Lennartsson *et al.* (1998).

9.4.4 Meta-analyses of herbivory

We can find examples where herbivory stops plant growth or reproduction, or has a negligible effect, or does just about everything in between. Meta-analyses are attempts to go beyond catalogues of particular cases by critically reviewing the studies in a given field and developing a statistically sound consensus overview of their findings. A number of meta-analyses of the effects of herbivory have been carried out: for the effects of sap feeders on woody plants (Zvereva *et al.*, 2010), the effects of herbivory on roots (Zvereva & Kozlov, 2012), and on plant stems (Stephens & Westoby, 2015). A large number of trends and contrasts emerge from these, on different classes of herbivore, different types of study, and so on, and it would not be feasible to list them all here, but a selection of results is shown in Figure 9.25.

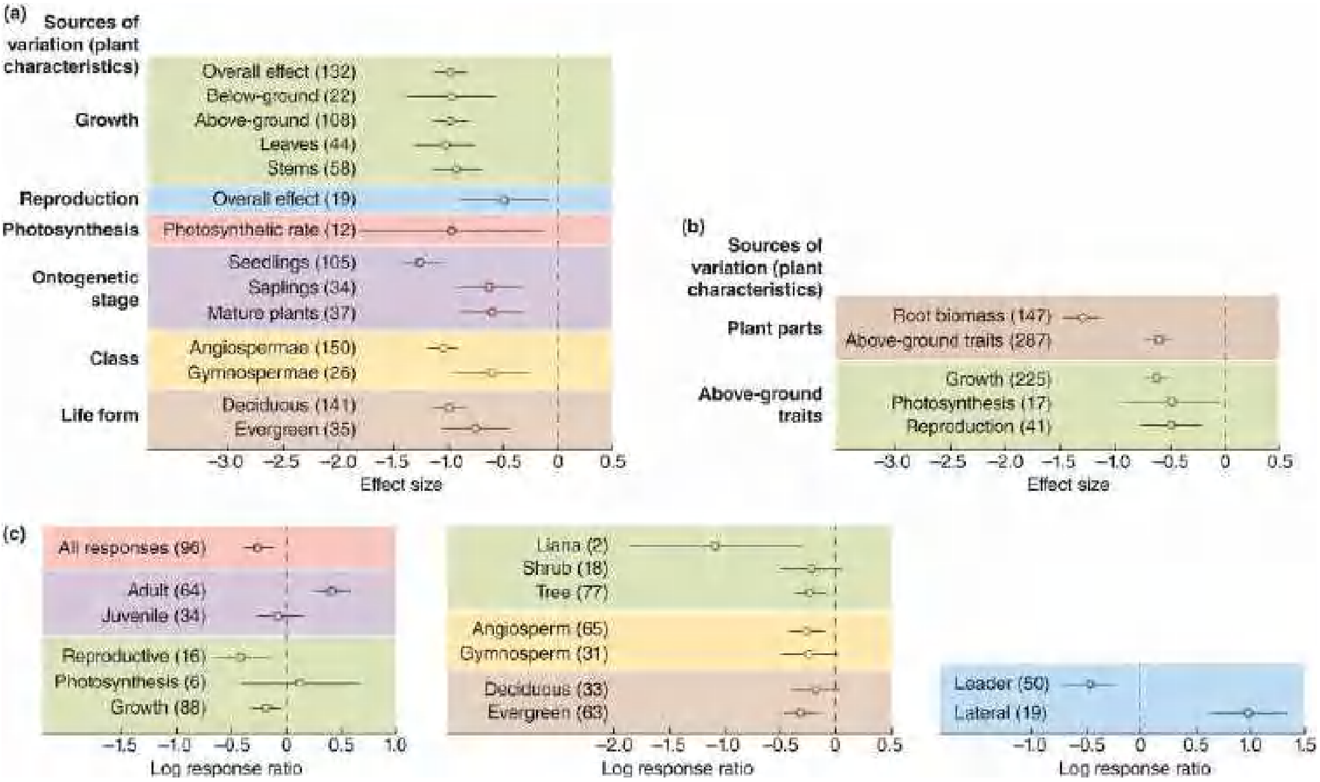


Figure 9.25 Some meta-analyses of herbivory. (a) The effects of sap feeders on woody plants. The effect size is measured by Hedges's *d* statistic, which is the difference between control and experimental (herbivory) data, divided by the pooled standard deviation, and weighted by the sample sizes in the different studies. Hence, negative values to the left of the dotted line indicate that plants subject to herbivory performed worse than controls; positive vales to the right indicate that they performed better. The bars are 95% CIs with the number of studies included in the analysis in parentheses. (b) Root herbivory. Statistics as in (a). (c) Stem herbivory. Here, the effect size is measured by the log response ratio, ln(experimental mean/control mean). Again, the bars are 95% CIs with the number of studies included in parentheses.

Source: (a) After Zvereva *et al.* (2010). (b) After Zvereva & Kozlov (2012). (c) After Stephens & Westoby (2015).

A number of common features emerge from these. The most obvious is that there is no shortage of studies demonstrating that herbivores have an adverse effect on plant performance; and this conclusion takes account of the fact that the authors tested, insofar as it is possible to do so, for publication bias (a tendency for studies to be disproportionately published the more 'positive' their results are), and they found none. It is also clear that this conclusion covers herbivores attacking the full range of plant parts. For both sap feeders and stem herbivores, effects are clearly, and not surprisingly, greatest when directed at the relatively unprotected early life stages of the plant (Figure 9.25a, c) (and this is a question not so easily addressed for root herbivory, where mature root systems only develop later in life).

Another recurring feature is that herbivory has less of an effect, and often even a positive effect, on photosynthetic rates compared with effects on growth or reproduction. This is a powerful reminder of how strong the compensatory responses of plants to herbivory may be, with photosynthetic rates in the immediate aftermath of herbivory often presaging ultimate reductions in growth or reproduction that are less profound than the immediate responses may have suggested. The relative strengths of effects on growth and reproduction are not so consistent. For the sap feeders, the effects on reproduction (–17%) were less than for growth (–29%; $P = 0.02$), though the difference between reproduction and growth of mature plants was not significant (P = 0.55; Figure 9.25a); for root herbivores, effects on reproduction were only marginally less than on growth (Figure 9.24b); while for stem herbivores, effects on reproduction were greater than on growth, though in this case separating out effects on adults and juveniles brought effects on reproduction and juvenile growth more into line (Figure 9.25c). What's perhaps most striking in the case of stem herbivores is that while the effects of herbivory on leader shoots was indeed strongly negative, those on lateral shoots was even more strongly positive. Plants often tolerate herbivory by changing their architecture, becoming bushier.

9.5 Animal defences

Animals have more options than plants when it comes to defending themselves, the most obvious being to run away. There are other behavioural strategies, too. Animals that withdraw to a prepared retreat (for example, rabbits and prairie dogs to their burrows, or snails to their shells), or which roll up and protect their vulnerable parts by a tough exterior (for example, armadillos and pill millipedes), reduce their chance of capture but stake their lives on the chance that the attacker will not be able to breach their defences. Other animals try to bluff themselves out of trouble by threat displays. The startle response of moths and butterflies that suddenly expose eye-spots on their wings is one example, and by 'playing dead', the opossum *Didelphis virginiana* may fail to stimulate a killing response.

chemical defences

Nonetheless, some animals do make use of chemicals. For example, defensive secretions of sulphuric acid of pH 1 or 2 occur in some marine gastropod groups, while other animals can tolerate the chemical defences of their plant food, store them and then use them in their own defence. A classic example is the monarch butterfly (*Danaus plexippus*; Figure 9.26a, b), whose caterpillars feed on milkweeds (*Asclepias* spp.). Milkweeds contain cardiac glycosides, which affect the vertebrate heartbeat and are poisonous to mammals and birds. Monarch caterpillars can store the poison, and it is still present in the adults, which in consequence are completely unacceptable to bird predators. A naive blue jay (*Cyanocitta cristata*) (i.e. one that has not tried a monarch butterfly before) will vomit violently after eating one, and once it recovers will reject all others on sight. In contrast, monarchs reared on cabbage are edible (Brower & Corvinó, 1967).

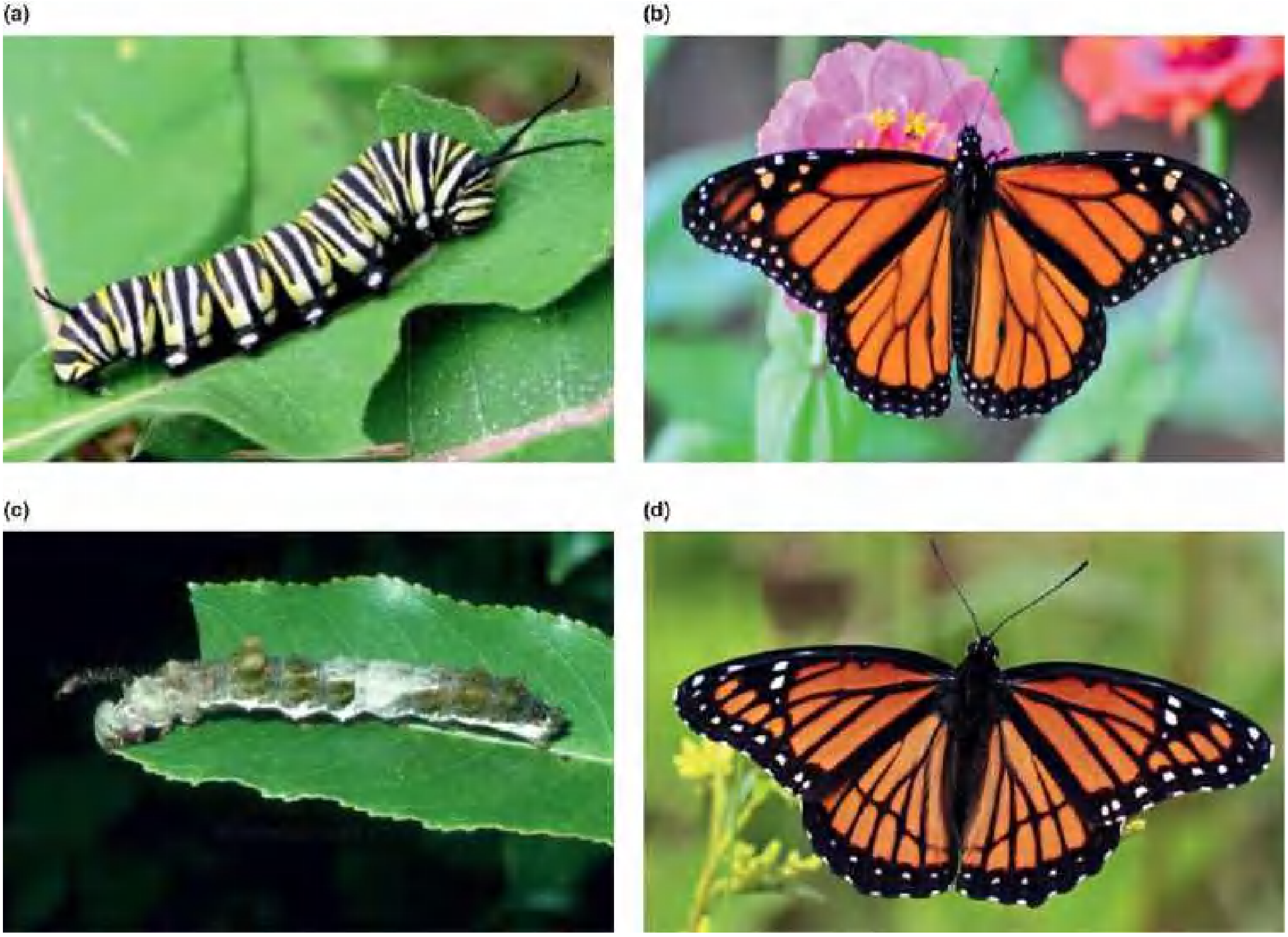


Figure 9.26 Mimicry in butterflies. (a) Larva of the monarch butterfly, *Danaus plexippus*, displaying aposematic colouration. (b) A monarch adult, also aposematically coloured. (c) Larva of the viceroy butterfly, *Limenitis archippus*, cryptically coloured to resemble a bird dropping. (d) A viceroy adult, a Batesian mimic of the monarch.

Chemical defences may be particularly important in modular animals, such as sponges, which lack the ability to escape from their predators. Despite their high nutritional value and lack of physical defences, most marine sponges appear to be little affected by predators (Kubanek *et al.*, 2002). In recent years, several triterpene glycosides, closely related to the plant defences we have discussed, have been extracted from sponges, including from *Ectyoplasia ferox* in the Caribbean. In a field study, crude extracts of refined triterpene glycosides from this sponge were presented in artificial food substrates to natural assemblages of reef fishes in the Bahamas. Strong antipredatory affects were detected when compared with control substrates (Figure 9.27). Notably, the triterpene glycosides also adversely affected competitors of the sponge, including organisms that overgrow them (bacteria, invertebrates and algae), and also other sponges (an example of allelopathy – see Section 8.3.3). All these enemies were apparently deterred by surface contact with the chemicals rather than by waterborne effects (Kubanek *et al.*, 2002).

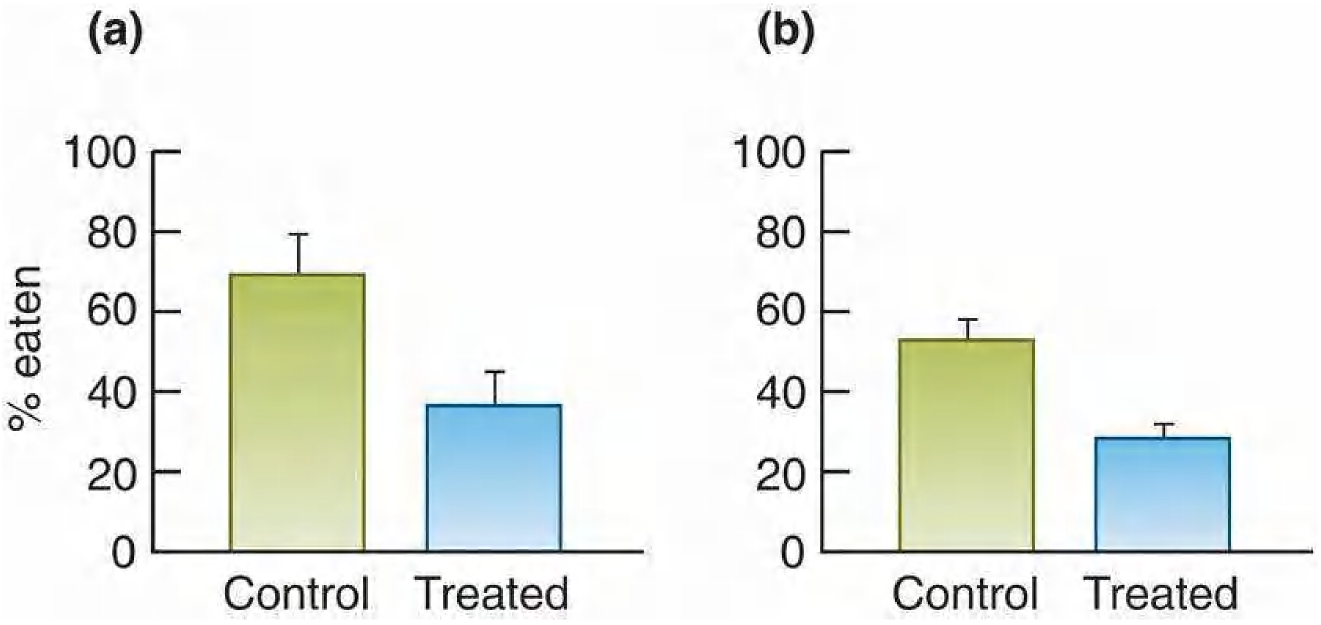


Figure 9.27 **Chemical defence in a sponge.** Results of field assays assessing antipredatory effects of compounds from the sponge *Ectyoplasia ferox* against natural assemblages of reef fish in the Bahamas. Means (+ SE) are shown for percentages of artificial food substrates eaten in controls (containing no sponge extracts) in comparison with: (a) substrates containing a crude sponge extract (*t*-test, $P = 0.036$) and (b) substrates containing triterpene glycosides from the sponge ($P = 0.011$).

Source: After Kubanek *et al*. (2002).

physical defences

Many animals utilise physical defences – spines, hard shells, and so on – as we have noted. Indeed, patterns we have discussed in the disposition of constitutive and inducible defences in plants have also been observed in the physical defences of animals. In a study of the shell morphology of marine snails, *Nucella lamellosa*, on San Juan Island, Washington, USA, juveniles were collected from sites with a low risk of predation by shell-crushing crabs, *Cancer productus* (swept regularly by currents or waves that make it difficult for the crabs to forage) and sites with a high risk ('quiet water' sites). These snails were then raised in an aquarium and either subjected to a 'crab' treatment, in which inflowing water contained 'effluent' from a crab being fed snails in an adjoining tank, or to a control treatment with no crab effluent. The consistent, predictable difference in predation risk between the different types of site generated an equally predictable difference in the snails' constitutive defences. Irrespective of experimental treatment, those from high risk sites had wider, shorter shells that the crabs find more difficult to crush (Figure 9.28). However, the snails' vulnerability to predation is also determined by the thickness of the shell around the shell's aperture (the apertural lip), and thickening of this turned out to be an inducible defence. Snails from both sites produced shells with thicker lips when subjected to indicators of contemporary crab predation (Figure 9.28), and indeed, this effect was greater for snails from the high-risk sites. Thus, not only were constitutive differences in defence supplemented by induced changes in the face of a perceived imminent risk of predation, but that inducibility was itself greater in snails that had evolved in the high-risk sites.

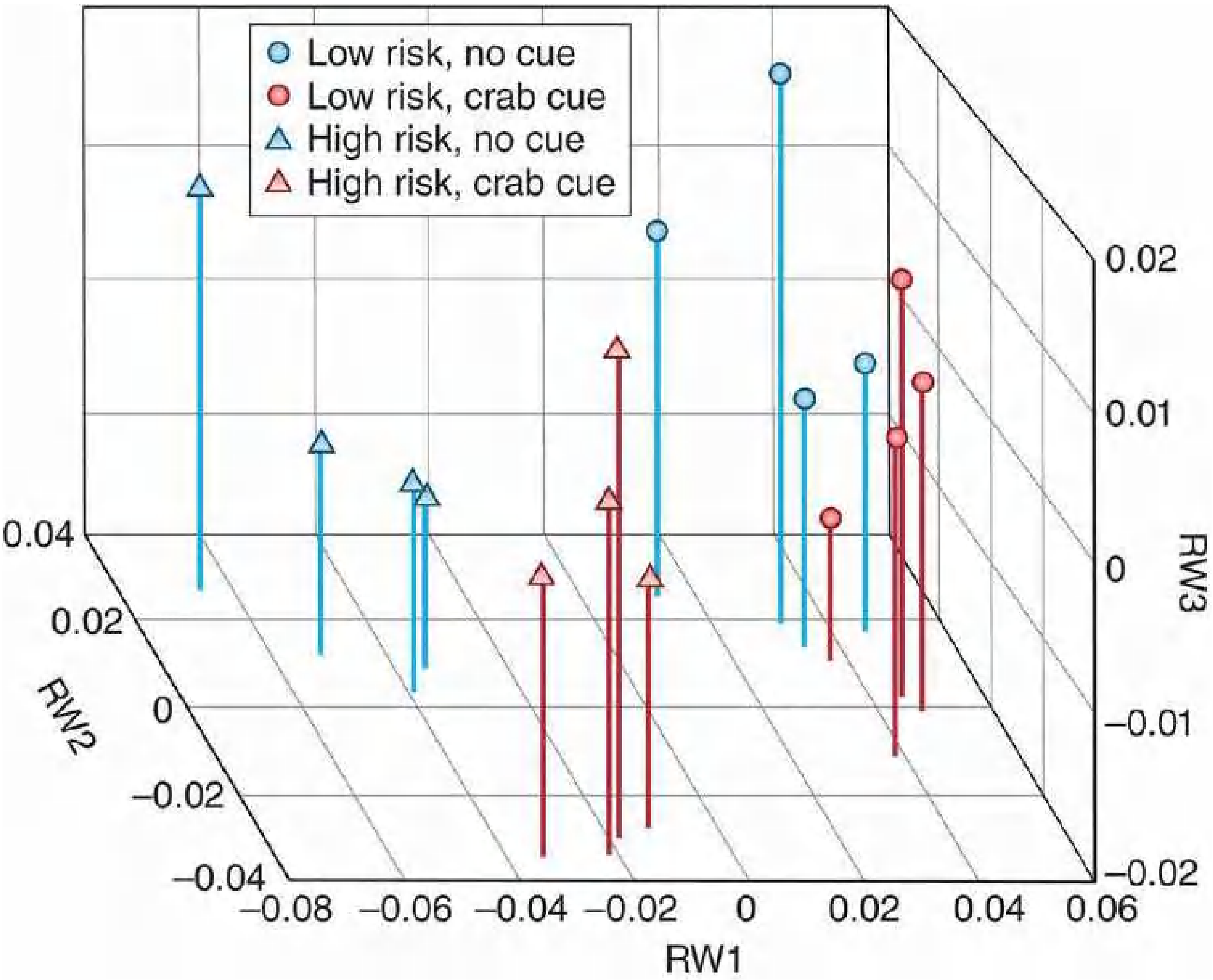


Figure 9.28 **Snail shell architecture varies with predation risk.** Three-dimensional plot of shape variation in the marine snail, *Nucella lamellosa*, collected from sites at high risk (triangles) and low risk (circles) of predation by crabs, *Cancer productus*, and then allowed to develop in aquarium water with (red symbols) and without (blue symbols) cues associated with crab predation. The shape variation analysis reduces total variation to a limited number of 'relative warps' (RWs, similar to principal components – see, for example, Figure 7.21), numbered in order of decreasing importance in terms of the amount of total variation they explain. Along RW1 (accounting for 47% of shape variation) low-risk sites had narrow, tall shells (positive scores), whereas high-risk sites had wider, shorter shells (negative scores). Along RW2 (accounting for 18% of shape variation) animals not subject to crab attention had shells with thinner apertural lips (positive scores), while those subject to crab attention had shells with thicker apertural lips (negative scores). RW3 accounted for only 6% of shape variation and subsequent RWs fewer still.

Source: After Bordeau (2012).

crypsis

Animals have more flexibility than plants when it comes to colouration (they do not have to be green), and many of them use colour and pattern to avoid predation, either by blending into the background (*crypsis*), or by standing out from their background, advertising their unpalatability through warning signals (*aposematism*), or by mimicking some other organism or object that would itself be unpalatable or uneatable (Ruxton *et* al., 2004 .) Straightforward examples of crypsis are the green colouration of many grasshoppers and caterpillars, and the transparency of many planktonic animals that inhabit the surface layers of oceans and lakes. More dramatic cases are the sargassum fish (*Histrio pictus*), whose body outline mimics the sargassum weed in which it is found, or the caterpillar of the viceroy butterfly (*Limenitis archippus*) that resembles a bird dropping (Figure 9.26c). Cryptic animals may be highly palatable, but their morphology and colour (and their choice of the appropriate background) reduce the likelihood that they will be used as a resource.

aposematism

While crypsis may be a defence strategy for a palatable organism, noxious or dangerous animals are often aposematic – evolving bright, conspicuous colours and patterns to advertise their unpalatability. The monarch butterfly, discussed previously, with its cardiac glycosides, is aposematically coloured, as is its caterpillar (Figure 9.26a, b). The usual evolutionary argument for this runs as follows: conspicuous colouration will be favoured because experienced predators (those who have tried and rejected noxious prey) will recognise subsequent potential prey items as noxious, and will therefore avoid these, whereas the costs of 'educating' the predator will have been shared amongst the whole population of conspicuous prey. As long as the signal is honest – conspicuous signals accurately reflect an underlying noxiousness – it is easy to see why aposematism should evolve, since the interests of predator and prey are in alignment: predators need to avoid prey as much as prey need to avoid predation (Summers *et al.*, 2015).

Batesian and Müllerian mimicry

However, the adoption of memorable body patterns by distasteful prey also immediately opens the door for deceit by other species, because there will be a clear evolutionary advantage to a palatable prey, 'the mimic', if it looks like an unpalatable species, 'the model'. This is referred to as Batesian mimicry. Developing the story of the monarch butterfly a little further, the adult of the palatable viceroy butterfly mimics the distasteful monarch (Figure 9.26 d), and a blue jay that has learned to avoid monarchs will also avoid viceroys. There will also be an advantage to aposematically coloured, distasteful prey in looking like one another, in that a predator having learned to associate appearance with unpalatability in one species will not need to sample the other species. This is referred to as Müllerian mimicry. In fact, though, the line between Batesian and Müllerian mimicry is not a clear one, especially if species that are Müllerian mimics of one another are not equally noxious. In such cases, less noxious species are close to being mimics of more noxious models, and the term quasi-Batesian mimicry has been coined to describe the relationship between them, though direct demonstration of quasi-Batesian mimicry remains rare (Rowland *et al.*, 2010).

9.6 The effect of predation on prey populations

predation may occur at a demographically unimportant stage

It may seem that since the effects of predators are harmful to individual prey, the immediate effect of predation on a population of prey must also be harmful. But this may not be so. Firstly, predation is unlikely to affect prey dynamics if it occurs at a stage of the prey's life cycle that does not have a significant effect, ultimately, on prey abundance. For example, if plant recruitment is not limited by the number of seeds produced, then insects that reduce seed production are unlikely to have an important effect on plant abundance. Thus, for the nodding thistle, *Carduus nutans*, in southern France, sowing 1000 thistle seeds per square metre led to no observable increase in the number of thistle rosettes. It is therefore no surprise that the weevil *Rhinocyllus conicus* does not reduce recruitment of the thistle despite inflicting seed losses of over 90% (Crawley, 1989).

compensatory reactions amongst survivors

There may also be compensatory changes in the growth, survival or reproduction of the surviving prey: they may experience reduced competition for a limiting resource, or produce more offspring, or other predators may take fewer of the prey. In short, while predation is bad for the prey that are eaten, it may be good for those that are not. Indeed, whenever density is high enough for intraspecific competition to occur, the effects of predation on a population should be ameliorated by the consequent reductions in intraspecific competition. Outcomes of predation may, therefore, vary with relative food availability. This was demonstrated by Oedekoven and Joern (2000), who monitored grasshopper (*Ageneotettix deorum*) survivorship in caged prairie plots. These were either unfertilised or subject to nitrogen fertilisation to increase food quality, and they either included, or lacked, predatory lycosid spiders (*Schizocosa* spp.). With ‘normal’ food quality (no fertiliser), spider predation and food limitation were compensatory (Figure 9.29). The grasshoppers were competing for food, so the removal of predators increased competition, such that the same number of grasshoppers survived to the end of the 31-day experiment. However, when food quality was higher, there was much reduced competition. Thus, this time when the spiders were absent, the grasshoppers did not experience the same increase in competition and the number surviving was therefore much higher than where spiders were present.

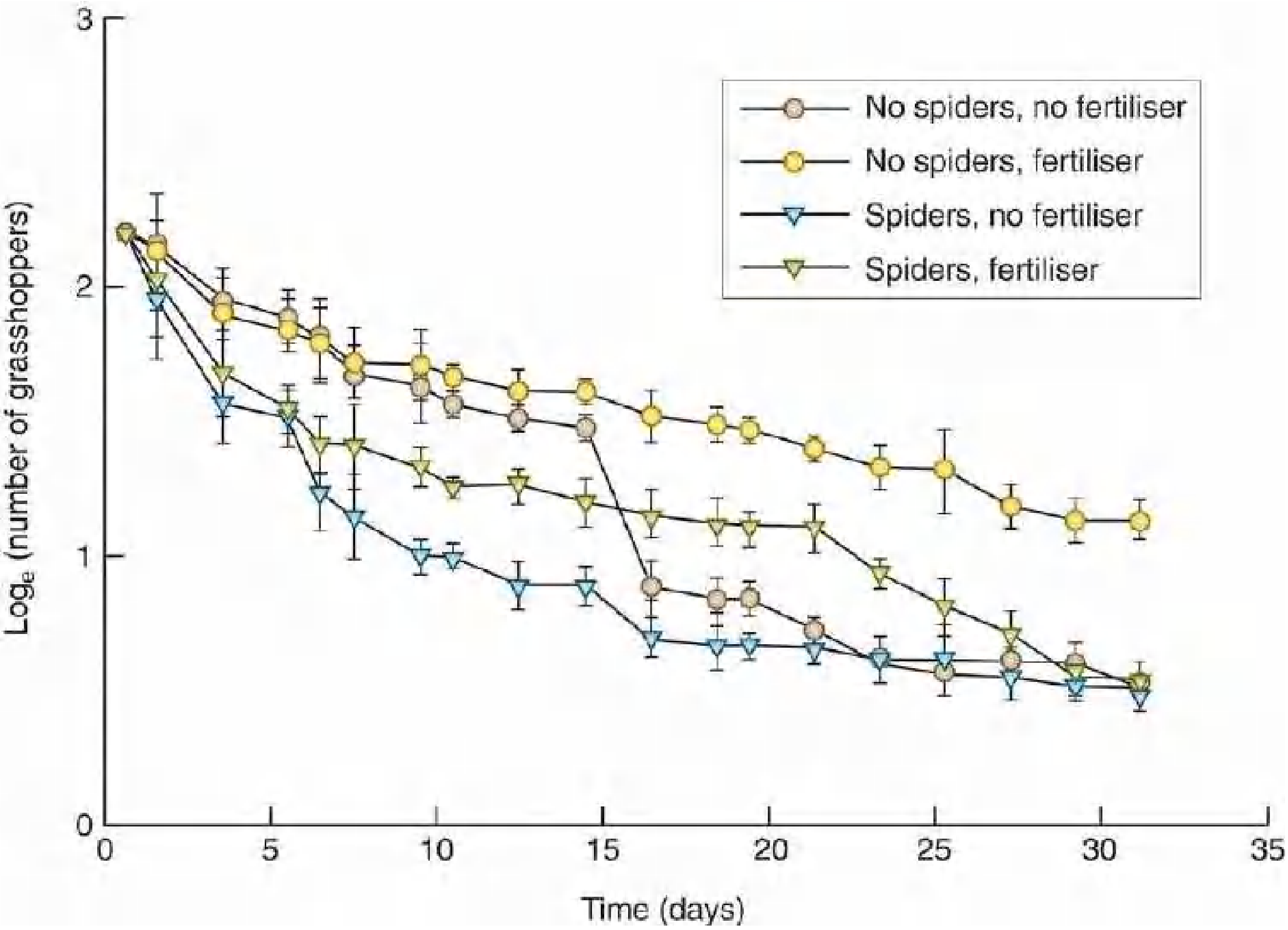


Figure 9.29 **The effects of predation may vary with food availability.** Trajectories of numbers of grasshoppers surviving (mean ± SE) for fertiliser and predation treatment combinations in a field experiment involving caged plots in the Arapaho Prairie, Nebraska, USA.

Source: After Oedekoven & Joern (2000).

predatory attacks are often directed at the weakest prey

Moreover, the individual prey that are killed (or harmed) are not always a random sample of the population as a whole, and may be those with the lowest potential to contribute to the population's future. For example, predation by many large carnivores is focused on the old (and infirm), the young (and naive) or the sick. This is apparent from a 30-year study of tawny owls, *Strix aluco*, themselves predators but here being preyed upon by northern goshawks, *Accipiter gentilis*, in Kielder Forest on the England–Scotland border (Hoy *et al*., 2015). The population sizes of these species were monitored over a period of around 30 years (Figure 9.30a), and these population estimates were used to test simulations of the effects of various patterns of selective predation. Female owls were more likely to be preyed upon by goshawks than the males, but most significantly, juvenile owls (i.e. those that were most likely to die anyway before they could reproduce) were almost four times more likely to be preyed upon than adults. Among adults, predation risk increased with age (Figure 9.30b), i.e. it was focused on those who had their best reproductive years behind them. Those most likely to contribute to future generations, young adults, were also most immune from predation.

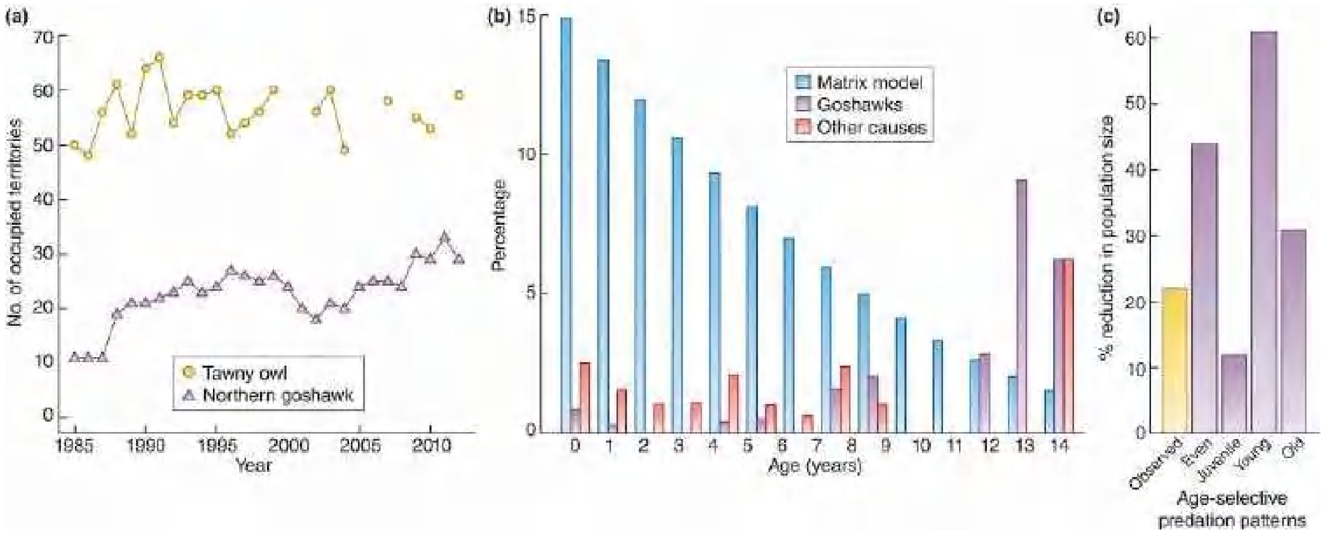


Figure 9.30 Goshawks preying on owls mostly take those least likely to contribute to owl population growth. (a) Abundance estimates for tawny owls and northern goshawks (both, numbers of territories occupied) in Kielder Forest, UK, 1985–2012. (b) The expected percentage of adult owls in different age classes when the population was at equilibrium, as predicted by a matrix model of the population (blue bars), and the percentage of owls that reached each age that were preyed upon by goshawks (purple bars) or died of other causes (red bars). (c) The effect of goshawk predation on owl abundance (percentage reduction in population size) in the matrix population model subjected either to the observed pattern of age-selective predation or four other patterns, as indicated.

Source: After Hoy *et al.* (2015).

The consequent relative insignificance of goshawk predation for owl population size was borne out, first, by simulation models of owl population dynamics in which the owls were subjected either to the observed pattern of selective predation or four other patterns, targeting juveniles, young adults or old adults, or with predation evenly distributed through the population (Figure 9.30c). The observed pattern had low impact, close to the pattern with least impact, simply targeting the juveniles. Moreover, statistical models aimed at accounting for the fluctuations in owl population size (Figure 9.30a) could find no support for an association with goshawk abundance, either currently or in the past, nor for that matter for an association with the abundance of voles, the owl's food.

Similar patterns may also be found in plant populations. The mortality of mature eucalyptus trees in Australia, resulting from defoliation by the leaf beetle, *Paropsis atomaria*, was restricted almost entirely to weakened trees on poor sites, or to trees that had suffered from root damage or from altered drainage following cultivation (Carne, 1969).

difficulties of demonstrating effects on prey populations

Overall, taking the step from noting that individual prey are harmed by predators to demonstrating that prey abundance is adversely affected is not easy. Of 28 studies in which herbivorous insects were experimentally excluded from plant communities using insecticides, 50% provided evidence of an effect on plants at the population level (Crawley, 1989). As Crawley noted, however, such proportions need to be treated cautiously. There is an almost inevitable tendency for negative results (no population effect) to go unreported, on the grounds that there was 'nothing' to report. Moreover, the exclusion studies often took seven years or more to show any impact on the plants: it may be that many of the negative studies were simply given up too early. Many more recent investigations have shown clear effects of seed predation on plant abundance (e.g. Kelly & Dyer, 2002 ; Maron *et al.*, 2002).

9.6.1 Intimidation: the non-consumptive effects of risk

On the other hand, there are important senses in which the effects of predators on prey can be more profound than is immediately apparent, with consequently harmful effects, potentially, on whole prey populations. We saw this for the effects of herbivory on plants in Section 9.4. We also saw in Section 9.2.4 that prey may be affected behaviourally, in their foraging, by the presence of predators, or even by the risk of predation. In fact, there is a wide range of such 'trait-mediated effects' of predators on prey (Werner & Peacor, 2003), including, for example, the development of protective morphological traits such as the spines of the cladoceran, *Daphnia*. Since such responses are likely to be costly to the prey, the potential exists for these, too, to reduce the prey's survival, fecundity and ultimately their abundance, though of course, since these trait-mediated effects have evolved to counter the effects of predation, they may also increase prey abundance.

To take one example, arctic ground squirrels (*Urocitellus parryii*) in the Yukon, Canada are preyed upon by a variety of predators, including Canada lynx (*Lynx canadensis*), coyotes (*Canis latrans*), great horned owls (*Bubo virginianus)*, red-tailed hawks (*Buteo jamaicensis*), and goshawks (*Accipiter gentilis*). A large-scale field experiment was carried out on four control and four experimental areas: one from which predators were excluded (effective for the mammals but only partly so for the birds), two in which the ground squirrels' food was supplemented, and one in which there was both predator exclusion and food supplementation. Results are shown in Figure 9.31. The ground squirrels alter their behaviour both in direct response to the presence of predators and in habitats where the risk of predation is higher. They also make a hormonal 'stress' response – the release of glucocorticoids – which reduces fecundity in mammals generally (Boonstra, 2013). It is striking, therefore, that predator exclusion had little effect on ground squirrel survival (and hence there was little evidence for predators having a direct, consumptive effect on the ground squirrels) but improved body condition increased the percentage weaning litters, and the litter size, and doubled population density relative to controls; while predator exclusion and food supplementation together often had even greater effects (though not on litter size (Figure 9.31a) where effects were inconsistent between years), and ultimately led to a massive increase in population density (Figure 9.31b). The indirect effects of predators therefore appeared to be greater than the direct effects. The results emphasise, too, that predators, through these indirect effects, can affect fecundity as well as survival. The importance of indirect effects of predation will also be apparent when we examine the cyclic dynamics of snowshoe hares (*Lepus americanus*) in Section 14.6.2.

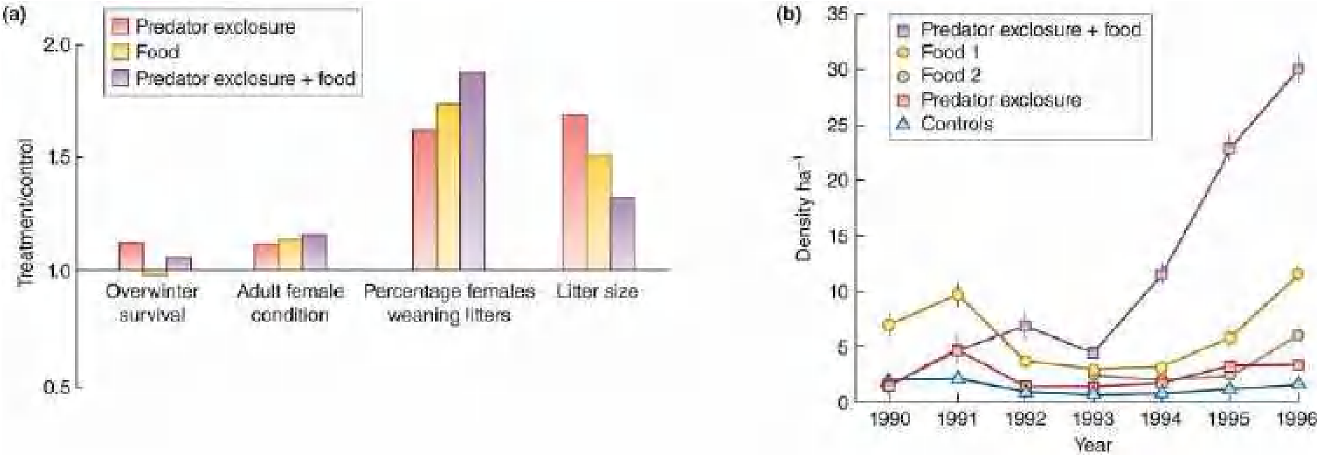


Figure 9.31 Predators have more effect on the fecundity of ground squirrels than they do on their survival. (a) The effects on female arctic ground squirrels (*Urocitellus parryii*), in the Yukon, of predator exclusion, food supplementation, and the two combined, on a range of metrics as indicated. These were averages of the ratios of treatment to control values over six years (1990–95). (b) The estimated spring breeding densities of ground squirrels over the six years in control and treatment plots, as indicated. Bars are SEs (sometimes so small as to be hidden by the points themselves).

Source: After Karels *et al*. (2000).

Preisser *et al*. (2005) carried out a meta-analysis, assessing the magnitudes of the effects of intimidation and consumption on prey abundance, that is, non-consumptive effects and direct consumptive effects. These were measured by Hedges' *d* statistic, the difference between treatment and control values, divided by their pooled standard error. They found that the mean magnitudes of non-consumptive, consumptive and total effects on prey abundance, relative to controls, were −0.61, −1.87 and −1.71, respectively – that is, the effects of non-consumptive effects, while smaller than those of consumptive effects, were nonetheless highly significant. Finding consumptive effects to exceed total effects is readily explained by noting, as anticipated above, that some of the non-consumptive effects on prey abundance were positive, not negative (their protective value exceeded their costs) and so will have counteracted the effects of consumptive effects. Indeed, non-consumptive effects made a greater contribution (positive and negative) to the total effect, overall, than did consumptive effects. Clearly, therefore, although ecologists have often ignored non-consumptive effects and focused only on the more obvious consumptive effects of predators, it is dangerous to do so, even when seeking to account for the negative effects of predation on prey abundance.

Thus, for a range of reasons, we should not assume that the consumptive effects of predators on prey can be equated directly with the harmful effects of predators on prey populations and the beneficial effects on predator populations. In the next chapter, we examine the effects of these and other aspects of predator–prey interactions on the population dynamics of both parties.

Chapter 10

The Population Dynamics of Predation

10.1 The underlying dynamics of consumer-resource systems: a tendency towards cycles

Having focused in the previous chapter on how predators and prey interact, we now turn to the population dynamics that emerge from those interactions. The approach will be to start by using simple models to establish some fundamental patterns. It will be apparent, however, that in order to do so we will be making simplifications that ignore many of the details of how predators and prey behave and respond to one another. In the remainder of the chapter, therefore, we will examine the most important of these predator and prey responses and their effects on the models, teasing out their separate effects before seeking to understand those effects in combination. Then, field and experimental data will be examined to see whether the deductions from models appear to be supported or refuted. In fact, simple models are most useful when their predictions are *not* supported by real data – as long as we can discover the reason for the discrepancy. Confirmation of a model's predictions provides consolidation; refutation with subsequent explanation is progress.

There have been two main series of models developed as attempts to understand predator–prey dynamics. Both will be examined here. The first (the Lotka–Volterra model, Section 10.1.1) is based on differential equations and hence applies most readily to populations in which breeding is continuous, but it also relies heavily on simple graphical models (Rosenzweig & MacArthur, 1963). The second (the Nicholson–Bailey model, Section 10.1.3) uses difference equations to model interactions between hosts and parasitoids with discrete generations. Despite this focus on a relatively narrow taxonomic group (though we have noted previously the very large number of important parasitoid species), these models have the advantage of having been subject to rigorous mathematical exploration. In fact, the two models have many similarities and produce similar dynamics, which is to be expected as they have the common aim of advancing our understanding of predator–prey dynamics by capturing the essence of predator–prey interactions. But they are only models. They will be useful in helping us interpret what we see in nature, but it is nature that we wish to understand.

10.1.1 The Lotka–Volterra model

The simplest differential equation model, named (like the model of interspecific competition) after its originators, Lotka and Volterra, will serve as a useful point of departure. The model has two components: P, the numbers present in a predator (or consumer) population, and N, the numbers or biomass present in a prey or plant population.

the Lotka–Volterra prey equation

We assume initially that in the absence of consumers the prey population increases exponentially (see Section 5.7):

$$dN/dt = rN, \tag{10.1}$$

where in this case, r is the intrinsic rate of natural increase of the prey. But prey individuals are removed by predators. The rate at which this occurs will depend on the frequency of predator–prey encounters, which will increase with the numbers of predators (P) and with the numbers of prey (N). However, the exact number encountered and successfully consumed will depend on the searching and attacking efficiencies of the predator, denoted by a, and sometimes called the 'attack rate'. The consumption rate of prey will thus be aPN, and overall:

$$dN/dt = rN - aPN. \tag{10.2}$$

the Lotka–Volterra predator equation

In the absence of prey, predator numbers are assumed to decline exponentially, through starvation:

$$dP/dt = -qP, \tag{10.3}$$

where q is the predator mortality rate. This is counteracted, though, by predator birth. The rate is assumed to depend simply on the rate at which food is consumed (see Equation 10.2), aPN, and on the predator's efficiency, f, at turning this food into predator offspring. Predator birth rate is therefore $faPN$, and overall:

$$dP/dt = faPN - qP. \tag{10.4}$$

Equations 10.2 and 10.4 constitute the Lotka–Volterra model.

The properties of this model can be investigated by finding zero isoclines. Zero isoclines were described for models of two-species competition in Section 8.4.1. Here, there are separate zero isoclines for the predators and prey, both of which are drawn on a graph of prey density (x-axis) against predator density (y-axis). Each is a line joining those combinations of predator and prey density that led either to an unchanging prey population ($dN/dt = 0$; prey zero isocline) or an unchanging predator population ($dP/dt = 0$; predator zero isocline). Having drawn, say, a prey zero isocline, we know that density-combinations to one side of it led to prey decrease, while combinations to the other side led to prey increase. Thus, as we shall see, if we plot the prey and predator zero isoclines on the same figure, we can begin to determine the pattern of the dynamics of the joint predator–prey populations.

In the case of the prey (Equation 10.2), when $dN/dt = 0$, it follows that

$$rN = aPN \tag{10.5}$$

or:

$$P = r/a. \tag{10.6}$$

properties revealed by zero isoclines

Thus, since r and a are constants, the prey zero isocline is a line for which P, predator abundance, is constant (Figure 10.1a). Below the line, predator abundance is low and the prey increase; above it, predator abundance is high and the prey decrease. Likewise, for the predators (Equation 10.4), when $dP/dt = 0$, it follows that

$$faPN = qP \tag{10.7}$$

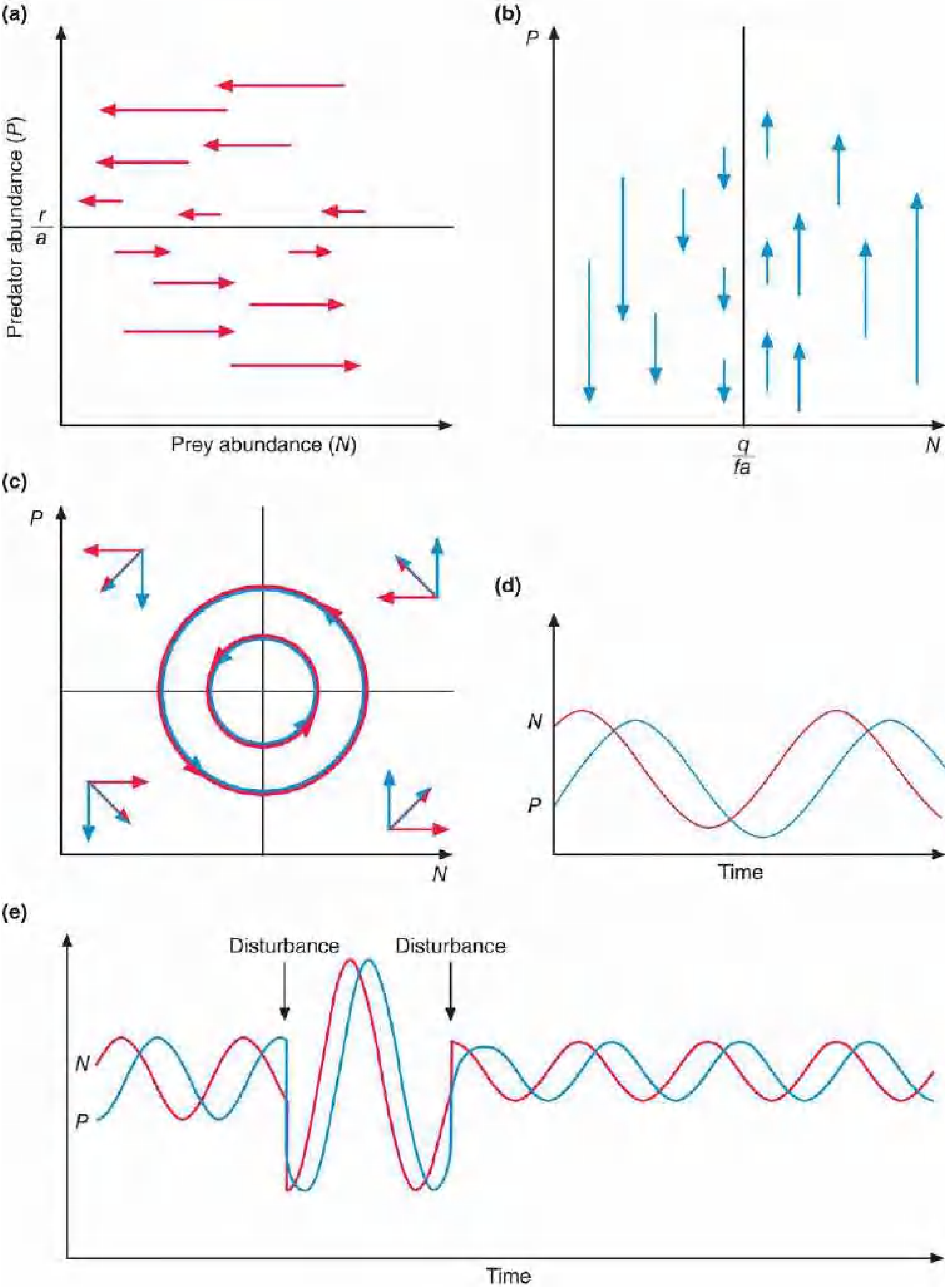

(a)
Predator abundance (P)
r/a
Prey abundance (N)
(b)
P
q/fa
N
(c)
P
N
(d)
N
P
Time
(e)
Disturbance
Disturbance
N
P
Time

Figure 10.1 The Lotka–Volterra predator–prey model. (a) The prey zero isocline, with prey (*N*) increasing in abundance (arrows left to right) at lower predator densities (low *P*) and decreasing at higher predator densities. (b) The predator zero isocline, with predators increasing in abundance (arrows pointing upwards) at higher prey densities and decreasing at lower prey densities. (c) When the zero isoclines are combined, the arrows can also be combined, and these joint arrows progress in anticlockwise circles. In other words, the joint population moves with time from low predator/low prey (bottom left in (c)), to low predator/high prey (bottom right), to high predator/high prey, to high predator/low prey and back to low predator/low prey. Note, however, that the lowest prey abundance ('9 o'clock') comes one-quarter of a cycle before the lowest predator abundance ('6 o'clock' – anticlockwise movement). These coupled cycles of predator–prey abundance, continuing indefinitely, are shown as numbers against time in (d). However, as shown in (e), these cycles exhibit neutral stability: they continue indefinitely if undisturbed, but each disturbance to a new abundance initiates a new, different series of neutrally stable cycles, around the same means but with a different amplitude.

or:

$$N = q/fa. \tag{10.8}$$

The predator zero isocline is therefore a line along which *N*, prey abundance, is constant (Figure 10.1b). To the left, prey abundance is low and the predators decrease; to the right, prey abundance is high and the predators increase.

Putting the two isoclines together (Figure 10.1c) shows the behaviour of joint populations. Predators increase in abundance when there are large numbers of prey, but this leads to an increased predation pressure on the prey, and thus to a decrease in prey abundance. This then leads to a food shortage for predators and a decrease in predator abundance, which leads to a relaxation of predation pressure and an increase in prey abundance, which leads to an increase in predator abundance, and so on (Figure 10.1d). Thus, predator and prey populations undergo 'coupled oscillations' in abundance, which continue indefinitely.

The Lotka–Volterra model, then, is useful in pointing to this underlying tendency for predator–prey interactions to generate fluctuations in the prey population tracked by fluctuations in the predator population. However, the detailed behaviour of the model should not be taken seriously, because the cycles it exhibits have 'neutral stability'. That is, the populations would follow precisely the same cycles indefinitely, but only until some external influence shifted them to new values, after which they would follow new cycles indefinitely rather than returning to the original pattern (Figure 10.1e). In practice, though, environments are continually changing, and populations would therefore no sooner start one cycle than they would be diverted to a new one. Hence, a population following the Lotka–Volterra model would *not* exhibit regular cycles, but would fluctuate erratically in the face of repeated disturbance.

For a differential equation model to show regular cycles of constant amplitude, the cycles must themselves be stable: when an external influence changes the population level, there must be a tendency to return to the original cycle. In fact, as we shall see, predator–prey models (once we move beyond the very limiting assumptions of Lotka–Volterra) are capable of generating a whole range of abundance patterns: stable-point equilibria, multigeneration cycles, one-generation cycles, chaos, etc. – a range repeated in surveys of real populations. The challenge for the models is to see what light they can shed on the behaviour of real populations.

10.1.2 Delayed density dependence

numerical responses

The underlying process generating coupled oscillations in predator–prey interactions is a series of time-delayed 'numerical responses' (i.e. changes in one species' abundance in response to the abundance of the other species). For example, there may be a time delay between 'many prey' and 'many predators' because the response of predator abundance to high prey abundance does not occur instantaneously but only after an increased predator reproduction rate gives rise to increased predator abundance. There may be another time delay between 'many predators' and 'few prey', and between 'few prey' and 'few predators', and so on. In practice, therefore, even where coupled oscillations exist, their exact shape is likely to reflect the varying delays, and strengths, of the different numerical responses. Certainly, the shapes of apparent coupled oscillations in real populations are varied, and not all are symmetric like those generated by the Lotka–Volterra model (see Figure 9.1).

the regulatory tendencies of delayed density dependence can be difficult to demonstrate

These responses show density dependence (see Section 5.2), in that they act to reduce the size of relatively large populations and allow relatively small populations to increase, but it is *delayed density dependence* (Varley, 1947). Its strength is related not to the current abundance (that would be *direct density dependenc*; see, for example, Figure 5.11) but to abundance at some time in the past (i.e. the delay-length ago). Compared with direct density dependence, delayed density dependence is relatively difficult to demonstrate. To see this, we can examine the coupled oscillations produced by a particular predator–prey model, shown in Figure 10.2a (Hassell, 1985). The details of the model need not concern us, but note that the oscillations are damped: they get gradually smaller over time until a stable equilibrium is reached. The prey population, subject to delayed density dependence, is regulated in size by the predator. In Section 5.3, we demonstrated density dependence by plotting k values against the log of density; but in the present case, when we plot the k values of predator-induced mortality against the log of prey density in that generation (Figure 10.2b), no clear relationship is apparent. On the other hand, when the same points are linked together, each generation to the next (Figure 10.2c), they can be seen to describe an anticlockwise spiral. This spiralling is characteristic of delayed density dependence. Here, because the oscillations are damped, the points spiral inwards to the equilibrium point. Moreover, when we plot the k values of predator-induced mortality against the log of prey density two generations previously (Figure 9.18d), the delayed density dependence is clearly revealed by the positive relationship characteristic of density dependence in general. Indeed, the fact that a two-generation delay gives a better fitting relationship than delays that are either shorter or longer, tells us that two generations is our best estimate of the delay in this case.

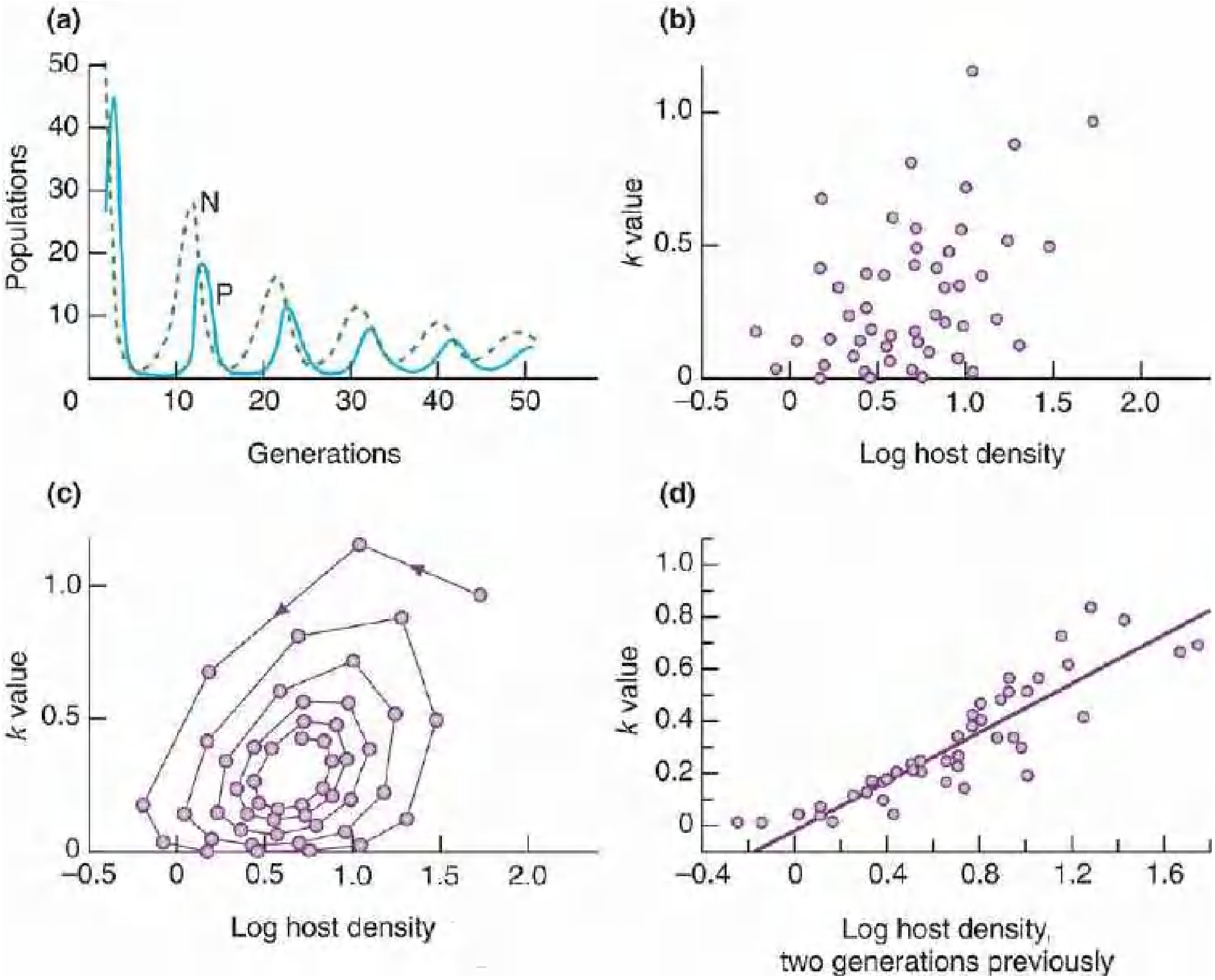


Figure 10.2 Delayed density dependence. (a) A parasitoid–host model followed over 50 generations: despite oscillations, the parasitoid has a regulatory effect on host population size. N, prey. P, predator. (b) For the same model, the *k* value of generation mortality plotted against the log of host density: no clear density-dependent relationship is apparent. (c) The points from (b) linked serially from generation to generation: they spiral in an anticlockwise direction – a characteristic of delayed density dependence. (d) The *k* value of generation mortality plotted against the log of host density two generations previously: a clear delayed density-dependent relationship is again apparent.

Source: (c) After Hassell (1985).

The regulatory effects of delayed density dependence are easily revealed for the model population of Figure 10.2, because it is not subject to the fluctuations of a natural environment, it is not subject to the density-dependent attacks of any other predator, and it is not subject to the inaccuracies of sampling error. However, data of this quality are rarely if ever available for natural or even experimental populations. We'll return in Chapter 14 to the question of uncovering delayed density-dependent effects and integrating them into an overall account of what determines abundance.

For now, though, note that what we have already seen highlights the relationship between 'regulation' and 'stability' in predator–prey interactions. Most natural predator and prey populations tend to exhibit lower amplitude and less regular fluctuations than those generated by the simplest models. Most of the rest of this chapter describes the search for explanations for these patterns. A population that remains roughly constant in size provides evidence for the effects of both regulatory and stabilising forces. The delayed density dependence of a predator–

prey interaction 'regulates' in the sense of acting strongly on large populations and only weakly on small populations. But if this leads to extravagant oscillations, it can hardly be said, typically, to 'stabilise'. What follows in this chapter, therefore, is often a search for stabilising forces that might complement the (delayed) regulatory forces that occur inherently in predator–prey interactions.

10.1.3 The Nicholson–Bailey model

Turning now to parasitoids, the basic model (Nicholson & Bailey, 1935) is again not so much realistic as a reasonable basis from which to start. Let H_t be the number of hosts, and P_t the number of parasitoids in generation t; r is the intrinsic rate of natural increase of the host. If H_a is the number of hosts attacked by parasitoids (in generation t), then, assuming no intraspecific competition amongst the hosts (exponential growth – see Section 5.6), and that each host can support only one parasitoid (commonly the case):

$$H_{t+1} = e^r(H_t - H_a), \tag{10.9}$$

$$P_{t+1} = H_a. \tag{10.10}$$

In other words, hosts that are not attacked reproduce, and those that are attacked give rise, in the next generation, not to hosts but to parasitoids.

To derive a simple formulation for H_a, we first consider E_t, the number of host–parasitoid encounters in generation t. Then, if A is the parasitoid's searching efficiency:

$$E_t = AH_tP_t \tag{10.11}$$

and:

$$E_t/H_t = AP_t. \tag{10.12}$$

Note the similarity to the formulation in Equation 10.2. Remember, though, that here, multiple events (attacks on hosts) can occur within a generation, and therefore it is necessary to consider the distribution of these attacks among hosts. In the Lotka–Volterra model, by contrast, time is modelled continuously, so the interaction between predators and prey depends on aPN, the instantaneous rate at which prey are captured.

a model based on random encounters ...

If encounters are assumed to occur more or less at random, then the proportions of hosts encountered zero, one, two or more times are given by the successive terms in the appropriate 'Poisson distribution' (see any basic statistics textbook). The term for the proportion not encountered at all, p_0, is given by e^{-E_t/H_t}. Thus the proportion that *is* encountered (one or more times) is $1 - e^{-E_t/H_t}$. The number encountered (or attacked) is then:

$$H_a = H_t\left(1 - e^{-E_t/H_t}\right). \tag{10.13}$$

Using this and Equation 10.12 to substitute into Equations 10.9 and 10.10 gives us:

$$H_{t+1} = H_t e^{(r - AP_t)} \tag{10.14}$$

$$P_{t+1} = H_t\left(1 - e^{(-AP_t)}\right). \tag{10.15}$$

... giving rise to (unstable) coupled oscillations

This is the basic Nicholson–Bailey model. Its behaviour is reminiscent of the Lotka–Volterra model but it is even less stable. An equilibrium combination of the two populations is a possibility, but even the slightest disturbance from this equilibrium leads to divergent coupled oscillations.

one-generation cycles

The coupled oscillations generated by the basic Lotka–Volterra and Nicholson–Bailey models are termed multigeneration cycles because there are several generations between successive peaks (or troughs) of abundance. These have lain at the heart of most attempts to understand cyclic predator–prey dynamics. However, other models of host–parasitoid (and other) systems are able to generate coupled oscillations just *one* host generation in length, and such cycles are sometimes, though quite rarely, observed (see, for example, Figure 9.1c). The dynamics of these 'generation cycles' are reviewed by Knell (1998).

10.1.4 Predator–prey cycles in nature: or are they?

The inherent tendency for predator–prey interactions to generate coupled oscillations in abundance, revealed in both the Lotka–Volterra and the Nicholson–Bailey models, suggests that we might expect to see such oscillations in real populations. However, many important aspects of predator and prey ecology have simply been left out of the models examined so far, and as subsequent sections will show, including them can greatly modify our expectations. And even if a population exhibits regular oscillations, this does not necessarily provide support for these or any other simple models. We saw cycles generated by intraspecific competition in Section 5.6, and we shall see several other routes to cycles in subsequent chapters. Hence, even when predators or prey exhibit regular cycles in abundance, it is never easy to demonstrate that these are *predator–prey* cycles.

hare and lynx: not the simple predator and prey they appear to be

The regular oscillations in the abundance of the snowshoe hare and the Canadian lynx, shown in Figure 9.1a, have often been said to epitomise predator–prey cycles. Recently, however, evidence has increasingly suggested that even this apparent exemplar is not as straightforward as it seems. Experimental manipulations carried out in the field are one powerful means of suggesting what forces are normally acting: if those forces are removed is the cycle eliminated? If the forces are increased is the cycle enhanced? In this case, a whole series of coordinated field manipulations has indicated that the cyclic hare is not simply a prey of the lynx (and other predators in the community), nor simply a predator of the plants it eats. The cycle can be understood only by taking account of its interactions both as a prey *and* as a predator (see Section 14.6.2). Modern statistical analysis of the time series of abundances supports this. The hare series incorporates a relatively complex 'signal', suggesting the influence of both its predators and its food; whereas the lynx series has a simpler signal, suggesting only the influence of its prey, the hare (Stenseth *et al.*, 1997; see also Section 14.5). So, what has so often been described as a predator–prey cycle seems rather to comprise one predator linked to a species that is both predator and prey.

We return to the question of cycles – indeed, some of the same cycles just discussed – in Section 14.6, as part of a more general exploration of how the whole range of biotic and abiotic factors come together to determine the level and pattern of a population's abundance.

10.2 Patterns of consumption: functional responses and interference

consumers often need to exceed a threshold of consumption

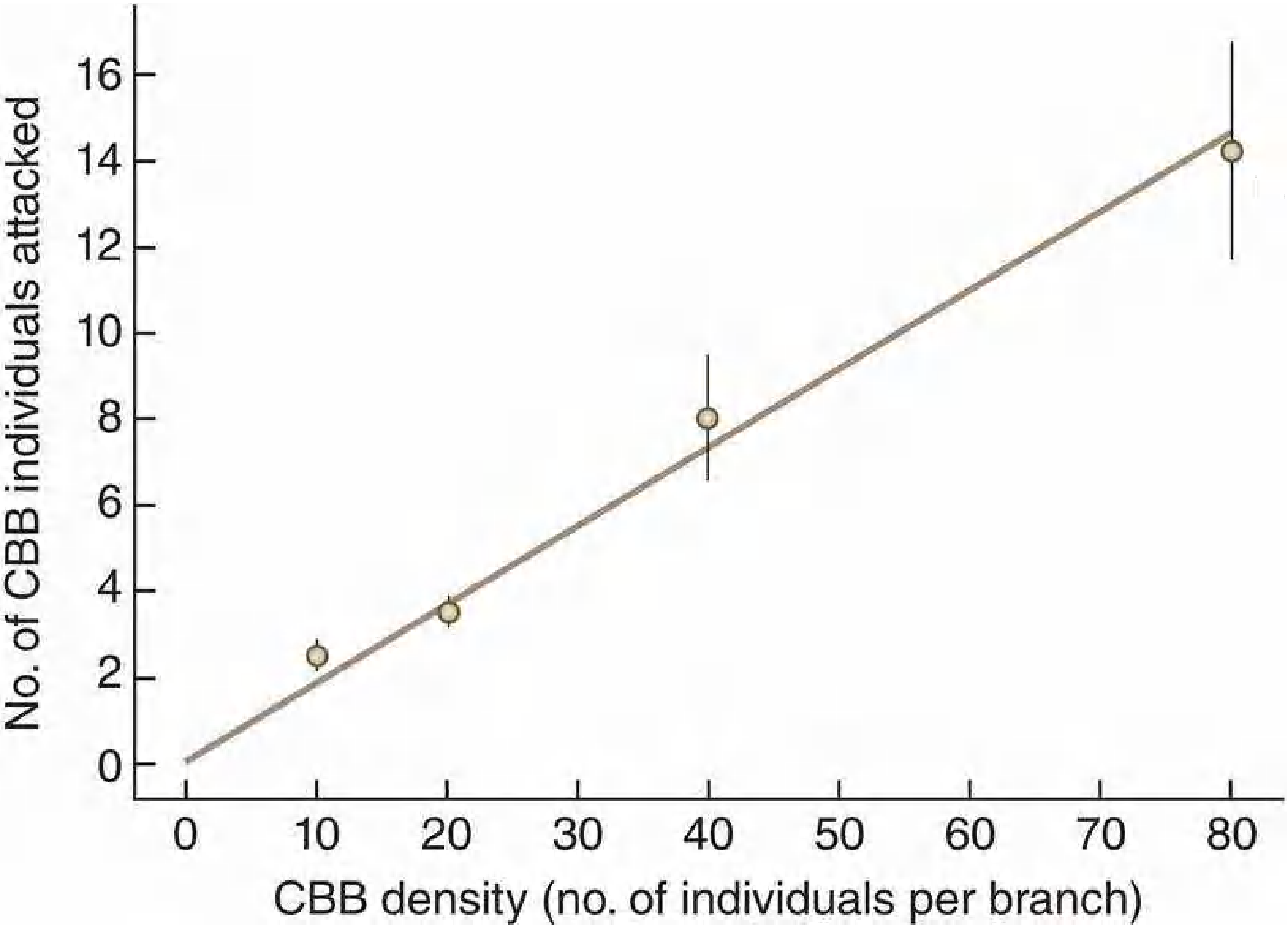
Having established some baseline dynamics from very simple models, we can now begin to consider features of real interactions that could be incorporated into those models. We start with patterns of consumption.

The beneficial effects that food has on individual predators are obvious. Generally speaking, the more food predators consume, the greater their rates of growth, development and birth, and the lower their rates of mortality. This, after all, is at the heart of intraspecific competition (see Chapter 5): high densities, implying small amounts of food per individual, lead to low growth rates, high death rates, and so on. Similarly, many of the effects of migration considered in Chapter 6 reflect the responses of individual consumers to the distribution of food availability. However, there are a number of ways in which the relationships between consumption rate and consumer benefit can be more complicated than they initially appear. In the first place, all animals require a certain amount of food simply for maintenance, and unless this threshold is exceeded the animal will be unable to grow or reproduce. Here, low consumption rates, rather than providing a small benefit to the consumer, simply alter the rate at which the consumer starves to death.

10.2.1 The type 1 functional response

More generally, consumers do not necessarily simply consume more, and benefit more, in direct proportion to the amount of food available. This dependence of consumption rate on prey abundance is known as the *functional response* of the predator (Solomon, 1949). The response reflects the behaviour of individual predators. However, it is of course the combined effect of a population of predators that affects the population dynamics.

Holling (1959) was the first to distinguish three 'types' of functional response. The most basic, 'type 1' functional response is that assumed by the Lotka–Volterra and Nicholson–Bailey equations: consumption rate rises linearly with prey density. This was indicated by the constant, a, in Equation 10.2, and the functional response in the simple model is thus aN. An example of a type 1 response is illustrated in Figure 10.3. The number of individual coffee berry borers, *Hypothenemus hampei*, attacked by the ant, *Azteca sericeasur*, rose in direct proportion to the number of borers present.

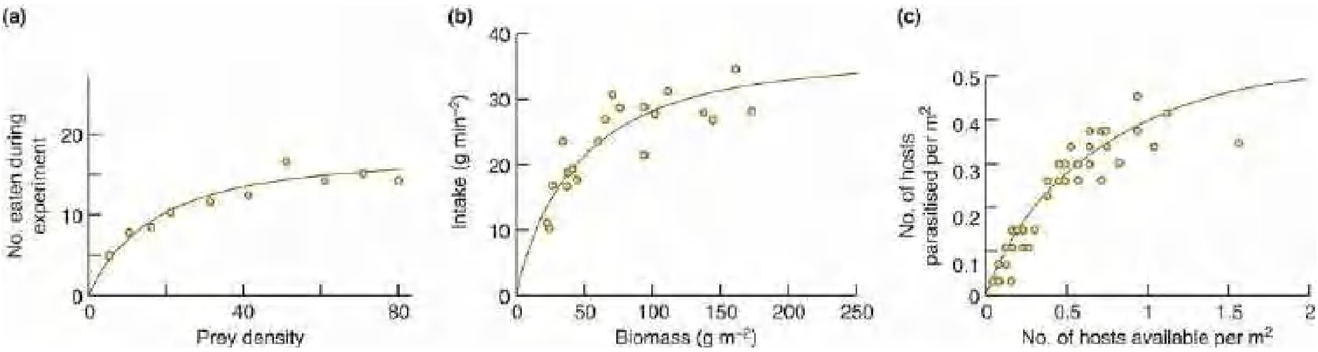


[Figure 10.3](#) **A type 1 functional response**, illustrated for *Azteca sericeasur* responding to different densities of the beetle pest, the coffee berry borer (CBB), *Hypothenemus hampei*. Bars are SEs.

Source: After Morris *et al*. (2015).

10.2.2 The type 2 functional response

The most frequently observed functional response is the 'type 2' response, in which consumption rate rises with prey density, but gradually decelerates until a plateau is reached at which consumption rate remains constant irrespective of prey density. Type 2 responses are shown for a carnivore, a herbivore and a parasitoid in Figure 10.4.



[Figure 10.4](#) **Type 2 functional responses.** (a) Tenth-instar damselfly nymphs (*Ischnura elegans*) eating *Daphnia* of approximately constant size. (b) Wood bison (*Bison bison*) feeding on the sedge *Carex atherodes* presented at a range of sedge biomass densities. (c) The parasitoid *Microplitis croceipes* attacking the tobacco budworm *Heliothis virescens*.

Source: (a) After Thompson (1975). (b) After Bergman *et al*. (2000). (c) After Tillman (1996).

the type 2 response and handling time

The type 2 response can be explained by noting that a predator has to spend time *handling* each prey item it consumes (i.e. pursuing, subduing and consuming the prey item, and then preparing itself for further search). As prey density increases, finding prey becomes increasingly easy. Handling a prey item, however, still takes the same length of time, and handling overall therefore takes up an increasing proportion of the predator's time, until at high prey densities the predator is effectively spending all of its time handling prey. The consumption rate therefore approaches and then reaches a maximum (the plateau), determined by the maximum number of handling times that can be fitted into the total time available.

With this in mind, the simple functional response of the Lotka–Volterra model, aN, can be replaced by the more realistic

$$\frac{aN}{1 + ahN} \tag{10.16}$$

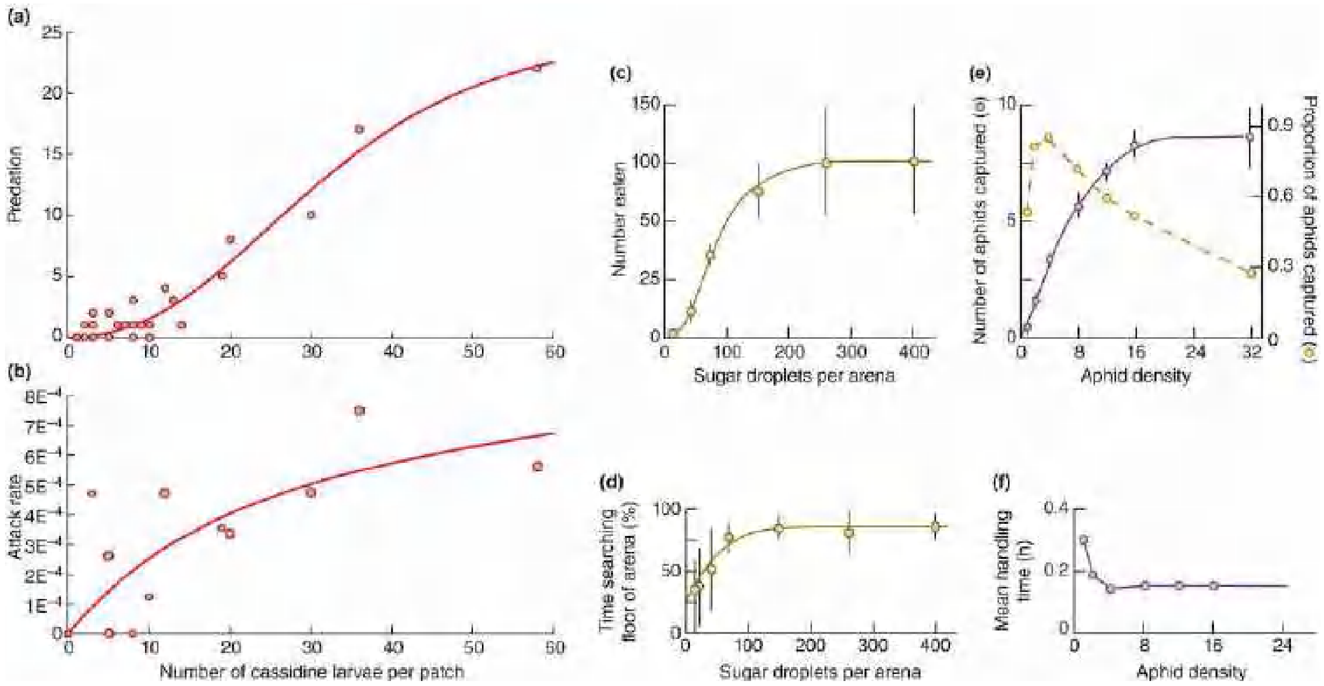
where h is the handling time. This captures the idea that as prey abundance, N, increases, the consumption rate reaches a plateau, given by the reciprocal of the handling time, $1/h$; and we can see that a type 1 functional response can be recovered from this simply by assuming that there is no handling time. A fuller derivation of the type 2 response is given by Holling (1959).

other routes to a type 2 response

On the other hand, the existence of a handling time is not the only route to a type 2 functional response, that is, a consumption rate curve that saturates as prey density increases. For instance, if the prey are of variable profitability, then at high densities the diet may tend towards a decelerating number of highly profitable items (Krebs *et al*., 1983); or a predator may become confused and less efficient at high prey densities.

10.2.3 The type 3 functional response

Finally, type 3 functional responses are illustrated in Figure 10.5. At high prey densities they are similar to a type 2 response, and the explanations for the two are the same. At low prey densities, however, the type 3 response has an accelerating phase, where an increase in density leads to a more than linear increase in consumption rate. Overall, therefore, a type 3 response is 'S-shaped' or 'sigmoidal'.



[Figure 10.5](#) Type 3 (sigmoidal) functional responses. (a) The paper wasp, *Polistes dominulus* responding to changing densities of the shield beetle, *Cassida rubiginosa*. (b) The basis of the response in (a): the attack rate of the wasp increases, though at a decelerating rate, with beetle density (larger symbols indicate multiple (2–8) data points). (c) The bluebottle fly, *Calliphora vomitoria*, feeding on sugar droplets. (d) The basis of the response in (c): searching efficiency of *C. vomitoria* increases with 'prey' (sugar droplet) density. (e) The wasp, *Aphelinus thomsoni*, attacking sycamore aphids, *Drepanosiphum platanoidis*: note the density-dependent increase in prey mortality rate at low prey densities (dashed yellow line) giving rise to the accelerating phase of the response curve (solid purple line). (f) The basis of the response in (e): handling time in *A. thomsoni* decreases with aphid density.

Source: (a, b) After Schenk & Bacher (2002). (c, d) After Murdie & Hassell (1973). (e, f) After Collins *et al.* (1981).

variations in searching efficiency or handling time

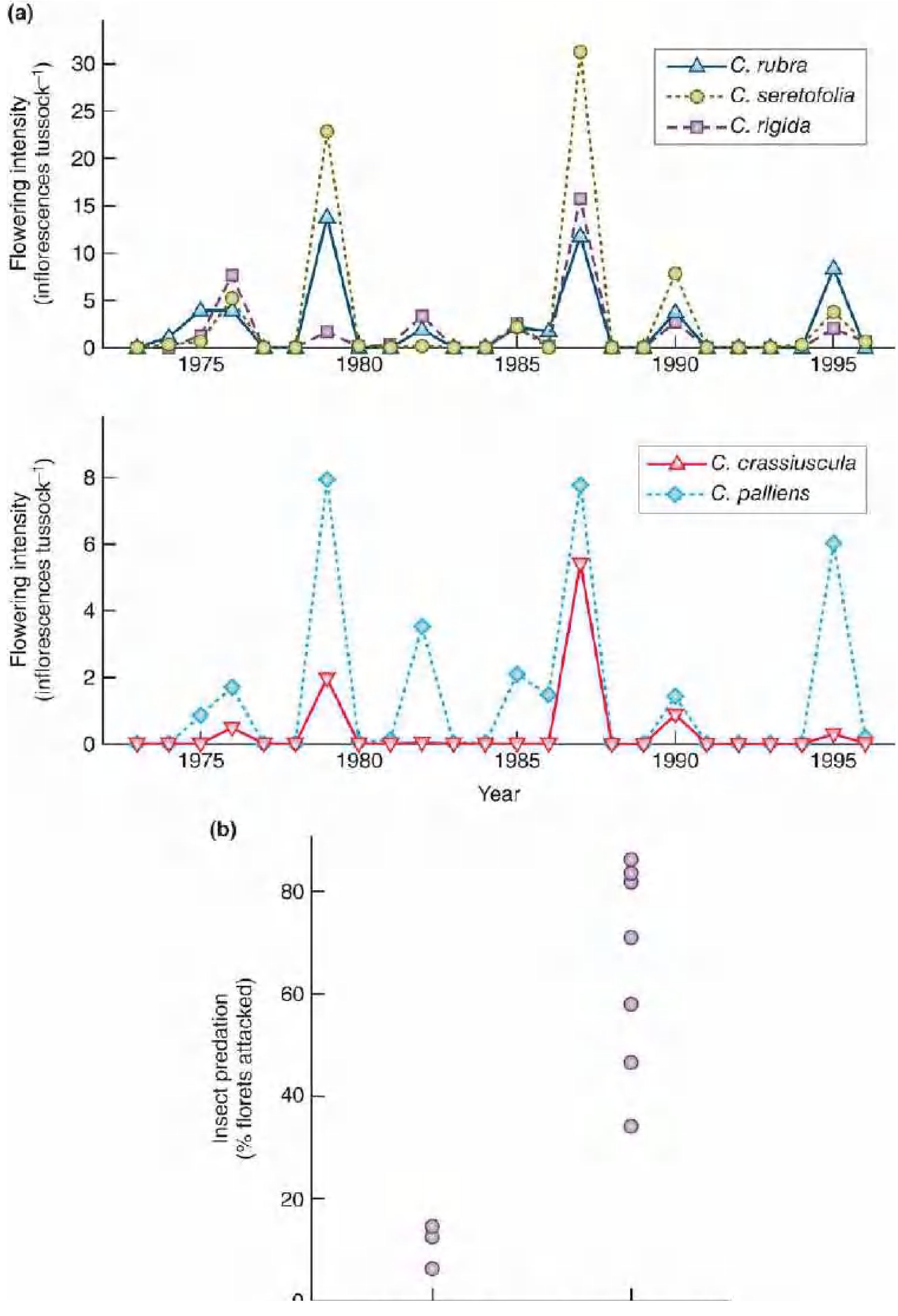
Generally, this sigmoidal shape and the type 3 functional response will arise whenever an increase in food density leads to an increase in the consumer's searching efficiency or attack rate, *a*, or a decrease in its handling time, *h*, since between them these two determine consumption rate (Equation 10.16). For example, a predator may only become interested in a particular species of prey once its relative abundance in the prey community is high enough (as we saw in Section 9.2.2 on 'switching'). Thus, the paper wasp *Polistes dominulus* (Figure 10.5a) attacks its beetle prey at an increasing rate as their density increases (Figure 10.5b). The bluebottle fly *Calliphora vomitoria* (Figure 10.5c) spends an increasing proportion of its time searching for food items as their density increases (Figure 10.5d), increasing efficiency, whilst the wasp *Aphelinus thomsoni* (Figure 10.5e) exhibits a reduction in mean handling time as the density of its sycamore aphid prey increases (Figure 10.5f). In each case, a type 3 functional response is the result

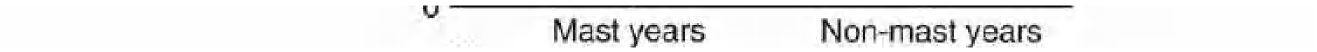
10.2.4 Individual and population-level satiation

mast years and the satiation of seed predators

One thing type 2 and type 3 functional responses have in common is a plateau at high prey densities. That is, consumers become satiated: simply incapable of eating any more. Satiation is worth a closer look in its own right.

Many plant species have ‘mast’ years. These are occasional years in which there is synchronous production of a large volume of seed, often across a large geographic area, with relatively little seed produced in the years in between (see Kelly & Sork, 2002). This is seen particularly often in tree species that suffer generally high intensities of seed predation (Silvertown, 1980), which is evolutionarily significant since the plants’ chances of escaping seed predation are likely to be much higher in mast years, partly because many individual predators may frequently be satiated, and partly because the predator population may be too small, and too slow in expanding, to fully exploit the glut of seeds available. An example of masting, not for trees but for five species of the tussock grass *Chionochloa*, including *C. pallens*, is shown in Figure 10.6a. The responses of seed predators of *C. pallens* are then shown in Figure 10.6b. The percentage of florets of *C. pallens* attacked by insects remained below 20% in mast years but went up to 80% or more in non-mast years. The fact that *C. pallens* and four other species show strong synchrony in masting further increases the benefit to each species in terms of escaping seed predation in mast years.


(a)
Flowering intensity (inflorescences tussock^{-1})
30
25
20
15
10
5
0
C. rubra
C. seretofolia
C. rigida
1975
1980
1985
1990
1995
Flowering intensity (inflorescences tussock^{-1})
8
6
4
2
0
C. crassiuscula
C. pallens
1975
1980
1985
1990
1995
Year
(b)
Insect predation (% florets attacked)
80
60
40
20
0



[Figure 10.6](#) **Masting in grasses and its negative effect on predation.** (a) The flowering rate for five species of tussock grass (*Chionochloa*) between 1973 and 1996 in Fiordland National Park, New Zealand. Mast years are highly synchronised in the five species, seemingly in response to high temperatures in the previous season, when flowering is induced. (b) Insect predation on florets of *C. pallens* in mast ($n = 3$) and non-mast years ($n = 7$) from 1988 to 1997 at Mount Hutt, New Zealand. A mast year is defined here as one with greater than 10 times as many florets produced per tussock than in the previous year.

Source: After McKone *et al.* (1998).

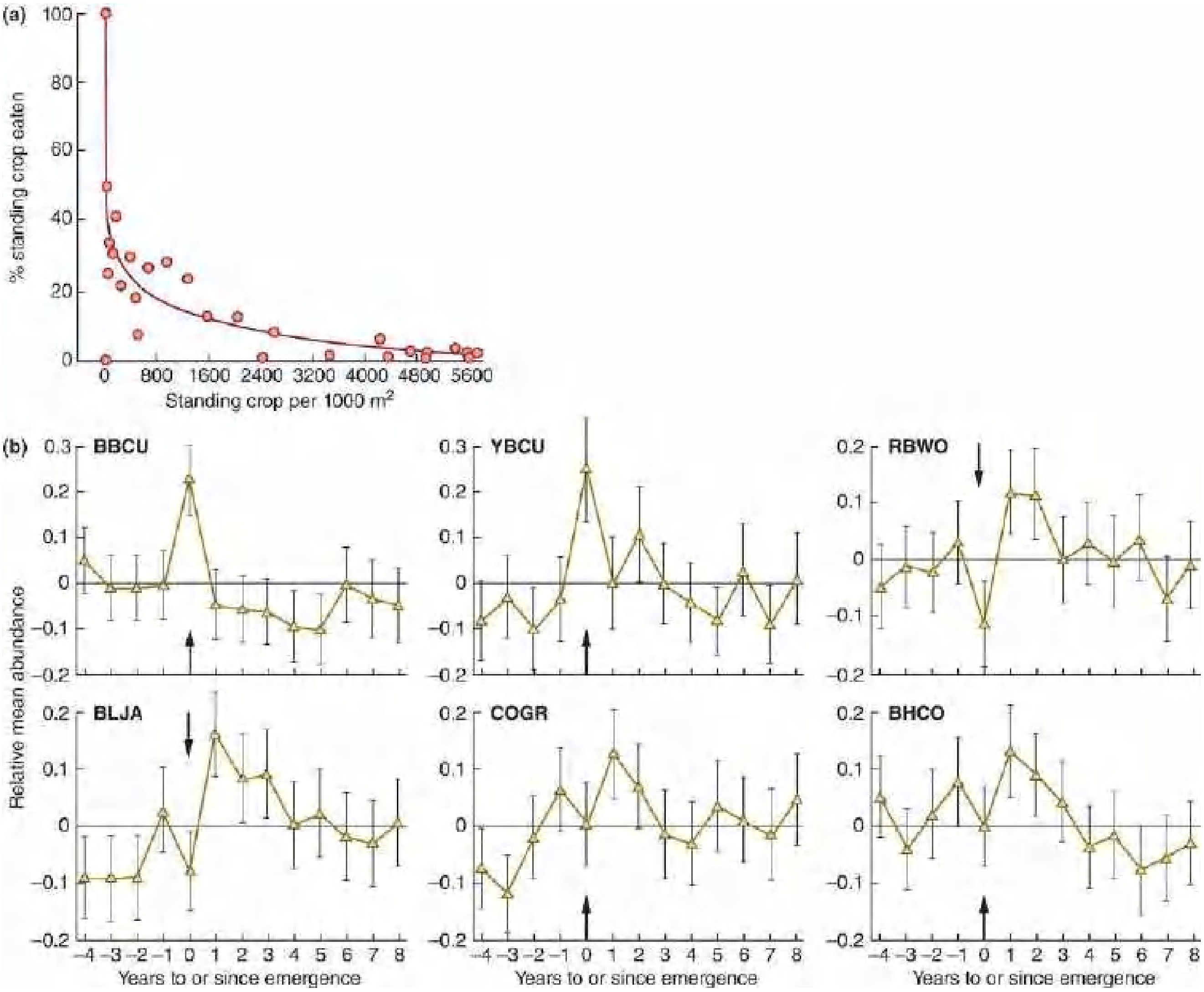
On the other hand, the production of a mast crop makes great demands on the internal resources of a plant. A spruce tree in a mast year averages 38% less annual growth than in other years, and the annual ring increment in forest trees may be reduced by as much during a mast year as by a heavy attack of defoliating caterpillars. The years of seed famine are therefore essentially years of plant recovery.

satiation and delayed density dependence ...

At the population level, masting returns us to the question of delayed density dependence in the numerical response of predators to prey, discussed previously, and highlights the particular importance of generation times. The seed predators are unable to do the maximum harm to the mast crop because their generation times are too long. A hypothetical seed predator population that could pass through several generations during a season would be able to increase exponentially on the mast crop and destroy it.

... as illustrated by periodical cicadas

We see something very similar in the remarkable case of ‘periodical’ cicadas (Figure 10.7). These insects (*Magicicada* spp.) emerge from their nymphal stages in the ground only very rarely, and synchronously – in fact, every 13 or 17 years, depending on the species. The adults then climb into the canopy, feed, mate and oviposit within three to four weeks, following which first instar nymphs hatch and return to the soil, for the whole 13- or 17-year cycle to start again. The unpredictability of mast seeding has therefore been replaced by a life cycle of predictable length, but one that, over evolutionary time, no predator has matched. Here, the fact that 13 and 17 are prime numbers is, perhaps, significant. A hypothetical ‘12-year cicada’ could be at least partially tracked by predators with two-, three-, four- or six-year life cycles. But for the real thing, it is 13 or 17 years – or nothing. Hence, when they did emerge at a study site in north-western Arkansas, the satiation of avian predators was apparent in the high proportion of them eaten before and after peak emergence, when their numbers were low, dropping to a very low proportion when their numbers were high around that peak (Figure 10.7a). And on a broader scale, throughout hardwood forests in the eastern USA, many species of birds experienced a peak in abundance one or two years *after* the cicada emergence (Figure 10.7b). A few others had local peaks in the year corresponding with the cicada emergence, but these were migrants attracted to the rare food glut, not local species increasing immediately in abundance through breeding (Figure 10.7b).



[Figure 10.7](#) Periodical cicadas satiate their predators and so avoid high rates of predation. (a) Percentage of periodical cicadas eaten by avian predators in relation to their abundance (standing crop). ($y = 74.6 - 19.6\log(x)$; $r^2 = 0.51$.) (b) Annual standardised abundance relative to their long-term mean (mean and 95% CIs), in relation to their long-term mean, in relation to the year of cicada emergence, for six bird species in the eastern USA. Two were migrants that increased in abundance in the same year by being attracted to the food glut: the black-billed cuckoo, *Coccyzus erythropthalmus* (BBCU) and the yellow billed cuckoo, *C. americanus* (YBCU). The other four increased in abundance through local increases in survival and/or birth rate, but only after a delay: the red-bellied woodpecker, *Melanerpes carolinus* (RBWO), the blue jay, *Cyanocitta cristata* (BLJA), the common grackle, *Quiscalus quiscula* (COGR) and the brown-headed cowbird, *Molothrus ater* (BHCO).

Source: (a) After Williams et al. ([1993](#)). (b) After Koenig & Liebhold ([2005](#)).

10.2.5 Food quality

food quality rather than quantity can be of paramount importance

Having focused here on the amount of food consumed, it is worth reminding ourselves of what was stressed in [Chapter 3](#): that the quantity of food consumed may be less important than its quality, which itself has both positive aspects (like the concentrations of nutrients) and negative aspects (like the concentrations of toxins). This is especially the case for herbivores. We can see the problem if we imagine ourselves provided with a perfectly balanced diet – diluted in an enormous swimming pool. The pool contains everything we need, and we can see it there before us, but we may starve to death before we can drink enough water to extract enough nutrients to sustain ourselves. In a similar fashion, herbivores may frequently be confronted with a pool of available nitrogen that is so dilute that they have difficulty processing enough material to extract what they need. Outbreaks of herbivorous insects, for example, may then be associated with rare elevations in the concentration of available nitrogen in their food plants, perhaps associated with unusually dry or, conversely, unusually waterlogged conditions (White, 2008).

For instance, in the case of cotton rats, *Sigmodon hispidus*, in Oklahoma, USA, analysis of their plant diets in several high and low density populations showed, first, that those faring best, i.e. those in the high density populations, consumed higher concentrations of a range of amino acids in the breeding season, the key phase for population recruitment. One example is shown in Figure 10.8a, for the amino acids methionine and cysteine (methionine is essential but in practice, cysteine may supply up to 50% of an animal's methionine requirement). In addition, though, those in the high-density populations were consistently consuming food with lower concentrations of phenolics, which can be both toxic and inhibitors of protein digestion (Figure 10.8b).

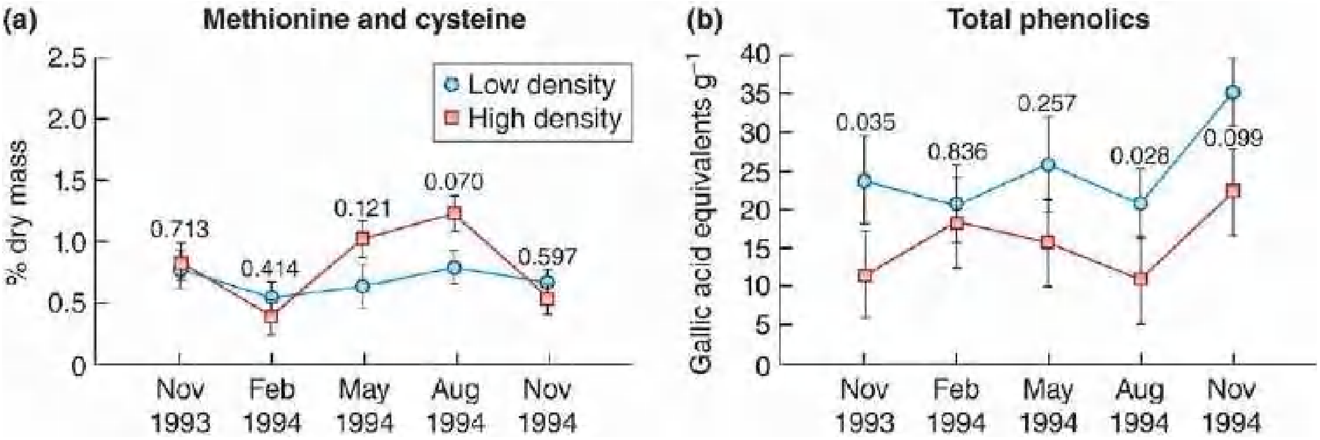


Figure 10.8 The composition of the food of cotton rats, *Sigmodon hispidus*, in Oklahoma is significantly better in high- than in low-density populations. (a) Concentrations of two key amino acids, methionine and cystine, were higher in the high-density populations throughout the summer breeding period. (b) Concentrations of phenolics, acting either as toxins or as protein digestive inhibitors, were consistently higher in the low-density populations. In both, bars are SEs and numbers are *P*-values for comparisons of concentrations.

Source: After Schetter *et al.* (1998).

In the case of the positive aspects of this, this association was further supported by experimental manipulation. In field enclosures, animals had their natural food supply supplemented either with additional general rations or with this plus a further supplementation of methionine (Table 10.1). Density, survival and recruitment were all enhanced by the provision of additional food; but density and recruitment were enhanced further still by the addition of methionine. In the case of herbivores especially, how much an animal eats may be less important than what it eats.

Table 10.1 **Cotton rats benefit from an improved quality of their diet.** The responses of cotton rats, *Sigmodon hispidus*, in field enclosures in Oklahoma, to additional food, and to additional food supplemented with the essential amino acid methionine, relative to unsupplemented controls.

Source: After Webb *et al*. (2005).

Demographic variable	Mixed ration versus control	Methionine versus control	Methionine versus mixed ration
Overall density	+	+	+
Female density	+	+	0
Male density	+	+	0
Overall survival	+	+	0
Female survival	+	+	0
Male survival	0	0	0
Recruitment	0	+	+

10.2.6 The effects of conspecifics – interference and ratio-dependent predation

One obvious omission from the predator–prey interactions in our basic models was any acknowledgement that prey abundance may be limited by other prey, and predator abundance by other predators. Prey are bound to be increasingly affected by intraspecific competition as their abundance increases. Predators, too, are likely to be competing for, not simply preying upon, their prey, as well as being limited at high densities by the availability of resting places or safe refuges of their own.

mutual interference

More specifically, predators have been assumed in the models discussed so far to consume prey at a rate that depends only on prey abundance. In the Lotka–Volterra model, for example, the consumption rate per predator is simply aN, and consumption rate with a type 2 functional response is $aN/(1 + ahN)$. Typically, however, consumption rate will also often depend on the abundance of the predators themselves. Most obviously, food shortage – the abundance of prey *per predator* – will commonly result in a reduction in the consumption rate per individual as predator density increases. However, even when food is not limited, the consumption rate can be reduced by a number of processes known collectively as mutual interference (Hassell, 1978). For example, many consumers interact behaviourally with other members of their population, leaving less time for feeding and therefore depressing the overall feeding rate. For instance, hummingbirds actively and aggressively defend rich sources of nectar and spend more time doing so the more competitor hummingbirds there are. Alternatively, an increase in consumer density may lead to an increased rate of emigration, or of consumers stealing food from one another, as do many gulls. All of these mechanisms give rise to a decline in predator consumption rate with predator density. Figure 10.9a, for example, shows this when the crab *Carcinus aestuarii* forages for the mussel *Musculista senhousia*. Figure 10.9b shows that the kill rate of wolves, *Canis lupus*, preying on moose, *Alces alces*, in Isle Royale National Park, Michigan, USA, was lowest when there were most wolves.

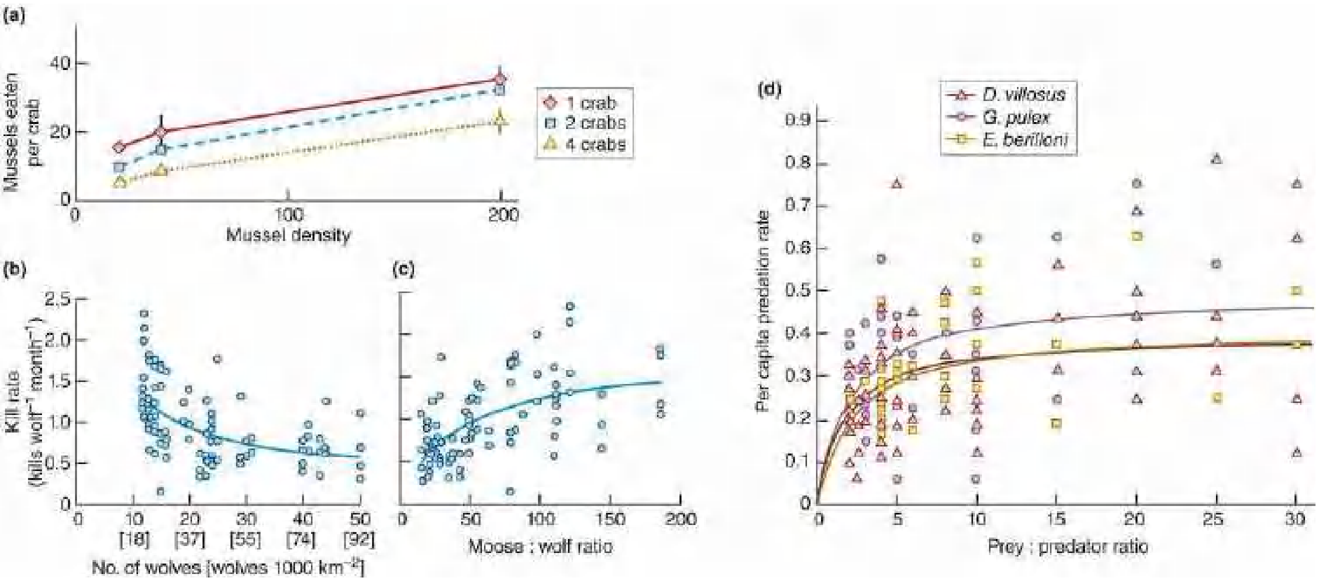


Figure 10.9 Mutual interference leads to a reduction in predation rates with predator density. (a) Mutual interference amongst crabs, *Carcinus aestuarii*, feeding on mussels, *Musculista senhousia*. The more crabs there were, the lower their per capita consumption rate. (b) Mutual interference amongst wolves, *Canis lupus*, preying on moose, *Alces alces*. (c) The same data but with wolf kill rate plotted against the moose : wolf ratio. The fitted curve assumes that the kill rate depends on this ratio, but also that the wolves may become 'saturated' at high moose densities (see Section 10.2.2). This curve fits better than any for which kill rate depends on either predator density (e.g. (b)) or prey density. (d) Relationship between the number of fly larvae eaten by three freshwater amphipods, *Dikerogammarus villosus* (red curve and triangles), *Gammarus pulex* (purple curve and dots) and *Echinogammarus berilloni* (yellow curve and squares), and the prey to predator ratio. All three curves were significantly better at fitting the data than those dependent on either prey or predator abundance alone.

Source: (a) After Mistri (2003). (b, c) After Vucetich *et al.* (2002). (d) After Médoc *et al.* (2015).

intimidation effects

We noted in Sections 9.2.4 and 9.6 that prey are often 'intimidated' by predators into altering their behaviour or even their morphology, both of which may serve to protect the prey from predation, at least partially. We noted, too, that if this is costly to the prey then it may lead to reductions in their survival, their fecundity and, potentially, their abundance. We can also see now that if these intimidatory effects tend to increase with predator density, as they are likely to do, then the consequences will be the same as we have seen for interference between predators – lower consumption rates at higher consumer densities.

ratio dependent predation

Arguably the simplest way of incorporating these ideas is to abandon altogether the assumption that consumption rate depends only on the absolute availability of prey, assuming instead that it depends on the ratio of prey to predators (i.e. on the average number of prey available *per predator*). The 'prey-dependent' predation of the Lotka–Volterra model is then replaced by *ratio-dependent predation* (Arditi & Ginzburg, 1989 , 2012). Evidence of ratio-dependent predation is illustrated, for example, for the wolf–moose study in Figure 10.9c, and for freshwater amphipods feeding on fly larvae in Figure 10.9d. In both cases, statistical models in which predation rate depended on the predator–prey ratio fitted the data far better than those in which the dependence was on either prey or predator numbers alone. On the other hand, ratio-dependent predation, just like prey-dependent predation, is an ideal, and cannot, therefore, be the whole story. It is unlikely that consumption rate depends only on prey numbers when the numbers of predator are high; but it is equally unlikely that consumption rate depends strictly on the predator/prey ratio when predator numbers are low and they do not compete with or interfere with one another. The question here is not which of the alternatives is 'true' but which is the more useful approximation and starting point for further development. The impassioned advocacy by Arditi and Ginzburg (2012) on behalf of ratio dependence has been opposed by an equally impassioned rebuttal from Abrams (2015).

10.3 The population dynamics of interference, functional responses and intimidation: equations and isoclines

Now, in this section, we can begin to incorporate some of the ideas from the previous section into the simple models from Section 10.1.

10.3.1 The population dynamics of interference

We begin with interference, taking the type 2 functional response (Equation 10.16) and replacing prey-dependent with ratio-dependent predation as a simple way of incorporating predator interference. The equation then becomes

$$\frac{\alpha N/P}{1+\alpha hN/P} = \frac{\alpha N}{P+\alpha hN}, \tag{10.17}$$

where the a has been replaced by an α to acknowledge that the units of the prey's attack rate are no longer the same. Then, if we combine this with intraspecific competition (interference) among the prey, modelled as a logistic equation (Section 5.7), we have a new set of predator–prey equations:

$$\frac{dN}{dt} = rN\left(\frac{K-N}{K}\right) - \frac{\alpha N}{P+\alpha hN} \tag{10.18}$$

$$dP/dt = f\frac{\alpha N}{P+\alpha hN}P - qP. \tag{10.19}$$

The behaviour of this and other variants of the Lotka–Volterra model (and indeed of the Nicholson–Bailey model) can be analysed mathematically (see, for example, Arditi & Ginzburg, 2012). As previously, however, we can explore this behaviour, and its dependence on the various biological realities we have now introduced, by looking directly at the zero isoclines that the equations generate. That is, we can draw up a catalogue of isoclines reflecting the various realities of functional responses, predator interference and so on, and then bring these together to explore how they might influence the dynamics of predator and prey populations.

a prey isocline with intraspecific competition

Starting with the prey (Figure 10.10a), we can assume that at low prey densities there is no intraspecific competition, and the prey isocline is therefore horizontal as in the Lotka–Volterra model (Figure 10.1). But as density increases, it is more and more the case that prey densities that would otherwise allow the prey to increase must be placed on the decrease side of the isocline because of the effects of intraspecific competition. Hence, the isocline increasingly 'bends' until it reaches the prey axis at the carrying capacity, K_N; that is, where the prey can only just maintain themselves even in the absence of predators.

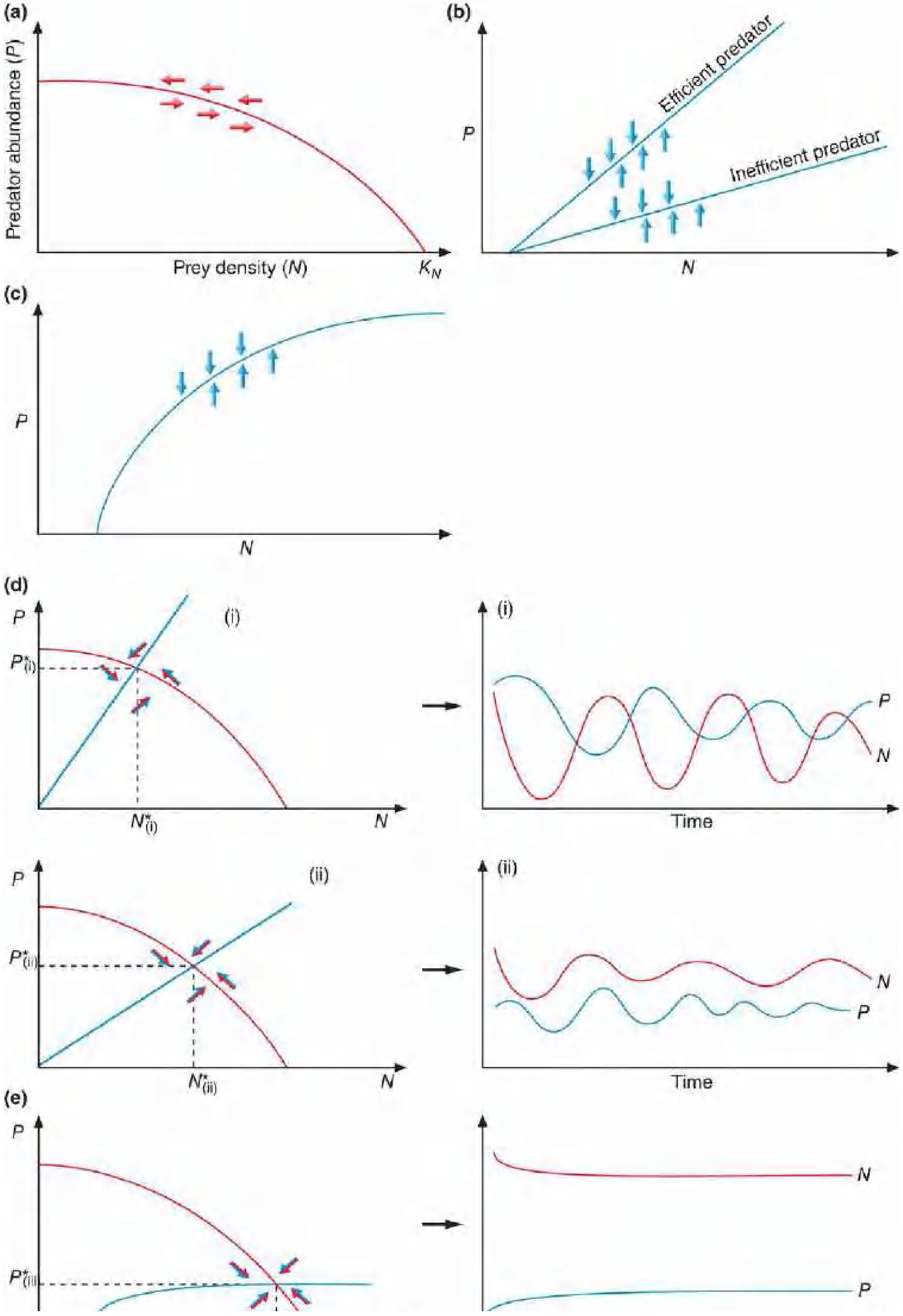

(a)
Predator abundance (P)
Prey density (N)
K_N
(b)
P
N
Efficient predator
Inefficient predator
(c)
P
N
(d)
P
(i)
$P^*_{(i)}$
$N^*_{(i)}$
N
(i)
P
N
Time
P
(ii)
$P^*_{(ii)}$
$N^*_{(ii)}$
N
(ii)
N
P
Time
(e)
P
$P^*_{(iii)}$
N
P

Figure 10.10 Prey and predator zero isoclines that incorporate crowding, and their effects on predator–prey dynamics. (a) A prey zero isocline subject to crowding. At the lowest prey densities this is the same as the Lotka–Volterra isocline, but when the density reaches the carrying capacity (K_N) the population can only just maintain itself even in the complete absence of predators. Arrows show combinations of predator and prey abundance where prey increase or decrease, as indicated. (b) Predator zero isoclines when there is ratio dependent predation. More efficient predators require smaller numbers of prey per predator to sustain them. Arrows show combinations of predator and prey abundance where predators increase or decrease, as indicated. (c) A predator zero isocline subject to crowding (see text). (d) The prey zero isocline combined with the predator zero isoclines with ratio dependent predation. P^* is the equilibrium abundance of predators, and N^* the equilibrium abundance of prey. Combination (i) is the least stable (most persistent oscillations) and has the most predators and fewest prey: the predators are relatively efficient. Less efficient predators, as in (ii), give rise to a lowered predator abundance, an increased prey abundance and less persistent oscillations. (e) With strong predator self-limitation (see (c)) oscillations can be eliminated altogether, but P^* tends to be low and N^* close to K_N.

a ratio-dependent isocline

For the predators, it is perhaps simplest to assume ratio-dependent predation, such that the predator zero isocline is no longer vertical as it was in Figure 10.1, where a threshold number of prey needs to be exceeded for predator numbers to increase, but a diagonal passing through the origin (Figure 10.10b) such that a predator–prey ratio needs to be exceeded.

a predator isocline with mutual interference

Alternatively, if we start with the Lotka–Volterra vertical isocline and incorporate mutual interference amongst the predators, then as this interference increases with predator density, individual consumption rates will decline, and additional prey will be required to maintain a predator population of any given size. The predator zero isocline will therefore depart increasingly from the vertical. Indeed, at high densities, competition for other resources may put an upper limit on the predator population (a horizontal isocline) irrespective of prey numbers (Figure 10.10c).

putting the isoclines together: crowding stabilises dynamics

The likely effects of prey crowding, and interference or ratio-dependent predation, can now be deduced by combining the predator and prey isoclines (Figure 10.10d and e). Oscillations are still apparent for the most part, but these are no longer neutrally stable. Instead, they are damped so that they converge to a stable equilibrium. Predator–prey interactions in which either or both populations are substantially self-limited are likely, therefore, to exhibit patterns of abundance that are relatively stable (i.e. in which fluctuations in abundance are relatively slight).

Furthermore, we can see that when the predator is relatively inefficient, or in other words when many prey are needed to maintain a population of predators (curve (ii) in Figure 10.10d), the oscillations are damped quickly but the equilibrium prey abundance (N^*) is not much less than the equilibrium in the absence of predators (K_N). By contrast, when the predators are more efficient (curve (i)), N^* is lower and the equilibrium density of predators, P^*, is higher – but the interaction is less stable (the oscillations are more persistent). Moreover, if the predators are very strongly self-limited (Figure 10.10e), then abundance may not oscillate at all; but P^* will tend to be low, whilst N^* will tend to be not much less than K_N.

the suppression–stability trade-off

Hence, for interactions where there is crowding, there appears to be a contrast between, on the one hand, those in which predator density is low, prey abundance is little affected and the patterns of abundance are stable, and then on the other, those in which predator density is higher and prey abundance is more drastically reduced, but the patterns of abundance are less stable. Other models, too, show this tendency for the stability of the interaction to be greatest when there is least suppression of prey abundance. Murdoch *et al.* (2003) refer to this as 'the suppression–stability trade-off'. As we discuss further below, this makes it particularly important to discover how stability might nonetheless be combined with profound prey suppression, particularly in the context of biological control (Section 15.2.4), where the central aim is just that: to reduce pest abundance to a level where its pest status is lost and to keep it there.

To quote examples of data proving the stabilising influence of self-limitation on predator–prey dynamics is difficult, simply because it is all but impossible to compare the dynamics of matched populations with and without such self-limitation. On the other hand, populations of predators and prey with relatively stable dynamics are commonplace, as are the stabilising forces of self-limitation. To take a more specific example, there are two groups of primarily herbivorous rodents that are widespread in the Arctic: the microtine rodents (lemmings and voles) and the ground squirrels. The microtines are renowned for their dramatic, cyclic fluctuations in abundance (see Chapter 14), but the ground squirrels have populations that remain remarkably constant from year to year, especially in open meadow and tundra habitats. There, significantly, they appear to be strongly self-limited by food availability, suitable burrowing habitat and their own spacing behaviour (Karels & Boonstra, 2000).

how important is mutual interference in practice?

On a cautionary note, however, Umbanhowar *et al.* (2003), for example, failed to find evidence of mutual interference in a field study of the parasitoid *Tachinomyia similis* attacking its moth host *Orgyia vetusta*. And more generally, the strength of mutual interference may often have been exaggerated by forcing predators to forage in artificial arenas at densities much higher than those they experience naturally. This is a useful reminder of the general point that an ecological force that is powerful in models or in the laboratory may nonetheless often be trivial, in practice, in natural populations. There can be little doubt, though, that self-limitation in its various forms frequently plays a key role in shaping predator–prey dynamics.

10.3.2 The population dynamics of functional responses

type 3 responses stabilise but may be unimportant in practice

Different types of functional response have different effects on population dynamics, potentially at least. A type 3 response means a low predation rate at low prey densities. In terms of isoclines, this means that prey at low densities can increase in abundance virtually irrespective of predator density, and that the prey zero isocline will therefore rise vertically at low prey densities (Figure 10.11a). This has the potential to stabilise an interaction (Figure 10.11a, curve (i)), but for this the predator would have to be readily capable of maintaining itself at low prey densities, which seems to contradict the whole idea of a type 3 response (ignoring prey at low densities). Hence, curve (ii) in Figure 10.11a is likely to apply, and the stabilising influence of the type 3 response may in practice be of little importance. On the other hand, if a predator has a type 3 response to one particular type of prey because it switches its attacks amongst various prey types, in other words, if it is a generalist predator, then the population dynamics of the predator would be independent of the abundance of any particular prey type, and the vertical position of its zero isocline would therefore be the same at all prey densities. As Figure 10.11b shows, this can lead potentially to the predators regulating the prey at a low and stable level of abundance.

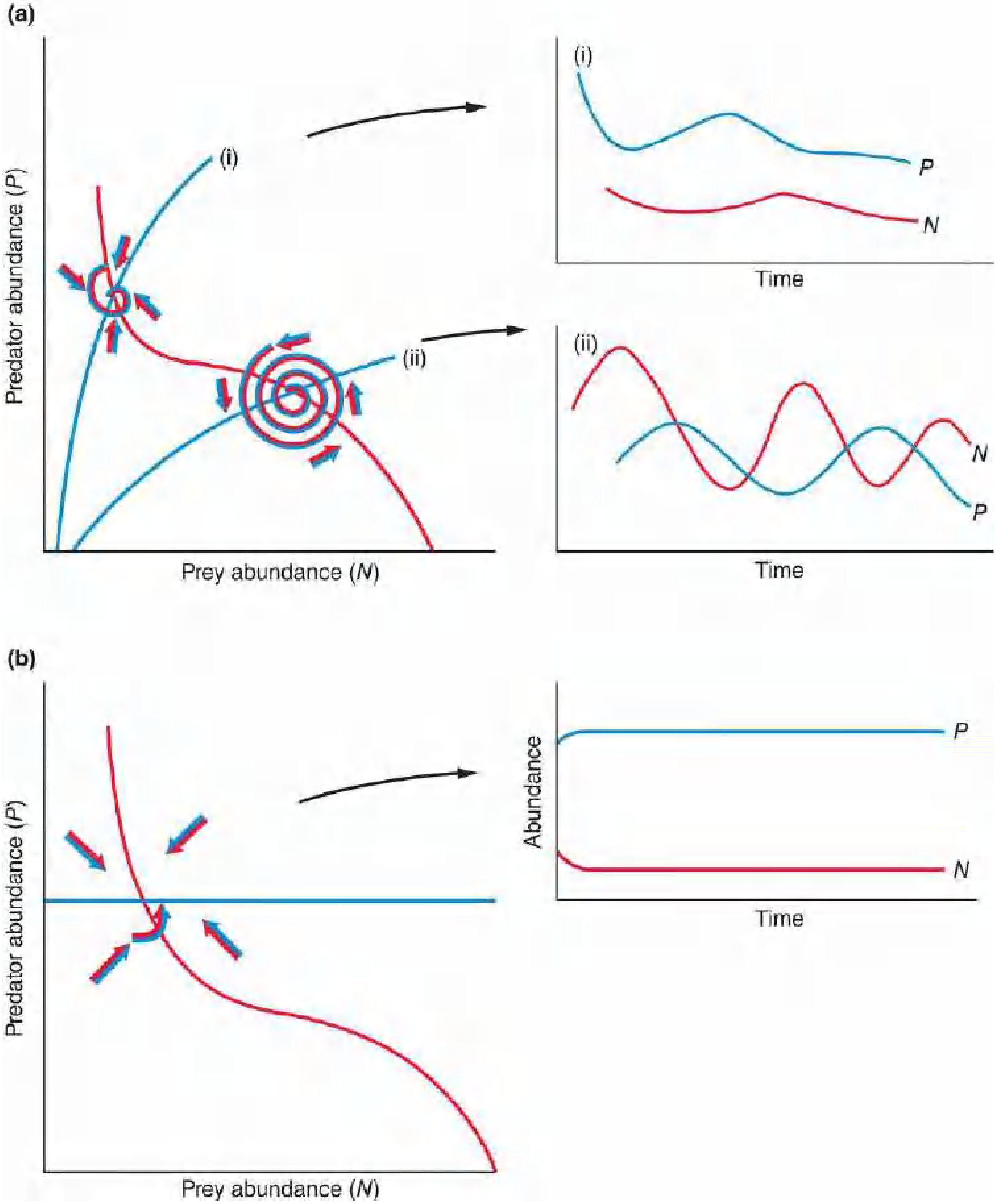

(a)
Predator abundance (P)
Prey abundance (N)
(i)
(ii)
(i)
P
N
Time
(ii)
N
P
Time
(b)
Predator abundance (P)
Prey abundance (N)
Abundance
P
N
Time

Figure 10.11 Prey and predator zero isoclines that incorporate a type 3 functional response, and their effects on predator–prey dynamics. (a) The prey zero isocline (red line, left-hand panel) is that which is appropriate when consumption rate is particularly low at low prey densities, perhaps because of a type 3 functional response, or because of a prey refuge or a reserve of plant material that is not palatable. In each case, a small population of prey can increase in abundance irrespective of the number of predators. With a relatively inefficient predator, predator zero isocline (ii) is appropriate and the outcome is not dissimilar from Figure 10.10. However, a relatively efficient predator will still be able to maintain itself at low prey densities. Predator zero isocline (i) will therefore be appropriate, leading to a stable pattern of abundance in which prey density is well below the carrying capacity and predator density is relatively high. (b) When a type 3 functional response arises because the predator exhibits switching behaviour, the predator's abundance may be independent of the density of any particular prey type (main figure), and the predator zero isocline may therefore be horizontal (unchanging with prey density). This can lead to a stable pattern of abundance (right) with prey density well below the carrying capacity.

switching, generalist predators and stability in Fennoscandia

Klemola *et al.* (2002) argued that such switching by generalist predators, and the consequent type 3 responses, accounted for parallel trends in the stability of dynamics of two groups of herbivores in Fennoscandia, in the far north of Europe: the autumnal moth, *Epirrita autumnata*, and voles in the genera *Microtus* and *Myodes*, though the latter was called *Clethrionomys* at the time (Figure 10.12). In both cases, fluctuations were far more pronounced in the north than the south. In both cases, specialist predators linked directly to their prey are believed to be instrumental in generating those fluctuations. But in both cases, there are many more generalist predators in the south, where the dynamics are more stable, than in the north. In fact, in the case of the vole cycles, Hanski *et al.* (1991 ; see also Section 14.6.3) constructed a simple model supporting this idea, in which prey (voles) interacted with specialist predators (mustelids: stoats and weasels) and generalist and switching predators (including red foxes, badgers, cats, buzzards, owls and kestrels). As the number of generalist predators increased, oscillations in vole and mustelid abundance decreased in length and amplitude. Large enough densities of switching generalists stabilised the cycle entirely.

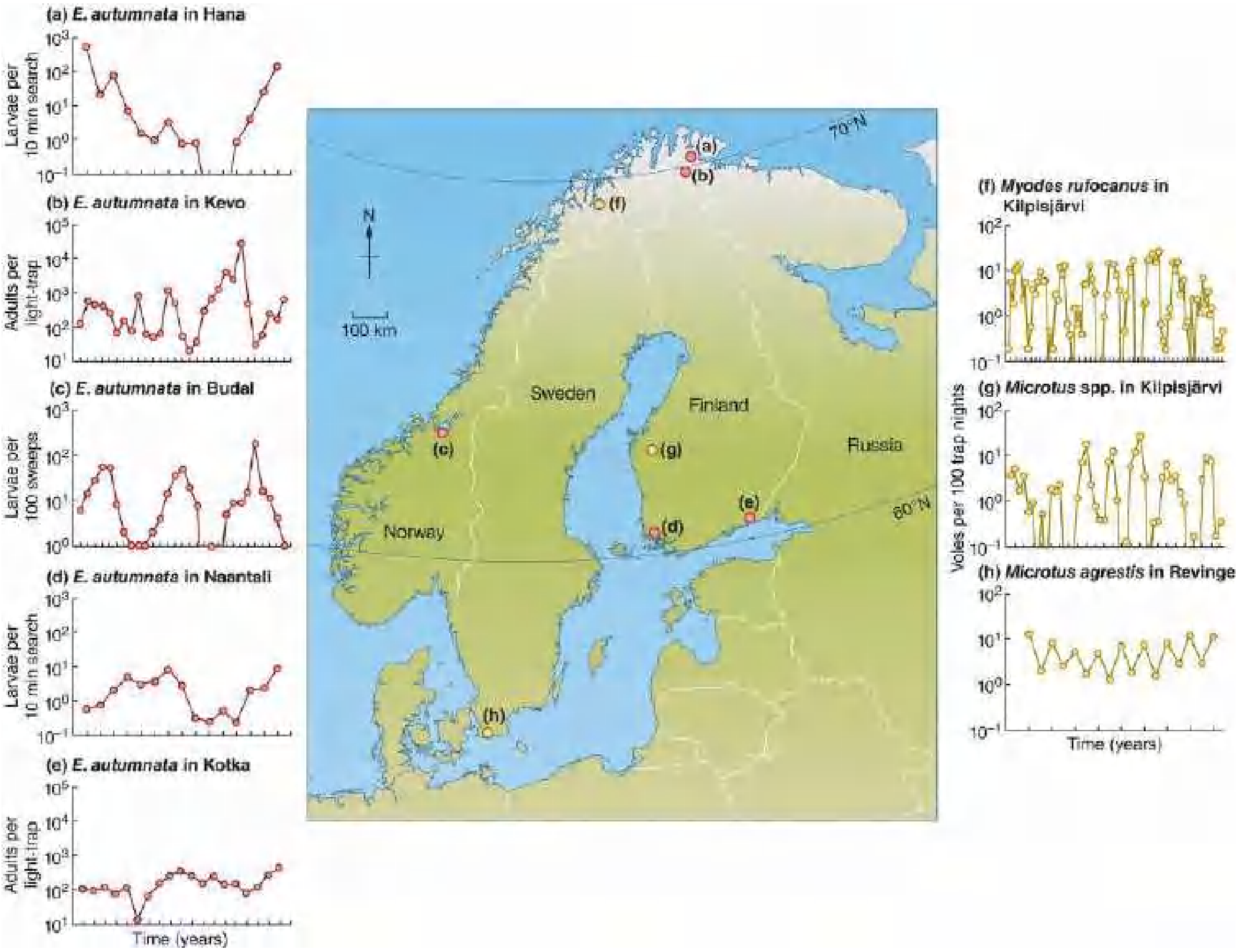


Figure 10.12 **Population fluctuations of both moths and voles are more pronounced in northern than in southern Fennoscandia.** Population fluctuations for the autumnal moth, *Epirrita autumnata*, and voles in the genera *Microtus* and *Myodes* (previously *Clethrionomys rufocanus*), at various sites in Fennoscandia. For details of individual sites and datasets (years covered, etc.), see the original article.

Source: After Klemola *et al*. (2002).

APPLICATION 10.1 Generalist, switching predators as effective biocontrol agents?

One context in which a stable low abundance of prey is positively desirable is when the prey species is a pest and the predator a biological control agent – a natural enemy of the pest employed to keep the pest in check (see Section 15.2.4). Symondson *et al*. (2002) reviewed 181 studies in which the effectiveness of a generalist predator, with its potential for switching, was tested by manipulating its abundance in the field and monitoring either any subsequent changes in pest abundance, or in the abundance of, or damage to, the crop on which the pest was feeding (Table 10.2). Success rate was high – around 75% or greater, whether studies involved a single generalist, an assemblage of generalists or an assemblage that included generalists. Switching, generalist predators certainly seem capable of holding their prey down at stable low abundances.

Table 10.2 Generalist predators can be effective in pest control. A summary of manipulative field studies that monitored the success of generalist predators used in biological control in reducing the abundance of their prey. 'Generalists' were defined, conservatively, as species preying on more than one Order of pest. Data were collated from studies between 1960 and 2001 from 22 countries and 42 crops and included at least 56 species of generalist predators. 'Others' were arthropod pests in the Thysanoptera, Orthoptera or Diplopoda.

Source: After Symondson *et al*. (2002).

	Percentage successful (no. of studies in brackets)			
Pest group	**Single species of generalist predator**	**Assemblages of generalist predators**	**Assemblages of natural enemies, including generalist predators**	**Totals**
Mollusca	89 (9)	100 (1)	100 (2)	92 (12)
Acari	79 (14)	75 (4)	100 (1)	79 (19)
Coleoptera	89 (9)	100 (4)	100 (1)	93 (14)
Diptera	70 (10)	43 (7)	50 (2)	58 (19)
Hemiptera	68 (31)	78 (18)	86 (14)	75 (63)
Lepidoptera	71 (17)	92 (12)	100 (7)	83 (36)
Others	67 (72)	83 (6)	–	72 (18)
Totals	74 (102)	79 (52)	89 (27)	77 (181)

type 2 responses (and the Allee effect) destabilise

Turning to type 2 responses, if the predator has a response that reaches its plateau at relatively low prey densities (well below K_N), then the prey zero isocline may rise initially as prey density increases, because the predators are becoming less efficient, but then fall eventually as the effects of competition amongst the prey become apparent. The prey isocline may therefore have a 'hump' overall. A hump will also arise if the prey are subject to an 'Allee effect', where they have a disproportionately low rate of recruitment when their own density is low, perhaps because mates are difficult to find or because a 'critical number' must be exceeded before a resource can be properly exploited (i.e. there is inverse density dependence at low population densities) (Lidicker, 2010). If the predator isocline crosses to the right of the hump, then the population dynamics of the interaction will be little affected; but if the isocline crosses to the left of the hump, then the outcome will be persistent rather than convergent oscillations, and the interaction could be *de*stabilised (Figure 10.13).

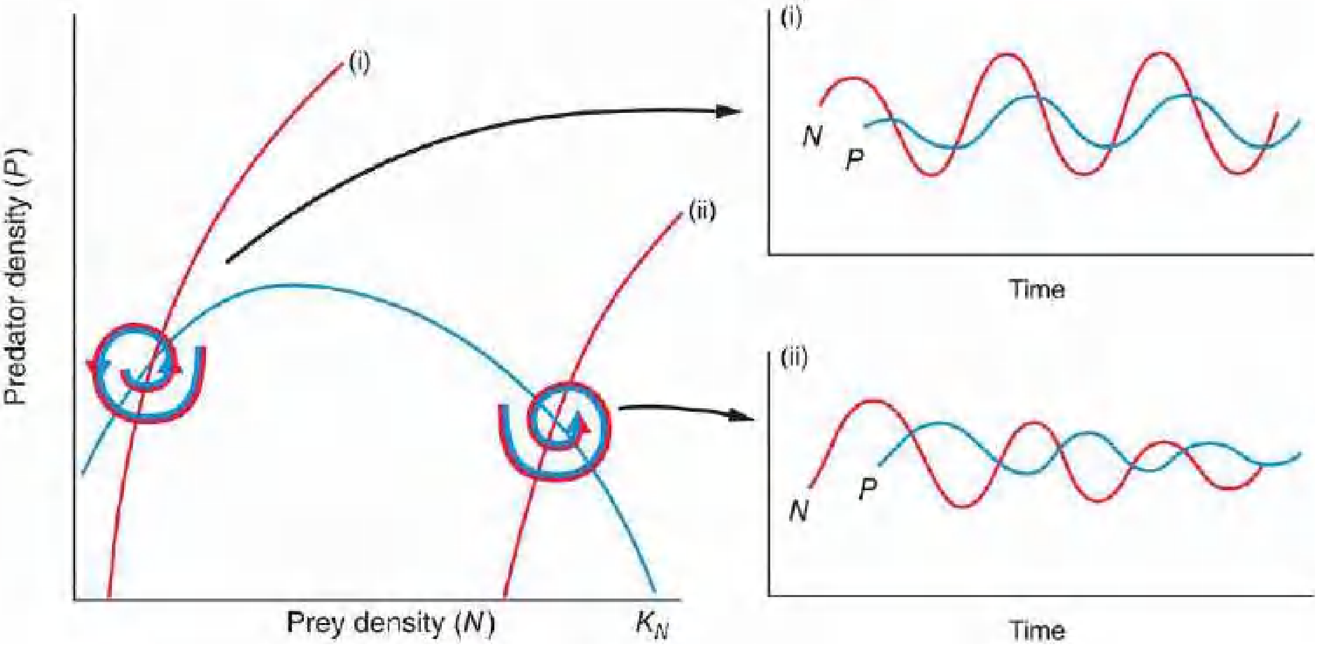


Figure 10.13 Possible effects of a 'humped' prey isocline, either as a result of a type 2 functional response or an Allee effect. If the predator is highly efficient, with its isocline crossing to the left of the hump, then the hump can be destabilising, leading to the persistent oscillations of a limit cycle (graph (i)). But if the predator is less efficient, crossing to the right of the hump, then the hump has little effect on the dynamics: the oscillations converge (graph (ii)).

APPLICATION 10.2 Human exploitation and a destabilising Allee effect

A destabilising Allee effect has not apparently been established for any 'natural' predator–prey interaction. On the other hand, when we ourselves are the predator, our isocline may cross that of the prey well to the left of their hump. For example, with exploited whale or fisheries populations, we frequently have the ability (i.e. the technology) to maintain effective predation at low prey densities. If the prey population also exhibits an Allee effect (whales and other large animals often have low birth rates at low densities), then the combination of this and persistent predation may all too readily drive a population towards extinction (Stephens & Sutherland, 1999 ; and see Section 15.3.4).

10.3.3 The population dynamics of intimidation

We saw in Section 9.6.1 that although the idea that 'predators eat prey: it's as uncomplicated as that' has an attractive simplicity, the truth is more complex. There is now plenty of evidence to suggest that non-consumptive effects – the effects on prey of predator intimidation, whether that comes directly from the predators or only from the types of habitat that predators favour – are far too important to ignore. It is therefore also important that we understand the role that these non-consumptive effects might play in driving predator–prey dynamics. Figure 10.14 sets out, schematically, how these roles might arise.

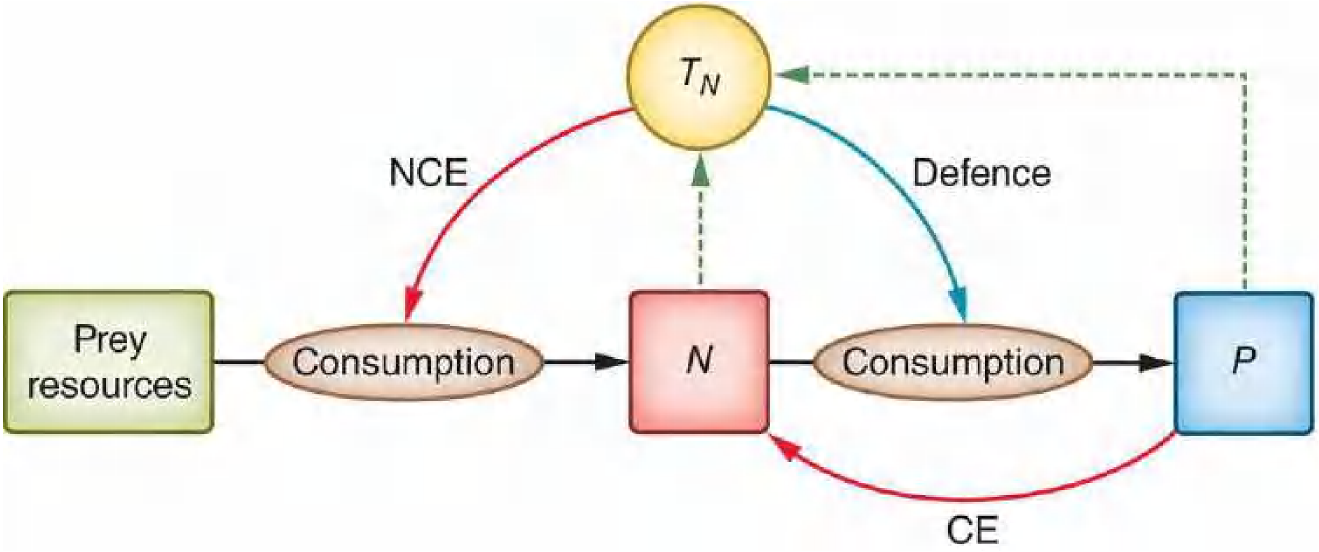


Figure 10.14 Schematic model of the role of non-consumptive effects in predator prey interactions. Prey (abundance N) consume their own resources and are consumed themselves by predators (abundance P). Hence the predators have direct negative consumptive effects (CE) on the prey. But the prey also respond to predators by expressing behavioural or morphological traits, T_N, that may increase in intensity with prey and/or predator abundance (the green dashed arrows). These traits will defend the prey from predator consumption. But the prey will also pay a cost in terms of their own resource acquisition and, potentially, their fitness. These are non-consumptive effects of predation (NCE).

Source: After González-Olivares *et al.* (2017).

We noted previously that the consequences of intimidation are likely to be similar to those of mutual interference: lower predation rates at higher predation densities. This suggests immediately that non-consumptive effects will tend to stabilise predator–prey dynamics. There have also been many modelling studies looking at non-consumptive effects directly, mostly based on the simple models described above (see, for example, Ives & Dobson, 1987 ; Krivan, 2007 ; González-Olivares *et al.*, 2017), but we can understand the key points without going into mathematical details.

Firstly, as we have already noted, increased investment by prey in anti-predator behaviour with increasing predator abundance will give rise to predator zero isoclines like those in Figure 10.10b or c, with a shallower slope the more effective the prey are in resisting predation. This is a reflection of there being prey abundances that would otherwise allow predators to increase in abundance that will no longer do so. Similarly, this increased investment in anti-predator behaviour with increasing predator abundance means that there will be predator abundances that would otherwise reduce prey abundance that will no longer do so, especially at low prey abundance, giving rise to prey zero isoclines that rise to the left like the one in Figure 10.10a, with a steeper negative slope the more effective the prey are in resisting predation. Indeed, such resistance might be seen as providing the prey with a ‘refuge’ from predation – a fraction of the population that will escape predation whatever the abundance of predators. This gives rise to prey zero isoclines like those in Figure 10.11. As we have seen, all of these tend to dampen or even eliminate oscillations in abundance in predator–prey systems. These stabilising consequences of non-consumptive effects can be understood less formally by noting that they give rise to reductions in prey abundance without associated increases in predator abundance and the tendency to cycle that that would generate. Non-consumptive effects tend also to increase prey abundance and decrease the ratio of predator to prey abundance (Ives & Dobson, 1987).

On the other hand, as we can see in Figure 10.14, and as we noted in Section 9.6.1, the consequences of intimidatory effects are varied – for example, the non-consumptive effects of predation may simply add to the harmful consumptive effects – and the consequences for predator–prey dynamics may also be varied. González-Olivares *et al.* (2017), for instance, analysed a model in which the resistance afforded by induced prey traits was explicitly dependent on predator density (Figure 10.14) and found that some parameter values gave outcomes similar to those we have described, but others generated dynamics with two or even three alternative stable states (see Section 14.7). Nonetheless, overall, just as there has been a tendency to neglect non-consumptive effects generally, there is likely to have been a tendency to neglect their importance as a stabilising force in predator–prey dynamics.

10.4 Foraging in a patchy environment

A central assumption of the simple models is that of homogeneity – that all members of both the predator and prey populations experience the interaction in precisely the same way. In this section we question that assumption.

For all consumers, food is distributed patchily. The patches may be natural and discrete physical objects: a bush laden with berries is a patch for a fruit-eating bird; a leaf covered with aphids is a patch for a predatory ladybird. Alternatively, a 'patch' may only exist as an arbitrarily defined area in an apparently uniform environment; for a wading bird feeding on a sandy beach, different 10 m^2 areas may be thought of as patches that contain different densities of worms. In all cases though, a patch must be defined with a particular consumer in mind. One leaf is an appropriate patch for a ladybird; but for a larger and more active insectivorous bird, 1 m^2 of canopy or even a whole tree may represent a more appropriate patch.

aggregative responses

Ecologists have been particularly interested in patch preferences of consumers where patches vary in the density of food or prey items they contain. This preoccupation has given us many examples where predators show an *aggregative response*, spending more time in patches containing high densities of prey because these are the most profitable patches (Figure 10.15a–c), although, as we shall see, this is not always the case (Figure 10.15d, g) and may not necessarily translate into prey in the most profitable patches being at the greatest risk of predation (Figure 10.15f, h). We deal in this section with the behavioural responses of predators to patchily distributed prey before turning in the following section to the consequences of these for prey and predator population dynamics.

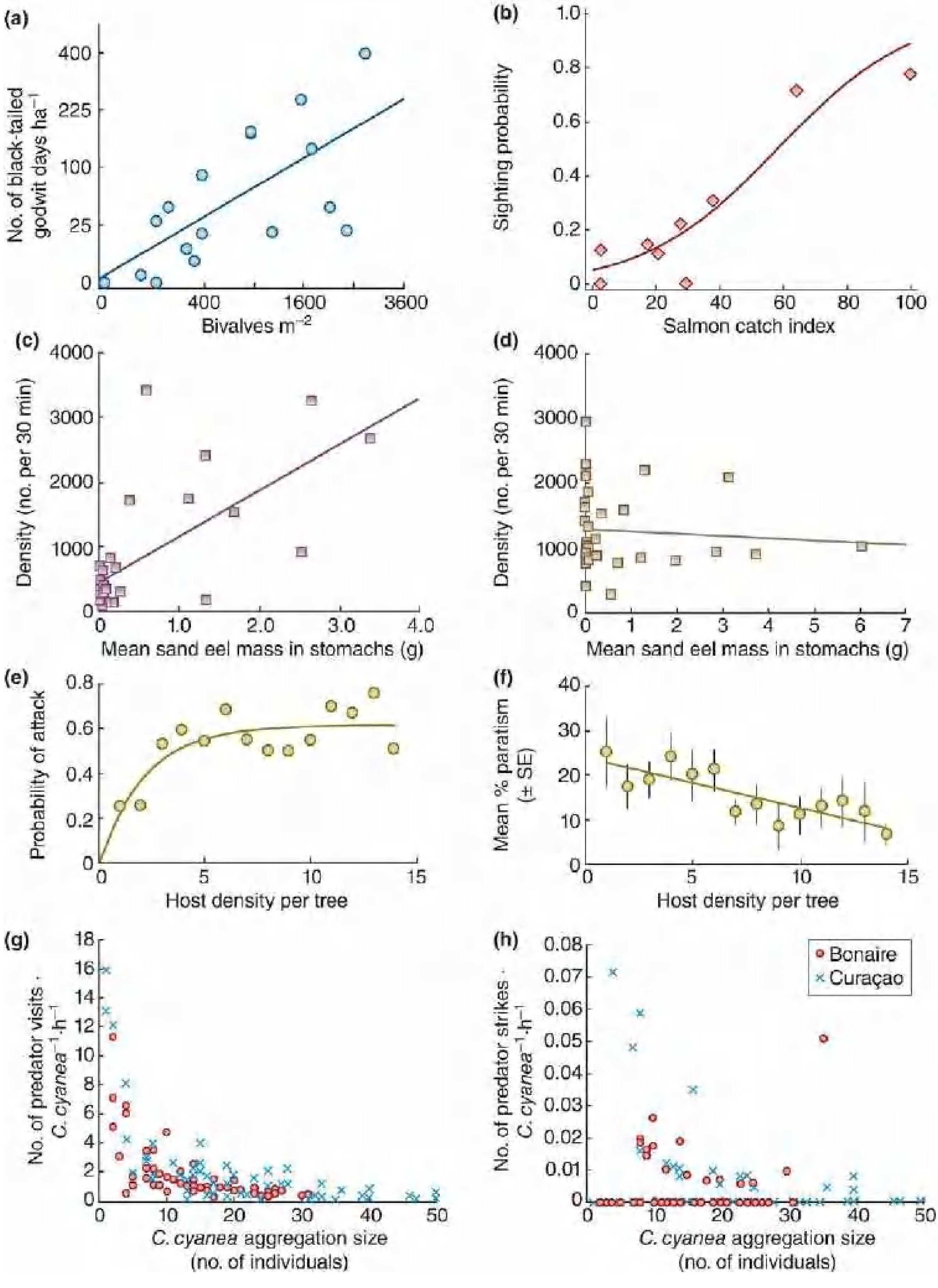
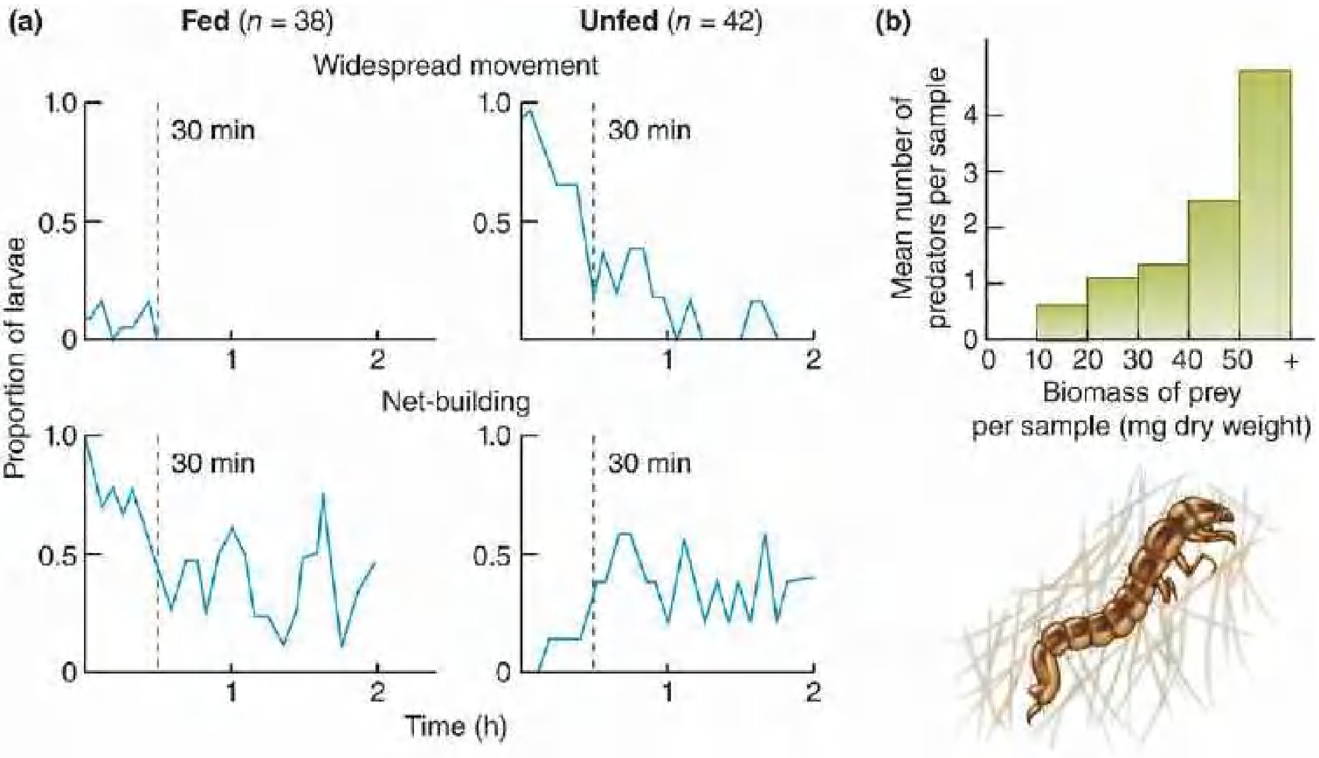

(a)
No. of black-tailed godwit days ha^{-1}
400
225
100
25
0
0
400
1600
3600
Bivalves m^{-2}
(b)
Sighting probability
1.0
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
Salmon catch index
(c)
Density (no. per 30 min)
4000
3000
2000
1000
0
0
1.0
2.0
3.0
4.0
Mean sand eel mass in stomachs (g)
(d)
Density (no. per 30 min)
4000
3000
2000
1000
0
0
1
2
3
4
5
6
7
Mean sand eel mass in stomachs (g)
(e)
Probability of attack
0.8
0.6
0.4
0.2
0
0
5
10
15
Host density per tree
(f)
Mean % paratism (± SE)
40
30
20
10
0
0
5
10
15
Host density per tree
(g)
No. of predator visits · C. cyanea^{-1}·h^{-1}
18
16
14
12
10
8
6
4
2
0
0
10
20
30
40
50
C. cyanea aggregation size (no. of individuals)
(h)
No. of predator strikes · C. cyanea^{-1}·h^{-1}
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
10
20
30
40
50
C. cyanea aggregation size (no. of individuals)
Bonaire
Curaçao

Figure 10.15 Aggregative responses of predators and parasitoids. (a) Black-tailed godwits, *Limosa limosa*, spend most time on patches of sand where their bivalve prey are most common. (b) Harbour seals, *Phoca vitulina*, are most likely to be seen in areas of the River Conon, Scotland where their salmonid prey are most common. (c) Whiting, *Merlangius merlangus*, are found at the greatest densities in areas of the northern North Sea where their sand eel prey are most common. (d) Unlike the pattern in (c), densities of haddock, *Melanogrammus aeglefinus*, show no relationship with the areas of the northern North Sea where their sand eel prey are most common. (e) The parasitoid, *Mastrus ridibundus*, is most likely to attack its codling moth host, *Cydia pomonella*, on trees where the moths are most common. (f) Despite the pattern in (e), codling moths are least likely to be attacked on trees where the moths are most common. (g) Predatory fish on coral reefs in the Netherlands Antilles are more likely to visit the smallest aggregations of their prey, the blue chromis, *Chromis cyanea*. (h) Because of the pattern in (g), blue chromis in the smallest aggregations are the most likely to be attacked.

Source: (a) After Gill *et al.* (2001). (b) After Middlemas *et al.* (2006). (c, d) After Temming *et al.* (2004). (e, f) After Bezemer & Mills (2001). (g, h) After Sandin & Pacala (2005).

10.4.1 Behaviour that leads to aggregated distributions

A consumer's aggregative response could arise simply by that consumer being able to perceive, at a distance, by sight say, or by smell, the existence of heterogeneity in the distribution of their prey. Often, however, the aggregation arises as a result of the responses of consumers once within a patch. There may, for example, be a change in the consumer's pattern of searching after encountering items of food. It may slow down or increase its rate of turning immediately following the intake of food, both of which lead to its remaining in the same vicinity ('area-restricted search'). Alternatively, or in addition, consumers may simply abandon unprofitable patches more rapidly than they abandon profitable ones. This was evident when carnivorous, net-spinning larvae of the caddis fly, *Plectrocnemia conspersa,* fed on chironomid (midge) larvae in a laboratory stream. Caddis flies in their nets were provided with one prey item at the beginning of the experiment and then fed daily rations of zero, one or three prey. They stayed longer at the higher feeding rates (Townsend & Hildrew, 1980). But they were also more likely to spin a net in the first place if they encountered a food item (Figure 10.16a). Overall, therefore, a net is more likely to be constructed, and less likely to be abandoned, in a rich patch, accounting for the directly density-dependent aggregative response in the natural stream environment observed for much of the year (Figure 10.16b).



[Figure 10.16](#) The behaviour of caddis fly larvae leads to their aggregation in prey-rich patches. (a) On arrival in a patch, fifth-instar *Plectrocnemia conspersa* larvae that encounter and eat a chironomid prey item at the beginning of the experiment ('fed') quickly cease wandering and commence net-building. Predators that fail to encounter a prey item ('unfed') exhibit much more widespread movement during the first 30 minutes of the experiment, and are significantly more likely to move out of the patch. (b) Directly density-dependent aggregative response of fifth-instar larvae in a natural environment expressed as mean number of predators against combined biomass of chironomid and stonefly prey per 0.0625 m^2 sample of streambed ($n = 40$).

Source: After Hildrew & Townsend (1980) and Townsend & Hildrew (1980).

thresholds and giving-up times

The difference in the rates of abandonment of patches of high and low profitability can be achieved in a number of ways, but two are especially easy to envisage. A consumer might leave a patch when its feeding rate drops below a threshold level, or a consumer might have a giving-up time – it might abandon a patch whenever a particular time interval passes without the successful capture of food. Whichever mechanism is used, or indeed if the consumer simply uses area-restricted search, the consequences will be the same: individuals will spend longer in more profitable patches, and these patches will therefore generally contain more consumers.

10.4.2 The optimal foraging approach to patch use

The advantages to a consumer of spending more time in higher profitability patches are easy to see. However, the detailed allocation of time to different patches depends on the precise differentials in profitability, the average profitability of the environment as a whole, the distance between the patches, and so on. The problem has been a particular focus of attention for optimal foraging theory, where we can ask how foragers should behave when they themselves deplete the resources of a patch, causing its profitability to decline with time. Amongst the many examples of this are insectivorous insects removing prey from a leaf, and bees consuming nectar from a flower.

the marginal value theorem

Early models found that the optimal stay-time in a patch should be defined in terms of the rate of energy extraction experienced by the forager at the moment it leaves a patch: the ‘marginal value’ of the patch (Charnov, 1976b ; Parker & Stuart, 1976). Charnov called the results the ‘marginal value theorem’. The models were formulated mathematically, but their salient features are shown graphically in Figure 10.17.

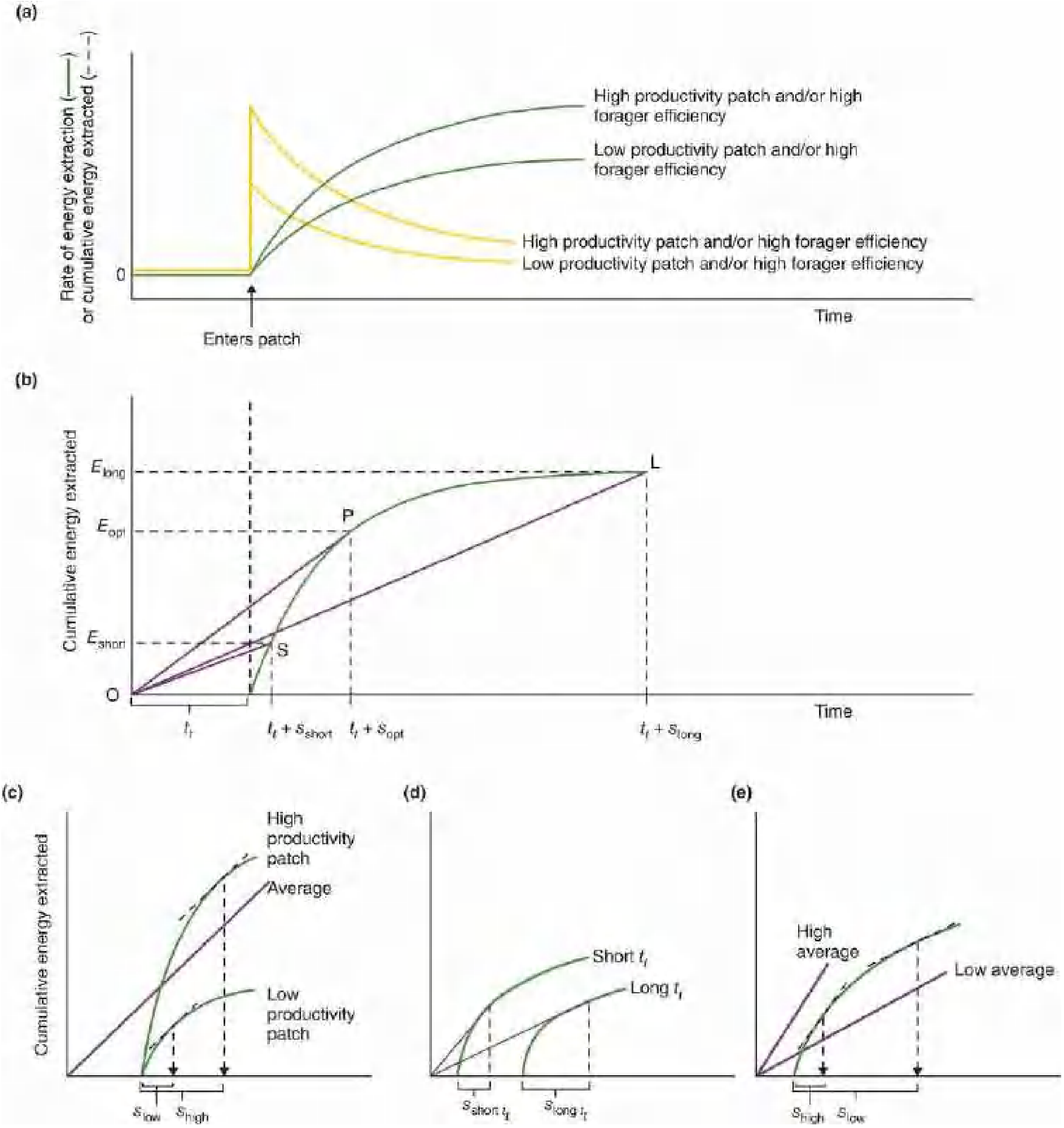

(a)
Rate of energy extraction (——) or cumulative energy extracted (– – –)
0
High productivity patch and/or high forager efficiency
Low productivity patch and/or high forager efficiency
High productivity patch and/or high forager efficiency
Low productivity patch and/or high forager efficiency
Time
Enters patch
(b)
Cumulative energy extracted
E_{long}
E_{opt}
E_{short}
O
L
P
S
t_t
$t_t + s_{short}$
$t_t + s_{opt}$
$t_t + s_{long}$
Time
(c)
Cumulative energy extracted
High productivity patch
Average
Low productivity patch
s_{low}
s_{high}
(d)
Short t_t
Long t_t
$s_{short\ t_t}$
$s_{long\ t_t}$
(e)
High average
Low average
s_{high}
s_{low}

Figure 10.17 The marginal value theorem. (a) When a forager enters a patch, its rate of energy extraction is initially high (especially in a highly productive patch or where the forager has a high foraging efficiency), but this rate declines with time as the patch becomes depleted. The cumulative energy intake approaches an asymptote. (b) The options for a forager. The solid green curve is cumulative energy extracted from an average patch, and t_t is the average travelling time between patches. The rate of energy extraction (which should be maximised) is energy extracted divided by total time, i.e. the slope of a straight (in this case purple) line from the origin to the curve. Short stays in the patch (slope = $E_{\text{short}}/(t_t + s_{\text{short}})$) and long stays (slope = $E_{\text{long}}/(t_t + s_{\text{long}})$) both have lower rates of energy extraction (shallower slopes) than a stay (s_{opt}) which leads to a line just tangential to the curve. s_{opt} is therefore the optimum stay-time, giving the maximum overall rate of energy extraction. *All* patches should be abandoned at the *same* rate of energy extraction (the slope of the line OP). (c) Low productivity patches should be abandoned after shorter stays than high productivity patches. (d) Patches should be abandoned more quickly when travelling time is short than when it is long. (e) Patches should be abandoned more quickly when the average overall productivity is high than when it is low.

The primary assumption of the model is that an optimal forager will maximise its overall intake of a resource (usually energy) during a bout of foraging, taken as a whole. The forager will sometimes move between patches, during which time its intake of energy will be zero, but once in a patch, the forager will extract energy in a manner described by the curves in Figure 10.17a. Its initial rate of extraction will be high, but as time progresses and the resources are depleted, the rate of extraction will steadily decline. Of course, the rate will itself depend on the initial contents of the patch and on the forager's efficiency and motivation (Figure 10.17a).

when should a forager leave a patch that it is depleting?

At what point should a forager leave a patch? If it left all patches early, then it would spend most of its time travelling between patches, and its overall rate of intake would be low. If it remained in all patches for long periods, then it would waste little time travelling, but it would spend extended periods in depleted patches. Its overall rate of intake would again be low. Some intermediate stay-time is therefore optimal. In addition, though, the optimal stay-time must clearly be greater for profitable patches than for unprofitable ones, and it must depend on the profitability of the environment as a whole.

Consider, in particular, the forager in Figure 10.17b. It is foraging in an environment where food is distributed patchily and where some patches are more valuable than others. The average travelling time between patches is t_t. This is therefore the length of time the forager can expect to spend on average after leaving one patch before it finds another. The forager in Figure 10.17b has arrived at an average patch for its particular environment, and it therefore follows an average extraction curve. In order to forage optimally it must maximise its rate of energy intake not merely for its period in the patch, but for the whole period since its departure from the last patch (i.e. for the period $t_t + s$, where s is the stay-time in the patch). As the figure shows, to do this it is necessary to maximise the slope of the line from O to the extraction curve. This is achieved simply by making the line a tangent to the curve (OP in Figure 10.17b). No line from O to the curve can be steeper, and the stay-time associated with it is therefore optimal (s_{opt}).

how to maximise overall energy intake

The optimal solution for the forager in Figure 10.17b, therefore, is to leave that patch when its extraction rate is equal to (tangential to) the slope of OP (i.e. it should leave at point P). In fact, the optimal solution for the forager is to leave all patches, irrespective of their profitability, at the same extraction rate (i.e. the same ‘marginal value’). This extraction rate is given by the slope of the tangent to the average extraction curve (e.g. in Figure 10.17b), and it is therefore the maximum average overall rate for that environment as a whole.

predictions of the marginal value theorem ...

The model therefore confirms that the optimal stay-time should be greater in more productive patches than in less productive patches (Figure 10.17c). Moreover, for the least productive patches (where the extraction rate is never as high as OP) the stay-time should be zero. The model also predicts that all patches should be depleted such that the final extraction rate from each is the same (i.e. the ‘marginal value’ of each is the same). It predicts that stay-times should be longer in environments where the travelling time between patches is longer (Figure 10.17d). And it predicts that stay-times should be longer where the environment as a whole is less profitable (Figure 10.17e).

... supported by some experiments ...

There are a number of studies that lend support to the marginal value theorem. One examined the behaviour of the parasitoid *Anaphes victus* attacking eggs of the beetle *Listronotus oregonensis* in a laboratory setting (Boivin *et al.*, 2004). Patches differed in quality by virtue of the varying proportions of hosts already parasitised at the start of the experiment, and in line with the theorem’s predictions, parasitoids stayed longer in the more profitable patches (Figure 10.18a). However, contrary to a further prediction, the marginal rate of fitness gain (the rate of progeny production in the final 10 minutes before leaving a patch) was greatest in the initially most profitable patches (Figure 10.18b). The animals did not stay long enough in the more profitable patches to reduce the marginal gains there to the levels of the less profitable ones.

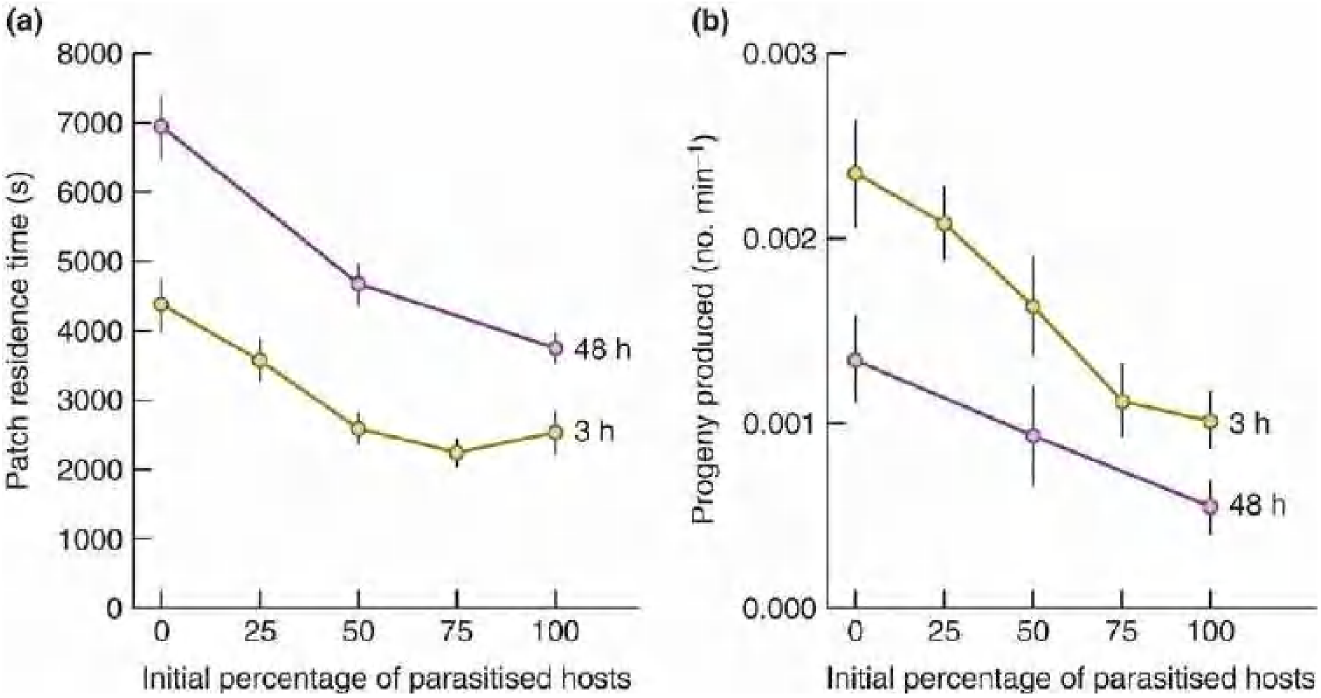


Figure 10.18 Experiments with parasitoids provide qualified support for the marginal value theorem. (a) When the parasitoid *Anaphes victus* attacked the beetle *Listronotus oregonensis* in patches of 16 hosts, a varying percentage of which had already been parasitised, parasitoids remained longer in the more profitable patches: those with the smaller percentage of parasitised hosts. This was true whether there was a short (3 h) or long (48 h) delay, initially, simulating different travel times between patches. (b) However, the marginal gain rate in fitness – the number of progeny produced per minute in the final 10 minutes before leaving a patch – was greatest in the initially most profitable patches. Bars are SEs.

Source: After Boivin *et al.* (2004).

... and by some field observations

A number of other studies have tested the marginal value theorem by providing animals with artificial patches either in the lab or in the field. However, tests in a more natural setting have been much rarer. One exception, using recent technological advances, monitored the foraging behaviour of Adélie penguins, *Pygoscelis adeliae*, feeding on patches of krill in the Southern Ocean. Accelerometers attached to the penguins' heads (one at the front, one at the rear) were able to monitor both prey capture events and the penguins' depth (Figure 10.19a). The saturating rate of return within a patch, assumed by the marginal value theorem, was confirmed (Figure 10.19b), though initially that rate accelerated. Crucially though, in line with the predictions of the marginal value theorem, patch residence time (dive duration) was greater in patches with higher productivity (those with higher krill capture-rates at the dive scale; Figure 10.19c) but was lower in environments that overall were more productive (with higher krill capture rates at the bout scale) (Figure 10.19d).

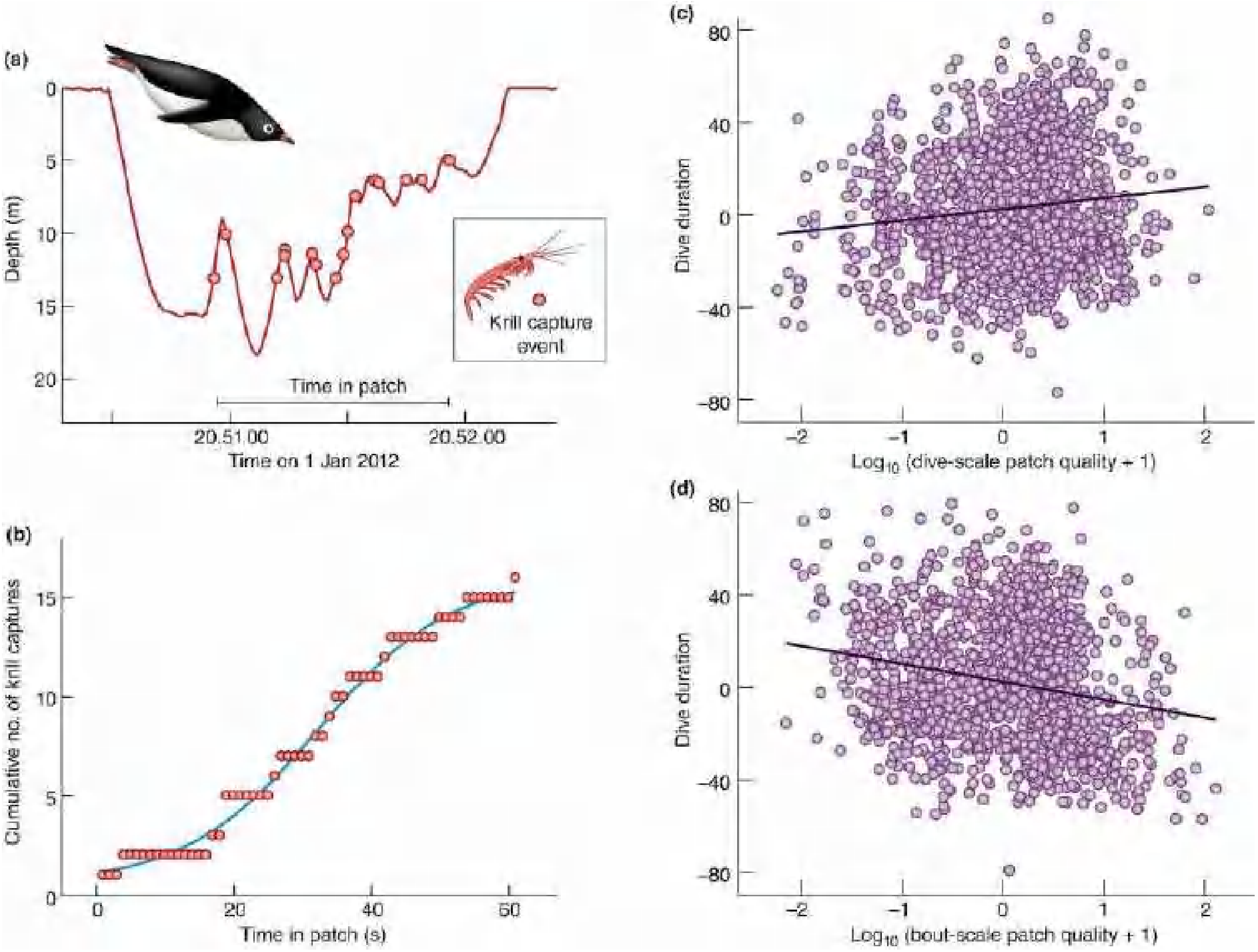
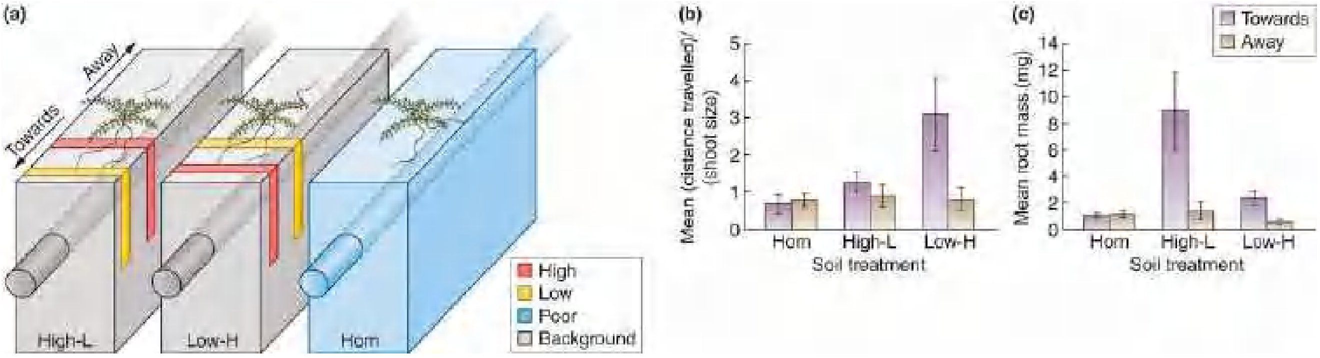


Figure 10.19 Optimal foraging by penguins feeding on patches of krill provide support for the marginal value theorem. (a) A typical time series of capture events; 'time in a patch' is the time between first and last capture. (b) The cumulative number caught showed diminishing returns by the end of a patch-visit, though the best-fit curve was sigmoidal overall. (c) Penguins spent more time in more productive patches (correlation = 0.13 in the best-fit partial regression model taking other factors, including dive depth, into account). (d) But penguins spent less time in patches when the overall productivity was higher (correlation = −0.21 in the same model).

Source: After Watanabe *et al.* (2015).

optimal foraging in plants

There has even been support for the marginal value theorem from the 'behaviour' (growth) of plants, as part of a more widespread and welcome attempt to carry 'foraging' across the taxonomic divide (Hutchings & de Kroon, 1994). In one example, the herbaceous perennial, *Achillea millefolium,* was grown in three soil environments that had the same overall concentration of nutrients but varied in their distribution: one had nutrients distributed homogeneously, one had a highly enriched patch close to one side of the plant and a low-concentration patch placed a little further away, and a third had the positions of the high- and low-concentration patches reversed (Figure 10.20a). The plant roots grew faster towards patches than they did in the homogeneous environment (Figure 10.20a), but crucially, having reached a patch, they moved quickly on from the low- to the high-concentration patch (Figure 10.20b), so that their stay-time was far longer in the high-concentration patch than in the low (Figure 10.20c).



[**Figure 10.20**](#) **Patterns of plant root growth provide support for the marginal value theorem.** (a) The experimental design in which plants, *Achillea millefolium*, were grown in three environments with the same overall concentration of nutrients: homogeneous (Hom, 4% manure), close to a high-concentration patch (66% manure) with a low-concentration patch (25%) beyond it (High-L), and close to a low-concentration patch with a high-concentration patch beyond it (Low-H). The cylinders indicate cameras with which root growth was monitored. (b) Plants roots grew at the same rate (standardised for plant size) away from patches and where there were none. But they grew fastest through the low-concentration patch to the high-concentration patch. (c) Consequently, they stayed longest, and grew most, within the high-concentration patch with the low-concentration patch beyond. Bars are SEs.

Source: After McNickle & Cahill (2009).

optimal patch use predictions modified by predation risk

A much fuller review of tests of the marginal value theorem is provided, for example, by Krebs and Kacelnik (1991). The picture this conveys is one of encouraging but not perfect correspondence. This is not surprising. The marginal value theorem provides us with a useful foundation, but it is not the whole story. In the first place, as with optimal diet theory, predation pressure should modify the predicted outcomes of optimal patch use. With this in mind, Morris and Davidson (2000) compared the giving-up food extraction rates of white-footed mice (*Peromyscus leucopus*) in a forest habitat (where predation risk is low) and a forest-edge habitat (where predation risk is high). They provided ‘patches’ (containers with millet grain) in 11 foraging sites in the two habitat types, and in both, some sites were in relatively open situations and others beneath shrubs. They then monitored the grain remaining at the time the patches were abandoned. As predicted (Figure 10.21), the mice abandoned patches at higher harvest rates where predation risk was highest: in vulnerable edge habitats rather than safe forest habitats, and particularly in open situations.

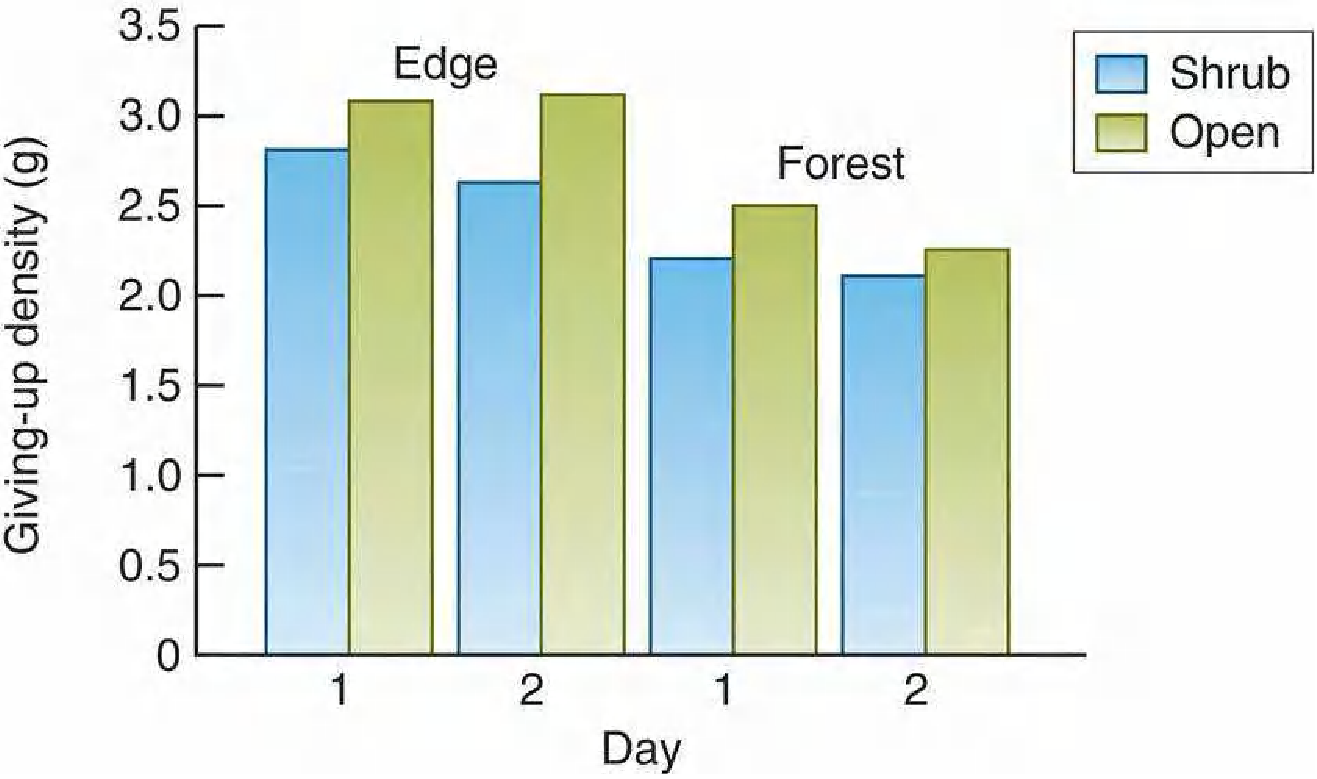


Figure 10.21 The use of patches differing in resource richness by mice is modified by predation risk. For white-footed mice, *Peromyscus leucopus*, the mass of millet grain remaining (giving-up density) was higher in patches in the open (riskier) than in paired patches located under shrubs (safer), and was higher in forest-edge habitat (higher predation) than in forest (lower predation).

Source: After Morris & Davidson (2000).

mechanistic models of patch use – complementary to tests of the marginal value theorem

A more fundamental reason for the mismatch between observed patterns of foraging and the predictions of the theory is that the animals, unlike the modellers, are not omniscient; they do not have perfect information about their environment, either at the outset or during the time they are foraging. In some cases, foragers appear to operate on the basis of fixed, mechanistic rules that determine when they leave a patch, but increasingly we find support for foragers behaving more flexibly, basing their decisions on their past experience. Thus, foragers may spend time learning about and sampling their environment, but even after they have done so, they are likely to be foraging with only imperfect information about the distribution of their prey.

For the results in Figure 10.18, for example, where the correspondence with the marginal value theorem's predictions was not perfect, Boivin *et al.* (2004) suggest that parasitoids seem to base their assessment of overall habitat quality on the quality of the first patch they encounter. That is, they 'learn', but their learned assessment may still be wrong. Such a strategy would be adaptive, though, if there was considerable variation in quality between generations (so that each generation had to learn anew), but little variation in quality between patches within a generation (so that the first patch encountered was a fair indication of quality overall).

A more thoroughgoing comparison of the predictions of two mechanistic models and those of the marginal value theorem used data on foraging in chacma baboons (*Papio ursinus*), both in natural and artificial settings (Marshall *et al.*, 2013). The two mechanistic models were themselves similar. The first was a 'Bayesian updating' model, in which foragers update their knowledge iteratively, with information from each successive bout of foraging being combined with prior knowledge that had itself been generated in a similar way from earlier bouts. The second was a learning model, in which foragers also integrate their past experiences into a current state of knowledge, but this time on the simpler basis of a moving average of past environments. The work was carried out in the Tsaobis Leopard Park, Namibia. Natural foraging patterns were determined by monitoring individuals for 20–30-minute periods, noting where they foraged and what they ate (54 individuals followed for a total of 1481 hours, containing 6175 patch visits), and combining this with information on the distribution and composition of the patches over a five month sampling period. Feeding experiments were also carried out in the field, but using the very simplified and artificial arrangement of food patches (tubs of maize kernels) shown in Figure 10.22.

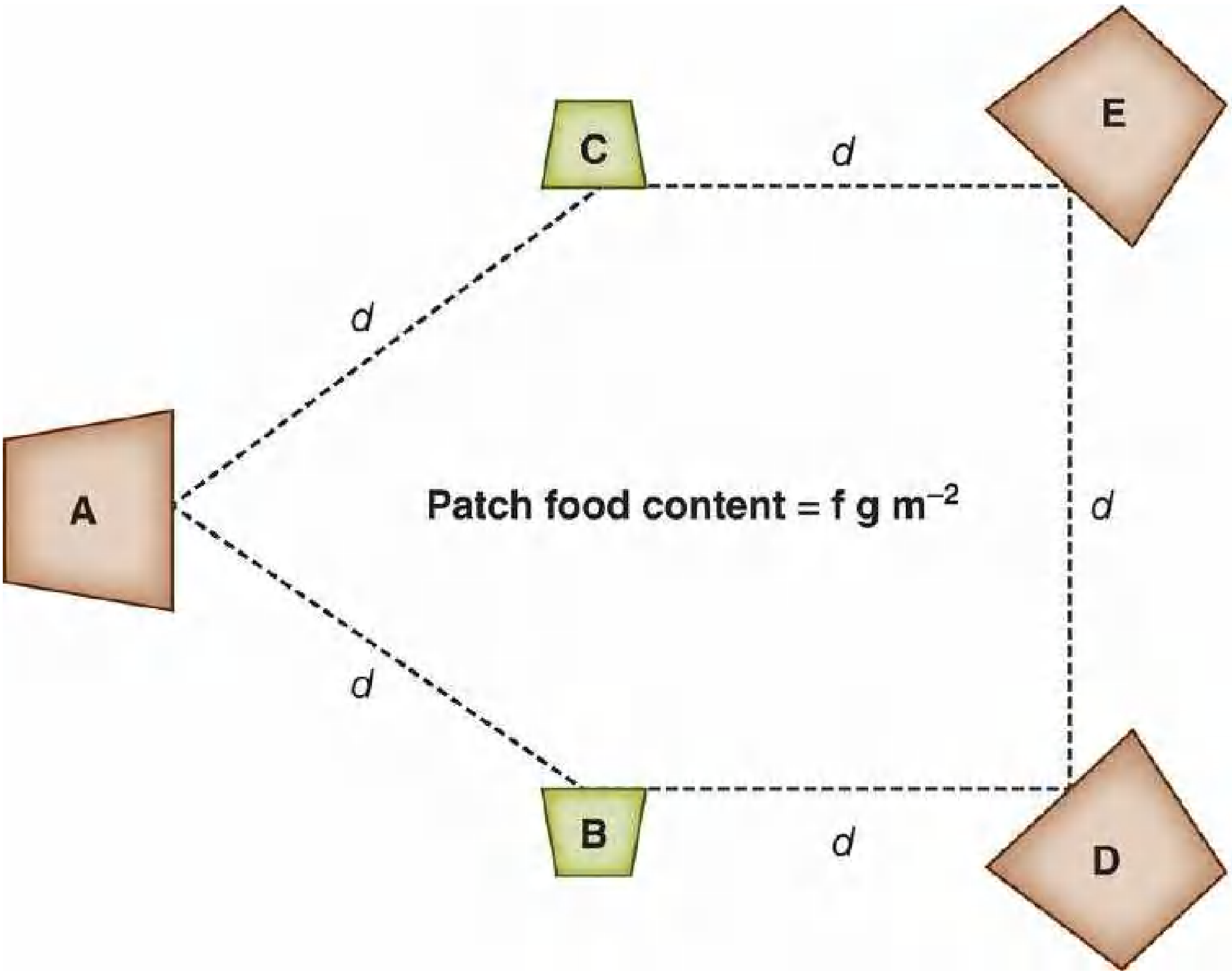


Figure 10.22 **Layout of a foraging experiment with chacma baboons.** Two separate baboon troops, *Papio ursinus*, were used, each, alternately, for two 14-day periods. Patch food content (loose dried maize kernels) was varied between 11.4 g m^{-2} (low, first 14-day period) and 17.4 g m^{-2} (high, second 14-day period). Interpatch distance (*d*) was was set at 25 m (short) for the initial seven days and 50 m (long) for the remaining seven days of the first period, reversed for the second period. Large patches (A, D and E) were 80 m^2 for the small troop and 96 m^2 for the large troop. Small patches (B and C) were 20 m^2 for the small troop and 27 m^2 for the large troop.

Source: After Marshall *et al*. (2013).

The study compared the three models in terms of how much support they received as being best at explaining the datasets. (The statistical method assigns weights to the explanatory power of the different models – the higher the weight (maximum 1.00), the greater the explanatory power and the stronger the support.) The marginal value theorem was best for natural foraging (weight

0.69), but the Bayesian updating model also received some support (weight 0.27). The learning model was much worse (weight 0.00). However, for experimental foraging, Bayesian updating was easily the best (weight 0.98) compared with both the marginal value theorem (weight 0.02) and the learning rule (again, weight 0.00). Perhaps equally important, though, for natural foraging the marginal value theorem and Bayesian updating models could also be used to predict the effects of overall patch quality and patch density on the residence times in patches, and in both cases the predictions of the two models were identical (Table 10.3). Also, the Bayesian updating model allows the influence of the previous patch to be assessed. This was an order of magnitude greater for experimental than for natural foraging (Table 10.3).

Table 10.3 Factors determining patterns in baboon foraging. The estimated parameter values (±SE) from models that receive statistical support when applied to baboon foraging in natural and experimental environments.

Source: After Marshall *et al.* (2013).

Natural environment			Experimental environment	
Predictors	**Marginal value theorem**	**Bayesian updating**	**Predictors**	**Bayesian updating**
Time in previous patch (s)		0.006 ± 0.02	Time in previous patch (s)	0.08 ± 0.02
Mean number of food items per patch	−0.11 ± 0.03	−0.11 ± 0.03	Mean weight of food per patch (g)	−0.56 ± 0.15
Mean number of food patches per km^2	−0.16 ± 0.02	−0.16 ± 0.03	Interpatch distance (m)	0.10 ± 0.04

All this suggests, for natural foraging at least, that patch use by the baboons comes close to what the marginal value theorem would predict is evolutionarily optimal, but also that behaviour by the baboons that followed a Bayesian updating process would allow them to come close to achieving this evolutionary optimum. They seem to do so by incrementally incorporating information from previous patches into the design of their foraging strategy. But whereas in the simple, largely unchanging experimental environment, the most recent experience is strongly influential, in the much more complex natural environment, recent experience carries relatively little weight compared with accumulated earlier experiences, allowing the baboons, by successive approximation, to construct a cognitive picture of the whole environment – precisely what the marginal value theorem assumes.

Thus, in spite of their limited information, animals seem often to come remarkably close to the strategy predicted by us for an omniscient forager operating according to the marginal value theorem. As Krebs and Davies (1993) point out, this is no more surprising than the observation that birds can fly without any formal qualification in aerodynamics. The marginal value theorem aims to identify the target that natural selection should carry foragers towards. The methods adopted by 'real' foragers have been favoured by natural selection precisely because they have carried their ancestors closer to that target. Optimal foraging and more mechanistic models are therefore compatible and complementary in explaining how a predator has achieved its observed foraging pattern, and why that pattern has been favoured by natural selection.

10.4.3 Ideal free and related distributions: aggregation and interference

the ideal free distribution ...

We can see, then, that consumers tend to aggregate in profitable patches where their expected rate of food consumption is highest. Yet we might also expect that consumers will compete and interfere with one another, as discussed previously, thereby reducing their per capita consumption rate. It follows from this that patches that are initially most profitable become immediately less profitable because they attract most consumers. We might therefore expect the consumers to redistribute themselves, and it is perhaps not surprising that the observed patterns of predator distributions across prey patches vary substantially from case to case. But can we make some sense of this variation in pattern?

... is a balance between attractive and repellent forces

In an early attempt to do so, it was proposed that if a consumer forages optimally, the process of redistribution will continue until the profitabilities of all patches are equal (Fretwell & Lucas, 1970 ; Parker, 1970). This will happen because as long as there are dissimilar profitabilities, consumers should leave less profitable patches and be attracted to more profitable ones. Fretwell and Lucas called the consequent distribution the ideal free distribution: the consumers are 'ideal' in their judgement of profitability, and 'free' to move from patch to patch. Consumers were also assumed to be equal. Hence, with an ideal free distribution, because all patches come to have the same profitability, all consumers have the same consumption rate. An opposite extreme would be a 'despotic' distribution, in which the best competitor settled in the best habitat and excluded all others, the second did the same from what was left, and so on, with the worst competitors probably excluded altogether. No doubt the truth often lies somewhere between these two extremes. For example, there are some simple cases where consumers appear to conform to an ideal free distribution insofar as they distribute themselves in proportion to the profitabilities of different patches (the ducks in Figure 10.23a). But even in such cases, one of the underlying assumptions is likely to have been violated: not all the ducks were equal competitors (Figure 10.23b).

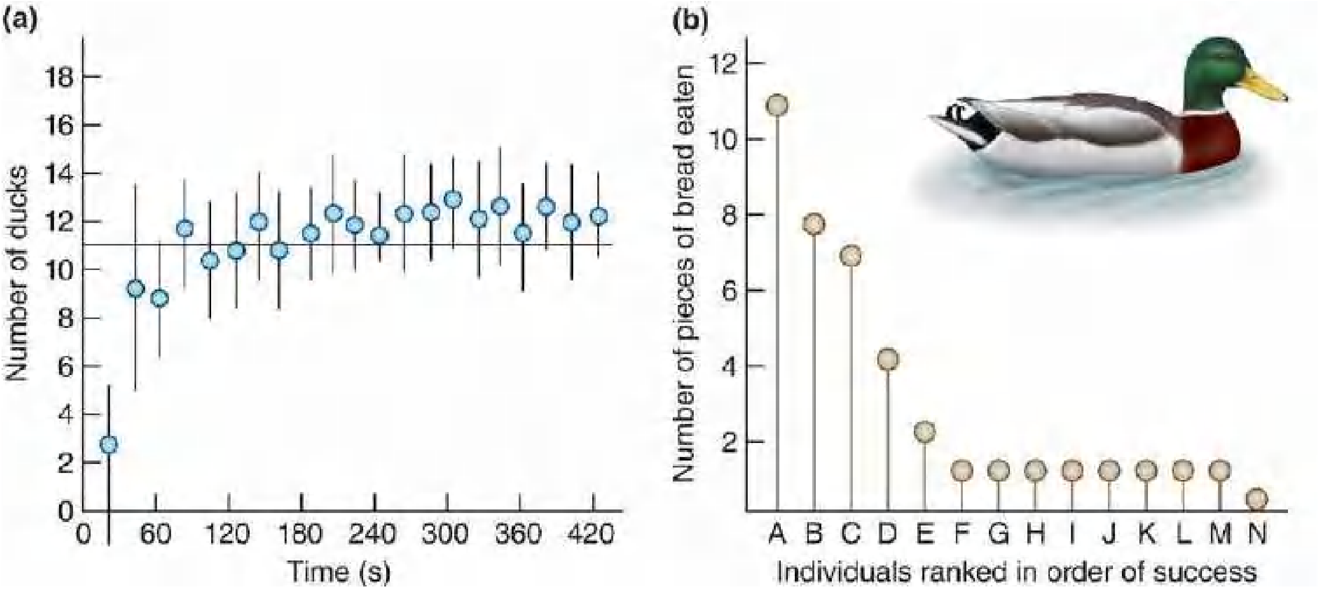


Figure 10.23 Ducks provide support for the ideal free distribution. (a) When 33 ducks were fed pieces of bread at two stations around a pond (with a profitability ratio of 2:1), the number of ducks at the poorer station, shown here, rapidly approached one-third of the total, in apparent conformity with the predictions of ideal free theory. (b) However, contrary to the assumptions and other predictions of simple theory, the ducks were not all equal.

Source: After Harper (1982), from Milinski & Parker (1991).

incorporating a range of interference coefficients

The early ideas have been much modified, taking account, for example, of unequal competitors (see Tregenza (1995) for a review). In particular, the distribution of predators among prey patches was put in a more ecological context by Sutherland (1983) when he explicitly incorporated predator handling times and mutual interference amongst the predators. Without going through the details of his argument, we can examine the consequences of his main finding, which was that predators should be distributed such that the proportion of predators in site i, p_i, is related to the proportion of prey (or hosts) in site i, h_i, by the equation:

$$p_i = k\left(h_i^{1/m}\right) \tag{10.20}$$

where m is the coefficient of interference, and k is a 'normalising constant' such that the proportions, p_i, add up to 1. It is now possible to see how the patch-to-patch distribution of predators might be determined jointly by interference and the selection by the predators of intrinsically profitable patches.

If there is no interference amongst the predators, then $m = 0$. All should exploit only the patch with the highest prey density (Figure 10.24), leaving lower density patches devoid of predators.

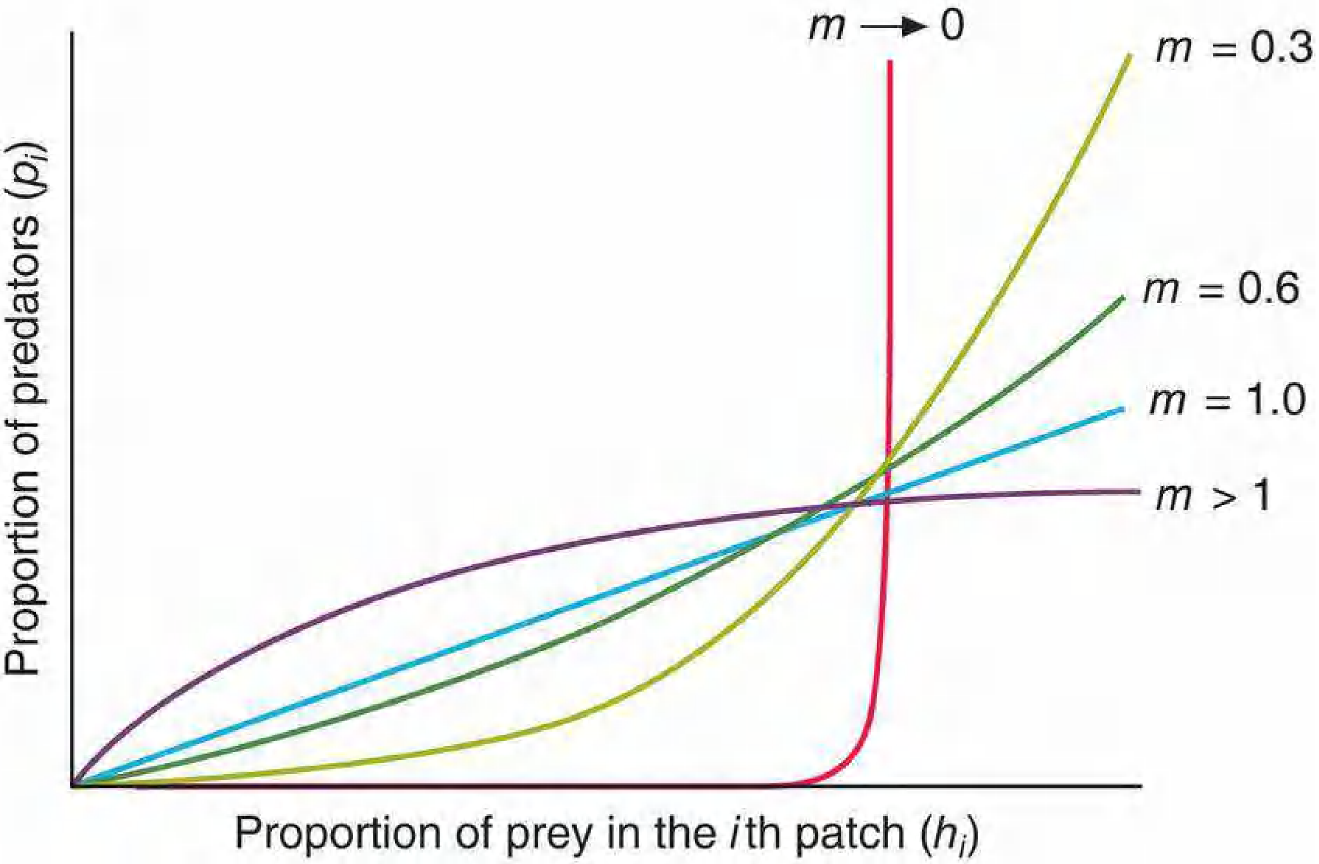


Figure 10.24 The effect of the interference coefficient, *m*, on the expected distribution of predators amongst patches of prey varying in the proportion of the total prey population they contain (and hence, in their 'intrinsic' profitability).

Source: After Sutherland (1983).

If there is a small or moderate amount of interference (i.e. $m > 0$, but $m < 1$ – a biologically realistic range), then high-density prey patches should still attract a disproportionate number of predators (Figure 10.24). In other words, there should be an aggregative response by the predators like the examples we saw in Figure 10.15 when we started to discuss predator aggregation; and we can also say that these predators' distribution shows *spatial density* *dependence*; that is, their density is highest where their prey’s density is highest. Indeed, with this level of interference, predator density actually accelerates with increasing prey density in a patch, as it does for the seals in Figure 10.15b. Hence, the prey’s risk of predation might itself be expected to be density dependent: the greatest risk of predation in the highest prey density patches.

With a little more interference ($m \approx 1$) the proportion of the predator population in a patch should still increase with the proportion of prey, but now it should do so more or less linearly rather than accelerating, such that the ratio of predators to prey is roughly the same in all patches (Figure 10.24, and the aggregative responses we saw in Figures 10.15a and c). Here, therefore, the risk of predation might be expected to be the same in all patches, and hence independent of prey density. Finally, with a great deal of interference ($m > 1$) the highest density prey patches should have the lowest ratio of predators to prey (Figure 10.24). The risk of predation might therefore be expected to be greatest in the lowest prey density patches and hence be inversely density dependent. Such inverse density dependence was apparent in Figure 10.15f and h.

aggregative responses and the aggregation of risk

It is clear, therefore, that the distributions of predators across prey patches reflect a shifting balance between the forces of attraction and of repulsion. Predators are attracted to highly profitable patches; but they are repelled by the presence of other predators that have been attracted in the same way. We have also seen that the relationship between these aggregative responses of predators and the aggregation of risk among prey is not straightforward (where, remember, 'aggregation of risk' means simply that the risk of predation is significantly greater in some patches than in others). This is summarised in Figure 10.25. When the distribution of predators over prey patches follows a perfect, straight line density-dependent relationship, the predator : prey ratio is the same in each host patch and the risk is likely to be the same in each host patch, too (Figure 10.25a). There is no aggregation of risk. On the other hand, when the aggregative response of the predators leads to a directly density-dependent relationship that accelerates (Figure 10.25b), there does appear to be aggregation of risk among prey. Moreover, there can be considerable aggregation of risk with either inverse spatial density dependence (Figure 10.25c) or no spatial density dependence of any sort (Figure 10.25d).

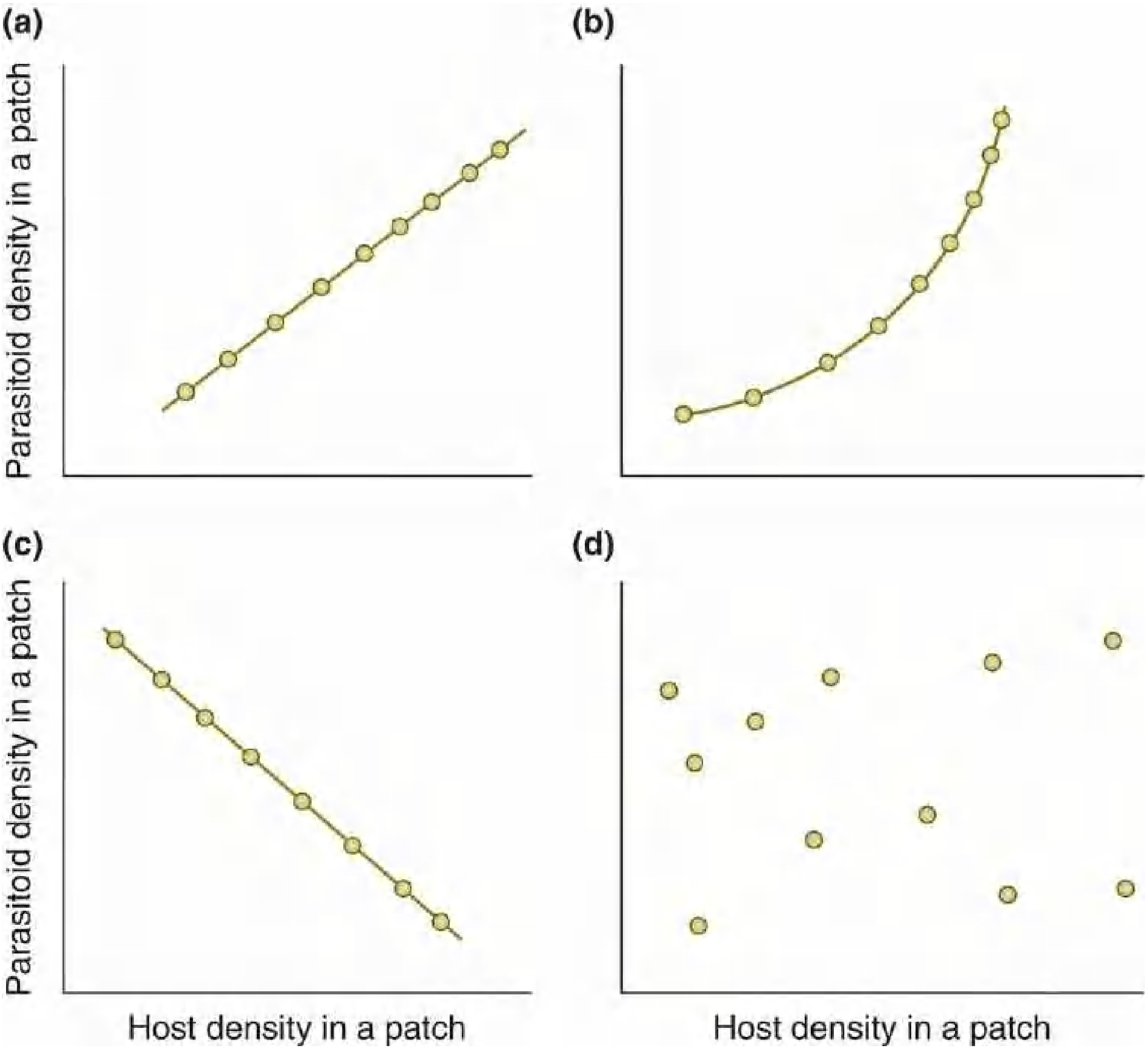


Figure 10.25 The aggregative responses of parasitoids and the aggregation of risk. (a) Parasitoids aggregate in high host-density patches, but the parasitoid : host ratio is the same in all patches (a perfect straight-line relationship), and hence the risk to hosts is apparently the same in all patches. (b) Parasitoid aggregation to high host-density patches now accelerates with increasing host density, and hosts in high-density patches are thus apparently at greater risk of parasitisation: there is aggregation of risk. (c) With perfect inverse density dependence (i.e. parasitoid aggregation in *low* host-density patches) the hosts in the low-density patches are apparently at a much greater risk of parasitisation: again there is aggregation of risk. (d) Even with no aggregative response (density independence) the hosts in some patches are apparently at a greater risk of parasitisation (are subject to a higher parasitoid : host ratio) than others: here too there is aggregation of risk.

learning and migration

Expected patterns are modified further still if we incorporate learning by the predators, or the costs of migration between patches (Bernstein *et al*., 1988 , 1991). If the predators' learning response is weak, for example, then they may be unable to track changes in prey density that result from patch depletion, and their distribution may drift from one that is density dependent to one that is independent of the density of prey. Similarly, when the cost of migration is very high, it does not pay predators to move whatever patch they are in, and mortality that would otherwise be directly density dependent may become inversely density dependent across all patches.

10.5 The population dynamics of heterogeneity, aggregation and spatial variation

Maintaining the approach we have followed throughout this chapter, having elaborated the variety of heterogeneities inherent in most predator–prey interactions, we now turn to examine the consequences of these for the dynamics of both predator and prey.

10.5.1 Aggregative responses to prey density

Clearly, there is no shortage of potential causes for the wide range of types of distributions of predators, and of mortality rates, across prey patches. Reviews of host–parasitoid interactions, for example (e.g. Pacala & Hassell, 1991), confirm that prey (hosts) in high-density patches are not necessarily the most vulnerable to attack (direct density dependence); percentage parasitism may also be inversely density dependent or density independent between patches. Indeed, the reviews suggest that only around 50% of the studies examined show evidence of density dependence, and in only around 50% of these is the density dependence direct, as opposed to inverse. Nonetheless, despite this variation in pattern, it remains true that the risk of predation often varies greatly between patches, and hence between individual prey.

Many herbivores also display a marked tendency to aggregate, and many plants show marked variation in their risk of being attacked. The cabbage aphid (*Brevicoryne brassicae*) forms aggregations at two separate levels. Nymphs quickly form large groups when isolated on the surface of a single leaf, and populations on a single plant tend to be restricted to particular leaves. When aphids attack only one leaf of a four-leaved cabbage plant (as they do naturally), the other three leaves survive; but if the same number of aphids are evenly spread over the four leaves, then all four leaves are destroyed (Way & Cammell, 1970). The aggregative behaviour of the herbivores affords protection to the plant overall.

But how might such heterogeneities influence the dynamics of predator–prey interactions?

10.5.2 Heterogeneity in predator–prey models

refuges, partial refuges and vertical isoclines

We can start by incorporating some relatively simple types of heterogeneity into the Lotka–Volterra isoclines. Suppose that a portion of the prey population exists in a refuge: for example, shore snails packed into cracks in the cliff-face, away from marauding birds, or plants that maintain a reserve of material underground that cannot be grazed. In such cases, the prey zero isocline rises vertically at low prey densities (as we saw previously in Figure 10.11), since prey at low densities, hidden in their refuge, can increase in abundance irrespective of predator density.

Even if predators tend simply to ignore prey in low-density patches, as we have seen in some aggregative responses, this comes close to those prey being in a refuge, in the sense that the predators do not (rather than cannot) attack them. The prey may therefore be said to have a ‘partial refuge’, and this time the prey isocline can be expected to rise almost vertically at low prey abundances. As we saw above, such isoclines have a tendency to stabilise interactions.

negative binomial encounters: pseudo-interference

Early analyses of the Nicholson–Bailey model agreed with this conclusion. May (1978), for example, argued simply that the distribution of host–parasitoid encounters was not random but aggregated, and could be described by a particular statistical model of aggregations, the negative binomial. The conclusion from an analysis of this model was, again, that the system is given a marked boost in stability by the incorporation of significant levels of aggregation, and in particular, aggregation appears capable of generating stable host abundances well below the host's normal carrying capacity.

How does this stability arise out of aggregation? The answer lies in what has been called 'pseudo-interference' (Free *et al*., 1977). With mutual interference, as predator density increases, predators spend an increasing amount of time interacting with one another, and their attack rate therefore declines. Here, attack rate also declines with parasitoid density, but as a result of an increasing fraction of encounters being wasted on hosts that have already been attacked (Morrison & Strong, 1981; Hassell, 1982). In other words, there is indirect interference, or competition, between the parasitoids. As a result, the chances of a host being parasitised may actually be lowest in patches to which most parasitoids have been attracted – precisely the pattern we saw for the codling moth example in Figure 10.15.

aggregation of risk strengthens direct (temporal) density dependences

The crucial point here is that with any spatial aggregation of risk among hosts, as parasitoid density increases, competition between them increases, and their effective attack rate, and hence their subsequent birth rate, decline rapidly – a *directly* density-dependent effect. This dampens both the natural oscillations in parasitoid density, and their impact on host mortality. Thus, the stabilising powers of this spatial phenomenon, aggregation of risk, arise not from any spatial density dependencies, but from its translation into direct, temporal density dependence (Taylor, 1993).

We also saw previously (Figure 10.25) that the relationship between the spatial density dependence in the responses of predators (their aggregative response) and the aggregation of risk among prey was itself not straightforward. There can be an aggregative response (behaviourally) but no aggregation of risk, and no aggregative response but nonetheless aggregation of risk. Hence, the link between any spatial density dependence in the aggregative response and the direct, temporal density dependence required to enhance stability is even more tenuous. Reinforcing this, in an analysis of 65 datasets, representing 26 different host–parasitoid combinations (Pacala & Hassell, 1991), only 18 appeared to have sufficient aggregation of risk to stabilise their interactions, and for 14 of these 18 cases, it was variation independent of host density that contributed most to the total, further weakening any imagined link between spatial density dependence and stability.

continuous redistribution of predators and prey

In host–parasitoid models, the parasitoids are assumed, in effect, to arrange themselves over host patches at the beginning of a generation and then to have to suffer the consequences of that arrangement until the beginning of the next generation. But suppose we move into continuous time – as appropriate for many parasitoids as it is for many other predators. Predators in a depleted or even a depleting patch should leave and redistribute themselves (see Section 10.4). Murdoch and Stewart-Oaten (1989) went to, perhaps, the opposite extreme by constructing a continuous-time model in which prey moved instantly into patches to replace prey that had been consumed, and predators moved instantly into patches to maintain a consistent pattern of predator–prey covariation over space. Here, the link between aggregation and stability is much less clear. Other, less 'extreme' continuous-time formulations (Ives, 1992), or those that combine discrete generations with redistribution within generations (Rohani *et al*., 1994), produce results that are themselves intermediate between the ‘Nicholson–Bailey extreme’ and the ‘Murdoch–Stewart-Oaten extreme’. Overall, models suggest a link between spatial aggregation and stability, but that link is neither certain nor straightforward.

10.5.3 Patch and lattice models

These continuous- and discrete-time approaches share a common perspective in seeing predator–prey interactions occurring within single populations, albeit populations with inbuilt variability. Many other models have been constructed that share more of a ‘metapopulation’ perspective (see Section 6.7), in which environmental patches support subpopulations that have their own internal dynamics, but are linked to other subpopulations by movement between patches. These can either be ‘patch’ models, in which subpopulations (patches) have no specific spatial location and so are equally connected to all other patches, or ‘lattice’ models, in which the patches do have a defined spatial arrangement and dispersal occurs only between neighbouring patches.

Briggs and Hoopes (2004) provide a useful review of stability in these and other spatial predator–prey models. There are many such models, exploring various aspects of within-patch dynamics and between-patch patterns of movement – very many more than there are studies that would allow those models to be tested with real data. But there are a number of key, recurring messages. The first is that patchiness and dispersal in their own right have no effect on the stability of the overall dynamics: if the patches are the same, and dispersal is uniform, stability is unaffected.

asynchrony between patches and limited dispersal combine to enhance stability

Overall, however, models with separate patches and dispersal between them do tend to have enhanced stability compared with the dynamics in individual patches (less violent fluctuations and a lower probability of extinction). The critical requirement for this is that there should be asynchrony between the dynamics of different patches. This leads, in the first place, to a simple statistical averaging: combining the highs and lows from different patches creates a time series in which the overall highs and lows are inevitably less extreme. In addition, a population at the peak of its cycle tends to lose more by dispersal than it gains (from other patches that are not at their peak), whereas a population in a trough tends to gain more than it loses. Dispersal and asynchrony together, therefore, give rise to net migration rates for patches that change in a density-dependent manner over time and hence stabilise the dynamics of individual patches. It follows from this that characteristics of patches that promote asynchrony themselves tend to stabilise the predator–prey dynamics of the overall system. Examples include fixed differences between the characteristics of the patches (and hence different dynamics within them) and initial differences in densities.

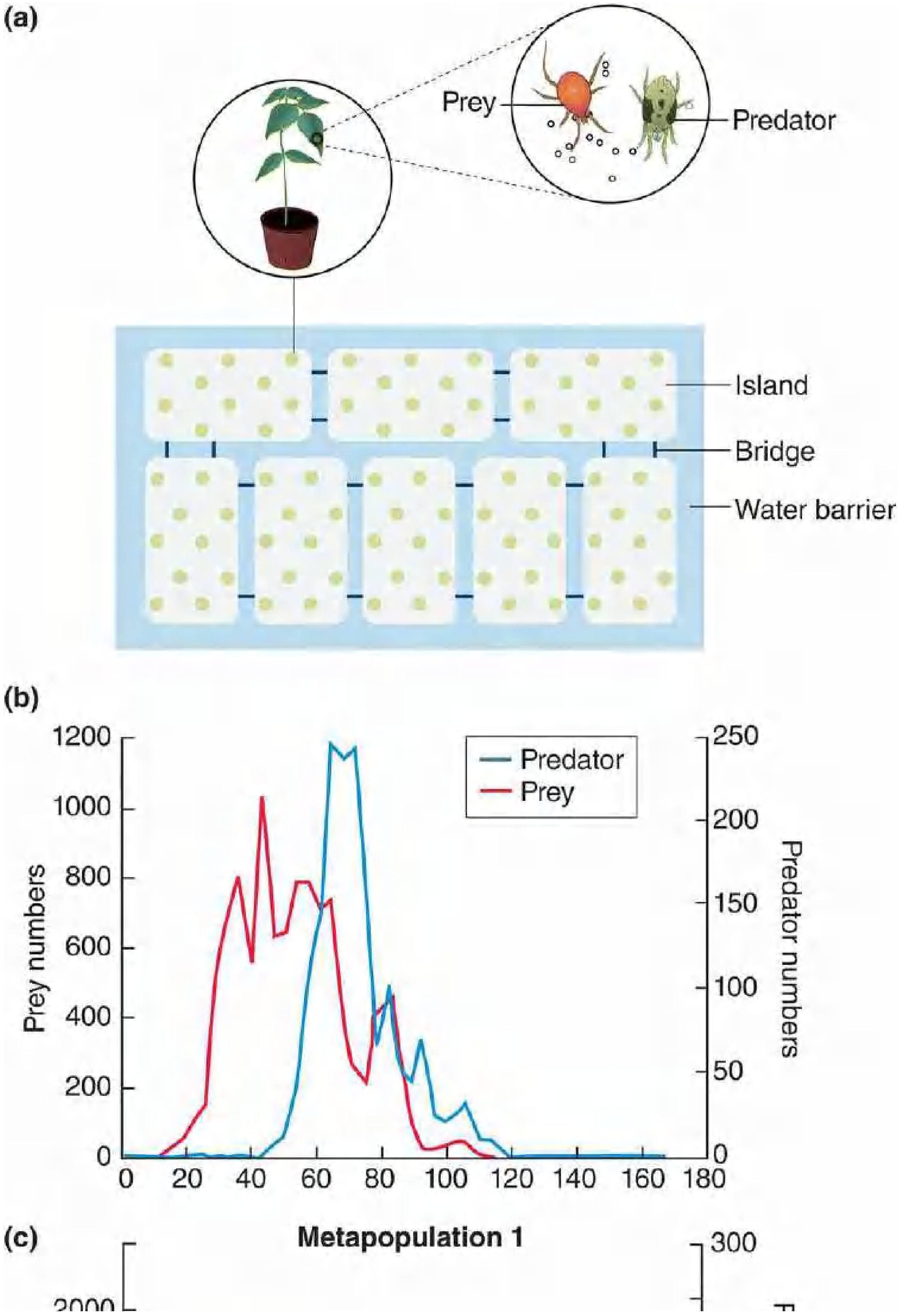
On the other hand, rates of dispersal between patches that are too high may ultimately lead to the whole metapopulation behaving like a single, unitary population, reducing differences between patch densities, synchronising their dynamics, and reducing stability. The very fact that a population is fragmented into subpopulations may also, in itself, contribute to a reduction in stability or persistence, since each fragment supports small populations, in which troughs in abundance are more likely, by chance, to become local extinctions. In this case, though, increased migration may enhance stability by migrants from persistent patches rescuing patches that have gone extinct. As a general rule, intermediate levels of movement between patches seem most likely to enhance overall persistence of a metapopulation.

10.5.4 Aggregation, heterogeneity and spatial variation in practice

metapopulation effects in mites, beetles and ciliates

What, then, can be said about the role of spatial variation in practice?

At the population level, the stabilising effects of heterogeneity were demonstrated famously, long ago, by Huffaker (Huffaker, 1958 ; Huffaker *et al*., 1963), who studied a system in which a predatory mite fed on a herbivorous mite, which fed on oranges interspersed amongst rubber balls in a tray. In a similar vein, Janssen *et al*. (1997) studied a laboratory system in which a predatory mite *Phytoseiulus persimilis* fed on a phytophagous mite, its prey, *Tetranychus urticae*. This latter fed on small bean plants collected together on styrofoam 'islands' separated by water, but with bridges connecting them across which the mites could walk: either a single large island containing 90 bean plants, or eight smaller islands each containing 10 plants (Figure 10.26a). Islands were thus connected, but their dynamics were at least semi-independent since the rate of transfer of mites between islands was low. When all the bean plants were collected together on a single island, the abundance of the prey mite increased, followed, after a delay, by a corresponding rise in the numbers of predators, which then led to a crash in the prey population, followed by a crash of the predators (Figure 10.26b). There was therefore only a single 'cycle' of predator and prey abundance, with the whole system lasting for just 120 days. The underlying predator–prey dynamics were unstable.


(a)
Prey
Predator
Island
Bridge
Water barrier
(b)
Predator
Prey
Prey numbers
Predator numbers
1200
1000
800
600
400
200
0
250
200
150
100
50
0
0
20
40
60
80
100
120
140
160
180
(c)
Metapopulation 1
300
2000

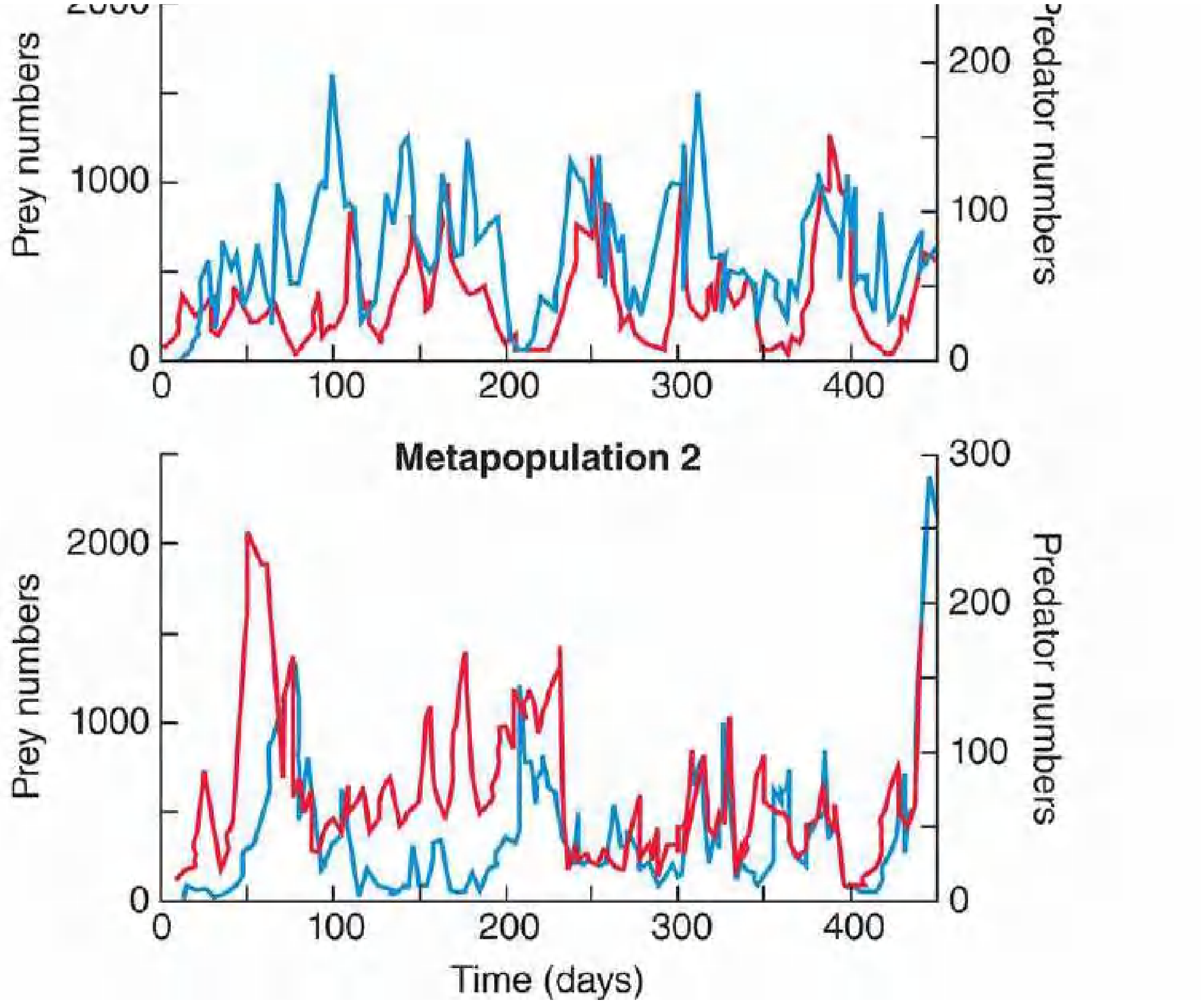


Figure 10.26 A metapopulation structure increases the persistence of predator–prey interactions for two species of mite. (a) An experimental set-up in which bean plants were collected together on styrofoam islands, surrounded by water but connected by bridges. The plants were fed upon by the mite *Phytoseiulus persimilis*, which was in turn preyed upon by another mite, *Tetranychus urticae*. (b) On a single island containing 90 plants, prey numbers rose, followed by predator numbers, after which prey numbers declined to extinction, again followed by the predator. (c) With the metapopulation structure shown in (a), both predator and prey persisted throughout the experimental period.

Source: After Janssen *et al.* (1997).

When the habitat was broken into eight small islands, however, in two separate runs of the system, both predator and prey persisted for more than a year – indeed until the experiments were terminated. In one case, there was clear evidence of persistent cycles in both predator and prey; in the other, the dynamics were much more erratic but still persistent (Figure 10.26c). Notably, though, on the individual small islands, there was not one example of the predator and prey populations persisting. Each went extinct at least once, requiring colonising mites walking over from an occupied island to rescue that particular population. Predators and prey were therefore ultimately doomed to extinction in each patch (island) – that is, the patch dynamics were unstable. But overall, at any one time, there was a mosaic of unoccupied patches, prey–predator patches heading for extinction, and thriving prey patches; and this mosaic was capable of maintaining persistent populations of both predators and prey, providing stability to the metapopulation as a whole (Figure 10.26c).

Others, too, have demonstrated the power of a metapopulation structure in promoting the persistence of coupled predator and prey populations when their dynamics in individual subpopulations are unstable. Figure 10.27a, for example, shows this for a parasitoid attacking its beetle host. Figure 10.27b shows similar results for prey and predatory ciliates (protists), where, in support of the role of a metapopulation structure, it was also possible to demonstrate asynchrony in the dynamics of individual subpopulations and frequent local prey extinctions and recolonisations (Holyoak & Lawler, 1996).

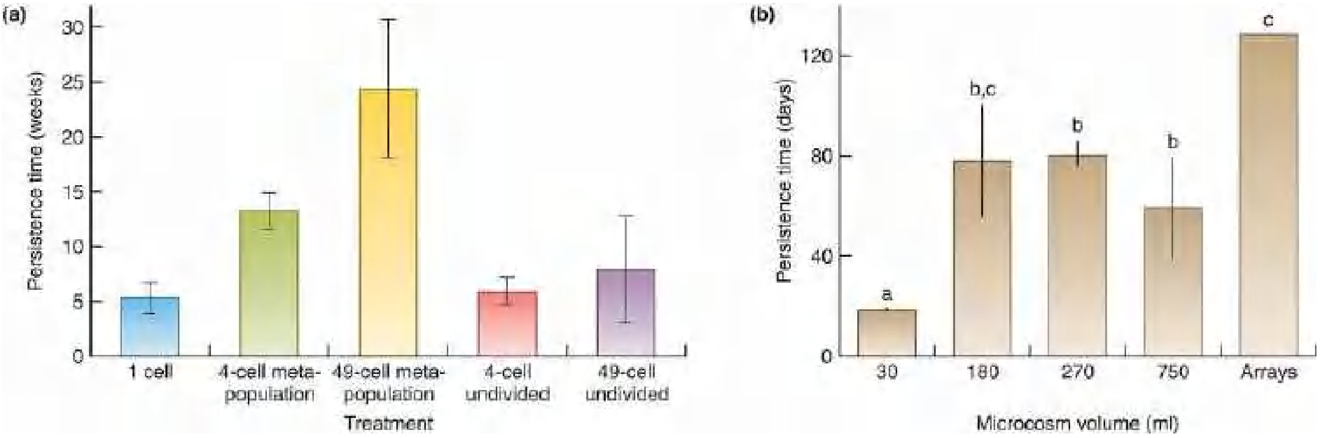


Figure 10.27 A metapopulation structure increases the persistence of predator–prey interactions for a parasitoid and its host. (a) The parasitoid *Anisopteromalus calandrae* attacking its bruchid beetle host *Callosobruchus chinensis*, living on beans either in small single ‘cells’ (short persistence time, left), or in combinations of cells (four or 49), which either had free access between them so that they effectively constituted a single population (persistence time not significantly increased, right), or had limited (infrequent) movement between cells so that they constituted a metapopulation of separate subpopulations (increased persistence time, centre). Bars are SEs. (b) The predatory ciliate *Didinium nasutum* feeding on the bacterivorous ciliate *Colpidium striatum* in bottles of various volumes, where persistence time varied little, except in the smallest populations (30 ml) where times were shorter, and also in ‘arrays’ of nine or 25 linked 30 ml bottles (metapopulations), where persistence was greatly prolonged: all populations persisted until the end of the experiment (130 days). Bars are SEs; different letters above bars indicate treatments that were significantly different from one another ($P < 0.05$).

Source: (a) After Bonsall *et al.* (2002). (b) After Holyoak & Lawler (1996).

Another study of ciliates focused more on the degree of fragmentation of the populations and the movement of animals between them. The set-up is shown in Figure 10.28. Populations of predators, *Didinium nasutum*, and prey, *Paramecium caudatum*, were followed in their liquid medium, first, in populations of different volumes: 4, 8, 16, 32 or 64 ml. Then, to include the effects of dispersal between such populations, they were monitored in either a single undivided population (64 ml), or with that population split into 2, 4, 8 or 16 equal-sized fragments, all adding to the same total size (64 ml), and with movement between fragments occurring through corridors that linked them (Figure 10.28a). All populations were replicated, usually five times, and all populations or subpopulations were seeded with one prey and 0.25 predators per millilitre. Dispersal rates were also measured, showing that prey dispersed more than four times as fast as the predators.

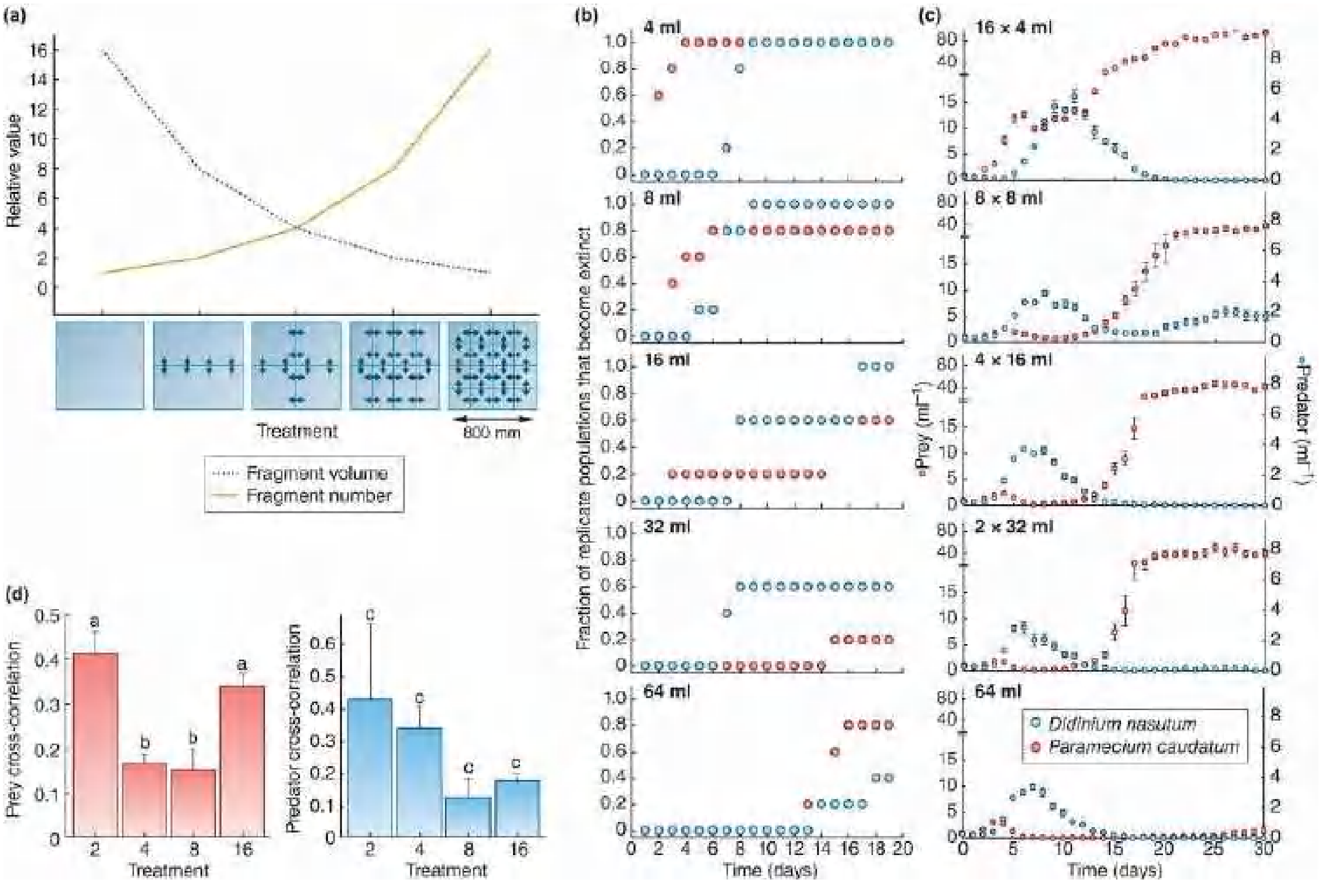


Figure 10.28 **The stability (persistence) of a ciliate predator–prey metapopulation is greatest at intermediate levels of fragmentation.** (a) The experimental set-up for predator and prey ciliates, *Didinium nasutum* and *Paramecium caudatum*, respectively, living in volumes of culture medium. The arrangement of subpopulations, with dispersal corridors (small blue arrows) between them, is shown below. Above that, relative values are shown for the fragment volume and the number of fragments. (b) The proportions of populations, similar to those in (a) but without corridors, that went extinct over time. From top to bottom: 4 ml, 8 ml, 16 ml, 32 ml and 64 ml. The smallest populations persisted least well. (c) Population trajectories over time (± SE) for the populations and subpopulations in (a). From top to bottom: 16 × 4 ml, 8 × 8 ml, 4 × 16 ml, 2 × 32 ml and 64 ml. Persistence was best with intermediate fragmentation (8 × 8 ml). (d) The cross correlations (± SE) between the numbers of prey (left) and predators (right) in the different subpopulations (fragments). Lower correlations equate to greater asynchrony. Bars with different symbols are significantly different from one another. Hence asynchrony was greatest with 8 × 8 ml fragments. (For predators, asynchrony increased with fragmentation if one outlying value from the two-fragment treatment was omitted.)

Source: After Cooper *et al.* (2012).

When the populations existed in isolation (no dispersal), extinction was clearly most likely in the smallest populations (Figure 10.28b). However, when there were corridors between the subpopulations, the pattern changed. In the unfragmented populations, both predator and prey declined to extinction in less than 15 days, with the exception of one replicate where the prey persisted (Figure 10.28c). At the other extreme, with 16 subpopulations, but also with two and four subpopulations, the predator population went extinct around this time, but then the prey population, released from predation and with higher dispersal rates, established a more or less steady population size thereafter (Figure 10.28c). Only at the intermediate level of fragmentation, with eight subpopulations, did both predator and prey coexist throughout the study. It is also noteworthy that asynchrony between the dynamics of the different subpopulations was greatest when there were eight fragments (Figure 10.28d). What we seem to see, therefore, is persistence arising out of a balance between the different stabilising and destabilising forces. The most fragmented populations were rendered unstable by the small size of their subpopulations and the homogenising effects of high dispersal. The less- and unfragmented populations failed to gain sufficient stability from their combinations of population asynchrony and dispersal. But with eight fragments, for this system at least, stabilising and destabilising forces were balanced.

real data confirm the complexity of natural systems

Such elaboration for the relationships between heterogeneity and stability in laboratory systems is valuable, and support from them for mathematical models is encouraging. But ultimately, the only thing that truly matters is what happens in nature. One major problem in making pronouncements about the stabilising role of aggregation of risk in natural systems is that although, as we have seen, there have been wide-ranging surveys of the data on spatial distributions of attacks, these data generally come from studies of very short duration – often of only one generation. We do not know if the observed spatial patterns are typical for that interaction; nor do we know if the population dynamics show the degree of stability that the spatial patterns might seem to predict. One investigation that did examine population dynamics and spatial distributions over several generations was that of Redfern *et al*. (1992), who made a seven-year (seven-generation) study of two tephritid fly species that attack thistles, and the parasitoids that attack those flies. For one fly host, *Terellia serratulae* (Figure 10.29a), there was evidence of year-to-year density dependence in the overall rate of parasitism (Figure 10.29b), but no strong evidence of significant levels of aggregation within generations, either overall (Figure 10.29c) or for parasitoid species individually. For the other species, *Urophora stylata* (Figure 10.29d), there was no apparent temporal density dependence but good evidence for the aggregation of risk (Figure 10.29e, f). And to repeat a pattern we have seen before, most heterogeneity was independent of host density. It cannot be said, however, that the patterns of this study fit neatly, overall, to the theory we have outlined. Firstly, both hosts were attacked by several parasitoid species – not one, as assumed by most models. Secondly, the levels of aggregation varied considerably and apparently randomly from year to year (Figure 10.29c, f): no one year was typical, and no single 'snap-shot' could have captured either interaction. Finally, while the relatively stable dynamics of *Terellia* may have reflected the more demonstrable direct density dependence in parasitism, this appeared to be quite unconnected to any differences in the aggregation of risk.

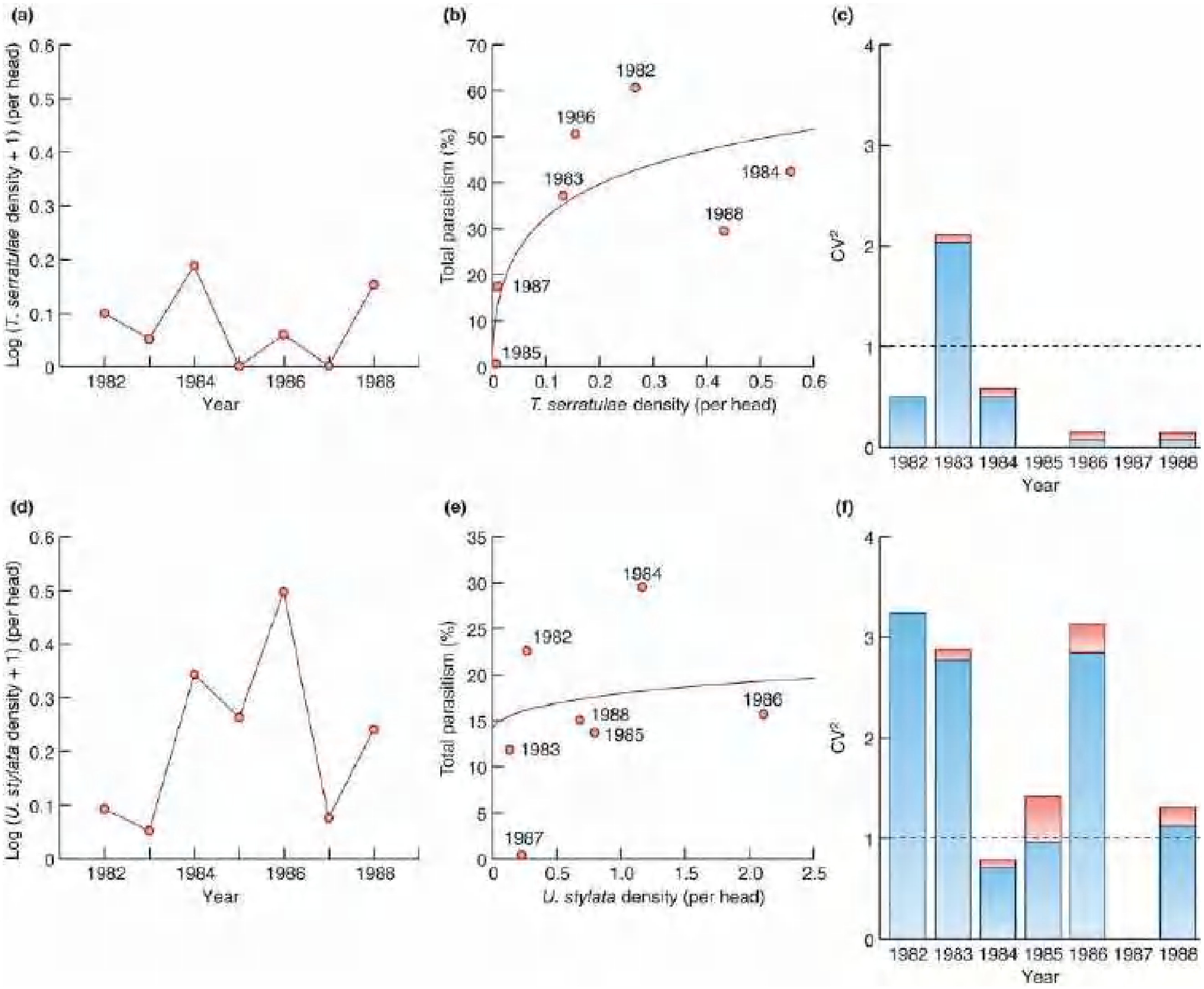


Figure 10.29 The complex interactions between density-dependence, aggregation of risk and stability in natural systems. Attacks by parasitoids on tephrytid flies (*Terellia serratulae* and *Urophora stylata*) that attack thistle flower-heads. The dynamics of the populations are shown for *T. serratulae* in (a) and for *U. stylata* in (d). Temporal density dependence of parasitoid attacks on *T. serratulae* (b) is significant ($r^2 = 0.75$; $P < 0.05$), but for *U. stylata* (e) it is not ($r^2 = 0.44$; $P > 0.05$); both fitted lines take the form $y = a + b \log_{10}x$. However, whereas for *T. serratulae* (c) there is little aggregation of risk of parasitoid attack within years (measured as $CV^2 > 1$ for aggregation), with *U. stylata* (f) there is far more, most of which is independent of host density (blue shading) rather than being host-density dependent (red shading).

Source: After Redfern *et al*. (1992).

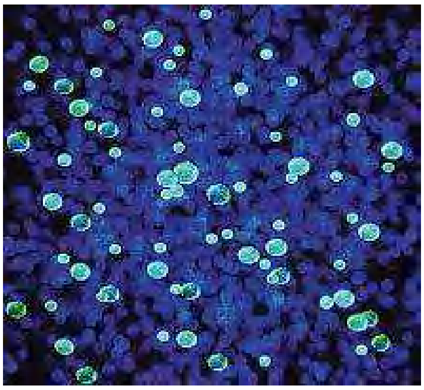
APPLICATION 10.3 What's required of a good biological control agent?

The effects of spatial heterogeneities on the stability of predator–prey dynamics are not only of purely scientific interest. They have also been the subject of lively debate (Hawkins & Cornell, 1999) in considering the properties and nature of biological control agents: natural enemies of a pest that are imported into an area, or otherwise aided and abetted, in order to control the pest (see Section 15.2.4). What is required of a good biological control agent is the ability to reduce the prey (pest) to a stable abundance well below its normal, harmful level, and we have seen that some theoretical analyses suggest that this is precisely what aggregative responses help to generate. Establishing such a link in practice, however, has not proved easy. Indeed, in their review of potentially desirable properties of biological control agents, especially parasitoids, Snyder and Ives (2009) conclude: "Despite a lot of interest in the ecological literature, we know of no clear empirical demonstrations of aggregation stabilising host–parasitoid dynamics, and there are considerable logistical difficulties to testing this idea". Clearly, with biological control, like predator–prey dynamics generally, building convincing links between patterns in population stability of natural populations and particular stabilising mechanisms – or combinations of mechanisms – remains a challenge for the future.

10.6 Beyond predator–prey

The simplest mathematical models of predator–prey interactions produce coupled oscillations that are highly unstable. However, by adding various elements of realism to these models it is possible to reveal the features of real predator–prey relationships that are likely to contribute to their stability. We have seen that a variety of patterns in the abundance of predators and prey, both in nature and in the laboratory, are consistent with the conclusions derived from models. Unfortunately, we are rarely in a position to apply specific explanations to particular sets of data, because the critical experiments and observations to test the models have rarely been made. Natural populations are affected not just by their predators or their prey, but also by many other environmental factors that serve to 'muddy the waters' when direct comparisons are made with simple models.

Moreover, the attention of both modellers and data gatherers (not that the two need be different) is increasingly being directed away from single- or two-species systems, towards those in which three species interact. For example, a pathogen attacking a predator that attacks a prey, or a parasitoid and a pathogen both attacking a prey/host. Interestingly, in several of these systems, unexpected dynamical properties emerge that are not just the expected blend of the component two-species interactions (Holt, 1997). We return to the problems of 'abundance' in a broader context in Chapter 14 and to 'community modules', comprised of more than two interacting species, in Chapter 16.



Chapter 11

Decomposers and Detritivores

11.1 Introduction

saprotrophs: detritivores and decomposers ...

When plants and animals die, their bodies become resources for other organisms. Of course, in a sense, most consumers live on dead material – the carnivore catches and kills its prey, and the living leaf taken by a herbivore is dead by the time digestion starts. The critical distinction between the organisms in this chapter, and herbivores, carnivores and parasites, is that the latter all directly affect the rate at which their resources are produced. Whether it is lions eating gazelles, gazelles eating grass or grass parasitised by a rust fungus, the act of taking the resource harms the resource's ability to regenerate new resource (more gazelles or grass leaves). By contrast, saprotrophs (organisms that make use of dead organic matter) do not control the rate at which their resources are made available or regenerate; they are dependent on the rate at which some other force (senescence, illness, fighting, the shedding of leaves by trees) releases the resource on which they live. (Necrotrophic parasites (see Section 12.2) are both heterotrophs and saprotrophs. They kill their host and then continue to extract resources from its dead body. Thus, the fungus *Botrytis cinerea* attacks living bean leaves but continues this attack after the host's death; while maggots of the sheep blowfly *Lucilia cuprina* may parasitise and kill their host, but then they continue to feed on the corpse.)

... do not generally control their supply of resources – 'donor control'

We distinguish two groups of saprotrophs: decomposers (the prokaryotic bacteria and archaea together with the eukaryotic fungi) and detritivores (animals that consume dead matter). Pimm (1982) described the relationship that generally exists between decomposers or detritivores and their food as *donor controlled*: the donor (prey; i.e. dead organic matter) controls the density of the recipient (predator; i.e. decomposer or detritivore) but not the reverse. This is fundamentally different from truly interactive predator–prey interactions. Indeed, while there is generally no direct negative feedback between decomposers/detritivores and the dead matter consumed (and thus donor-controlled dynamics apply), it is possible to see an indirect, positive 'mutualistic' effect through the release of nutrients from decomposing litter, which may ultimately *increase* the rate at which trees produce more litter. In fact, it is in nutrient recycling that decomposers and detritivores play their most fundamental role in ecosystems (see Chapter 21). In other respects, of course, the food webs associated with decomposition are just like food webs based on living plants: they have a number of trophic levels, including predators of decomposers and detritivores, and consumers of these predators, and thus they exhibit a range of trophic interactions (not just donor controlled).

decomposition defined

When an inorganic nutrient element is incorporated into an organic form – primarily during the growth of green plants – we refer to this as *immobilisation*. Conversely, decomposition involves the release of energy and the *mineralisation* of these chemical nutrients – the conversion of elements from an organic to inorganic form. Decomposition is defined as the gradual disintegration of dead organic matter and is brought about by both physical and biological agencies. Complex, energy-rich molecules are broken down by their consumers (decomposers and detritivores) into carbon dioxide, water and inorganic nutrients. Ultimately, the incorporation of solar energy in photosynthesis, and the immobilisation of inorganic nutrients into biomass, is balanced by the loss of heat energy and organic nutrients when the organic matter is mineralised. Thus, a given nutrient atom may be successively immobilised and mineralised in a repeated round of nutrient cycling. We discuss the overall role played by decomposers and detritivores in the fluxes of energy and nutrients at the ecosystem level in Chapters 20 and 21. In the present chapter, we introduce the organisms involved and look in detail at the ways in which they deal with their resources.

decomposition ... of dead bodies, ...

It is not only the bodies of dead animals and plants that serve as resources for decomposers and detritivores. Dead organic matter is continually produced during the life of both animals and plants and can be a major resource. Unitary organisms shed dead parts as they develop and grow – the larval skins of arthropods, the skins of snakes, the skin, hair, feathers and horn of other vertebrates. Specialist feeders are often associated with these cast-off resources. For example, there are specialist fungi that decompose feathers and horn, and sloughed off human skin is a resource for the household mites that are omnipresent inhabitants of house dust and cause problems for many allergy sufferers.

... of shed parts of organisms ...

The continual shedding of dead parts is even more characteristic of modular organisms. Some polyps on a colonial hydroid or coral die and decompose, while other parts of the same genet continue to regenerate new polyps. Most plants shed old leaves and grow new ones; the seasonal litter fall onto a forest floor is the most important of all the sources of resource for decomposers and detritivores, but the producers do not die in the process. Higher plants also continually slough off cells from the root caps, and root cortical cells die as a root grows through the soil. In addition, roots release soluble carbon-rich and nitrogen-rich compounds as exudates. This supply of organic material from roots produces the very resource-rich *rhizosphere*. Plant tissues are generally leaky, and soluble sugars and nitrogenous compounds also become available on the surface of leaves, supporting the growth of bacteria in particular, but also of archaea and fungi, in the *phyllosphere* (Whipps *et al.* 2008).

... and of faeces

Finally, animal faeces, whether produced by detritivores, microbivores, herbivores, carnivores or parasites, are a further category of resource for decomposers and detritivores. They are composed of dead organic material that is chemically related to what their producers have been eating.

The remainder of this chapter is in two parts. In Section 11.2 we describe the 'actors' in the saprotrophic 'play', and consider the relative roles of the bacteria, archaea and fungi on the one hand, and the detritivores on the other. Then, in Section 11.3, we consider, in turn, the problems and processes involved in the consumption by detritivores of plant detritus, faeces and carrion.

11.2 The organisms

11.2.1 Decomposers: bacteria, archaea and fungi

If scavengers do not take a dead resource immediately it dies (such as hyenas consuming a dead zebra), the process of decomposition usually starts with colonisation by bacteria, archaea and fungi. Other changes may occur at the same time: enzymes in the dead tissue may start to autolyse it and break down the carbohydrates and proteins into simpler, soluble forms. The dead material may also become leached by rainfall or, in an aquatic environment, may lose minerals and soluble organic compounds as they are washed out in solution.

bacteria, archaea and fungi are early colonists of newly dead material

Bacteria, archaea and fungal spores are omnipresent in the air, soil and water, and are usually present on (and often in) dead material before it is dead. They usually have first access to a resource. The early colonists tend to use soluble materials, mainly amino acids and sugars that are freely diffusible, but they lack the array of enzymes necessary for digesting structural materials such as cellulose, lignin, chitin and keratin. Many species of *Penicillium*, *Mucor* and *Rhizopus*, the so-called 'sugar fungi' in soil, grow fast in the early phases of decomposition. Together with bacteria and archaea having similar opportunistic physiologies, they tend to undergo population explosions on newly dead substrates, but because they are either immobile or move only passively through the soil, these populations collapse when the freely available resources are consumed, leaving very high densities of resting stages from which new population explosions may develop when another freshly dead resource becomes available. Hence, they may be thought of as the opportunist '*r*-selected species' among the decomposers (see Section 7.6.1). Many fungi, especially basidiomycetes, have an advantage over the immobile bacteria in the ability of their mycelial network to grow out from a colonised substrate in search of new resources (Boddy, 1999).

domestic and industrial decomposition

In nature, as in industrial processes such as the making of wine or sauerkraut, the activity of the early colonisers is dominated by the metabolism of sugars and is strongly influenced by aeration. When oxygen is in free supply, sugars are metabolised to carbon dioxide by growing microbes. Under anaerobic conditions, however, fermentations produce a less complete breakdown of sugars to by-products such as alcohol and organic acids that change the nature of the environment for subsequent colonisers. In particular, the lowering of the pH by the production of acids has the effect of favouring fungal as opposed to bacterial activity.

aerobic and anaerobic decomposition in nature

Anoxic habitats (those lacking oxygen) are characteristic of waterlogged soils and, more particularly, of sediments of oceans, estuaries and lakes. Aquatic sediments receive a continuous supply of dead organic matter from the water column above, but aerobic decomposition (mainly by bacteria and archaea) quickly exhausts the available oxygen because this can only be supplied from the surface of the sediment by diffusion. Thus, at some depth, from zero to a few centimetres below the surface, depending mainly on the load of organic material, sediments are completely anoxic. Sediments in the shallow 'shelf' waters close to land are richly supplied with organic matter, and the top layers are characterised by rapid depletion of oxygen (millimetres to centimetres). In sediments in the deeper ocean, and particularly in the abyssal depths, inputs of organic matter are low, oxygen is consumed slowly and it may diffuse to metres in depth (Figure 11.1). Below the oxic layer are found a variety of types of bacteria and archaea that employ various forms of anaerobic respiration (that is, they use terminal inorganic electron acceptors other than oxygen in their respiratory process – see Section 3.6). The types occur in a predictable pattern below the oxygen respiration zone, with denitrifying types first, metal oxide reducers next, then sulphate reducers and finally methanogenic archaea in the deepest zone (Orcutt *et al.*, 2011). Sulphate is comparatively abundant in seawater and so the zone of sulphate-reducing bacteria and archaea is particularly wide: in contrast, the concentration of sulphate in lakes is low, and methanogenesis by archaea plays a correspondingly larger role (Holmer & Storkholm, 2001).

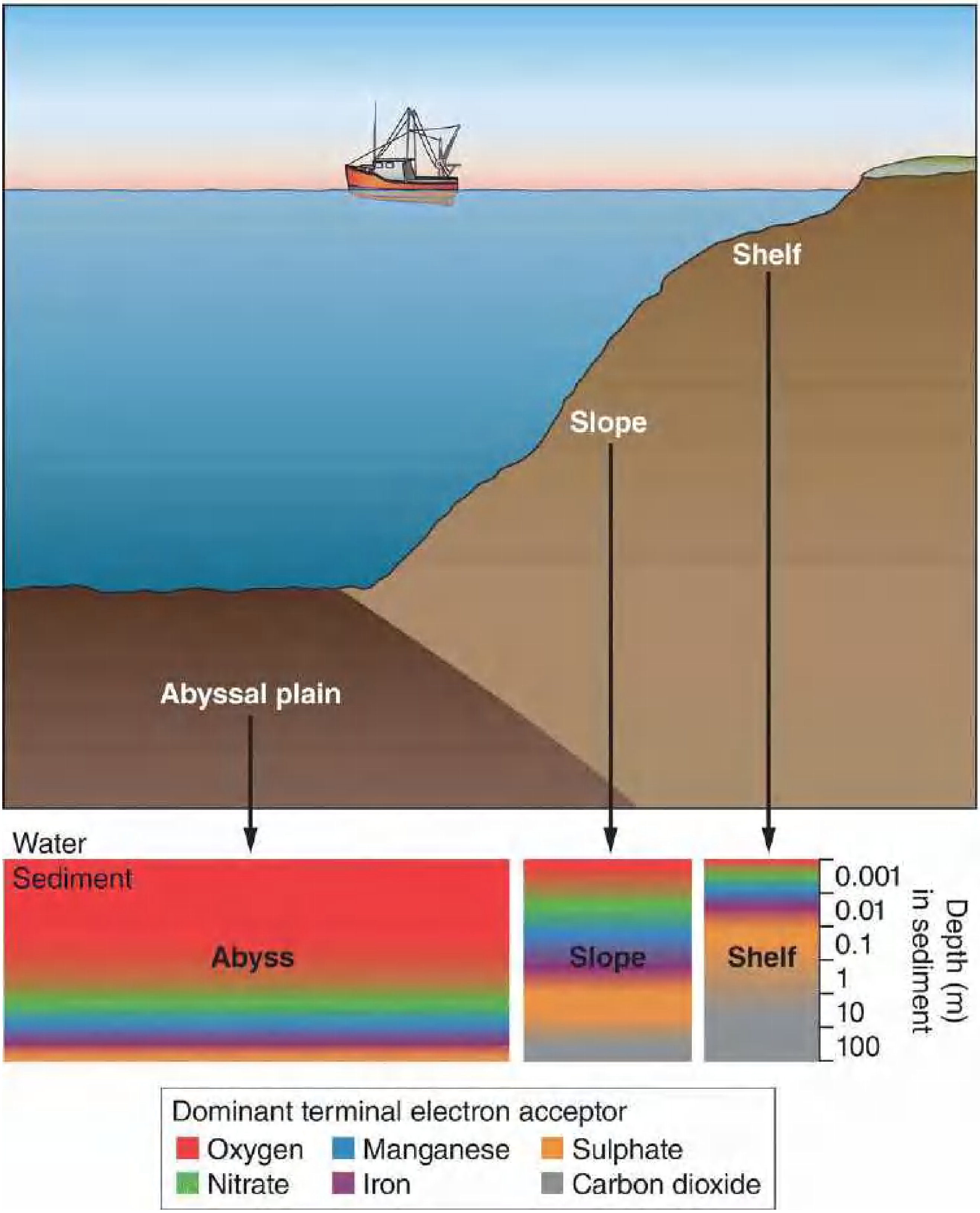


Figure 11.1 Zonation of microbial communities in marine sediment habitats. Schematic cross-section of marine sediment habitats (shallow shelf, slope and deep abyss) with representations of zonation of microbial communitites according to the dominant terminal electron acceptor used. Note that Archaea consume carbon dioxide during methanogenesis.

Source: After Orcutt *et al*. (2011).

decomposition of more resistant tissues proceeds more slowly … or not at all

In broad terms, the major components of dead terrestrial organic matter are, in a sequence of increasing resistance to decomposition: sugars < (less resistant than) starch < hemicelluloses, pectins and proteins < cellulose < lignins < suberins < cutins. Hence, after an initial rapid breakdown of sugar, decomposition proceeds more slowly, and involves microbial specialists that can use celluloses and lignins and break down the more complex proteins, suberin (cork) and cuticles. These are structural compounds, and their breakdown and metabolism depend on very intimate contact with the decomposers (most cellulases are surface enzymes requiring actual physical contact between the decomposer organism and its resource). The processes of decomposition may now depend on the rate at which fungal hyphae can penetrate from cell to cell through lignified cell walls. Indeed, more generally, we should not overemphasise the success with which decomposers and detritivores deal with their resources. It is, after all, the failure of organisms to decompose wood rapidly that makes the existence of forests possible. Deposits of peat, coal and oil are further testaments to the failures of decomposition.

succession of decomposing microorganisms

The changing nature of a resource during its decomposition is illustrated in Figure 11.2a for oak leaf litter (*Quercus petraea*) on the floor of a cool temperate deciduous forest in the Czech Republic. Overall leaf mass loss was 68% after two years of decomposition. In the first four months, pectins and hemicelluloses were most rapidly decomposed, as reflected by the pattern of activity of the extracellular enzymes responsible (e.g. β-glucosidase). In the intermediate phase (4–12 months), cellulose was most rapidly decomposed (reflected by endocellulase activity), while lignin was decomposed during both the intermediate and late phases (reflected by the activity of ligninolytic enzymes such as laccase) (Figure 11.2a). Fungal biomass (measured as content of ergosterol, a sterol found in fungal cell walls) peaked at about four months and remained at quite high levels for the rest of the study. Fungi and bacteria were both involved in decomposition, with fungi playing a relatively greater role than bacteria in the first eight months, as illustrated by the ratio of ergosterol to bacterial PLFA (phospholipid fatty acids specific to bacteria).

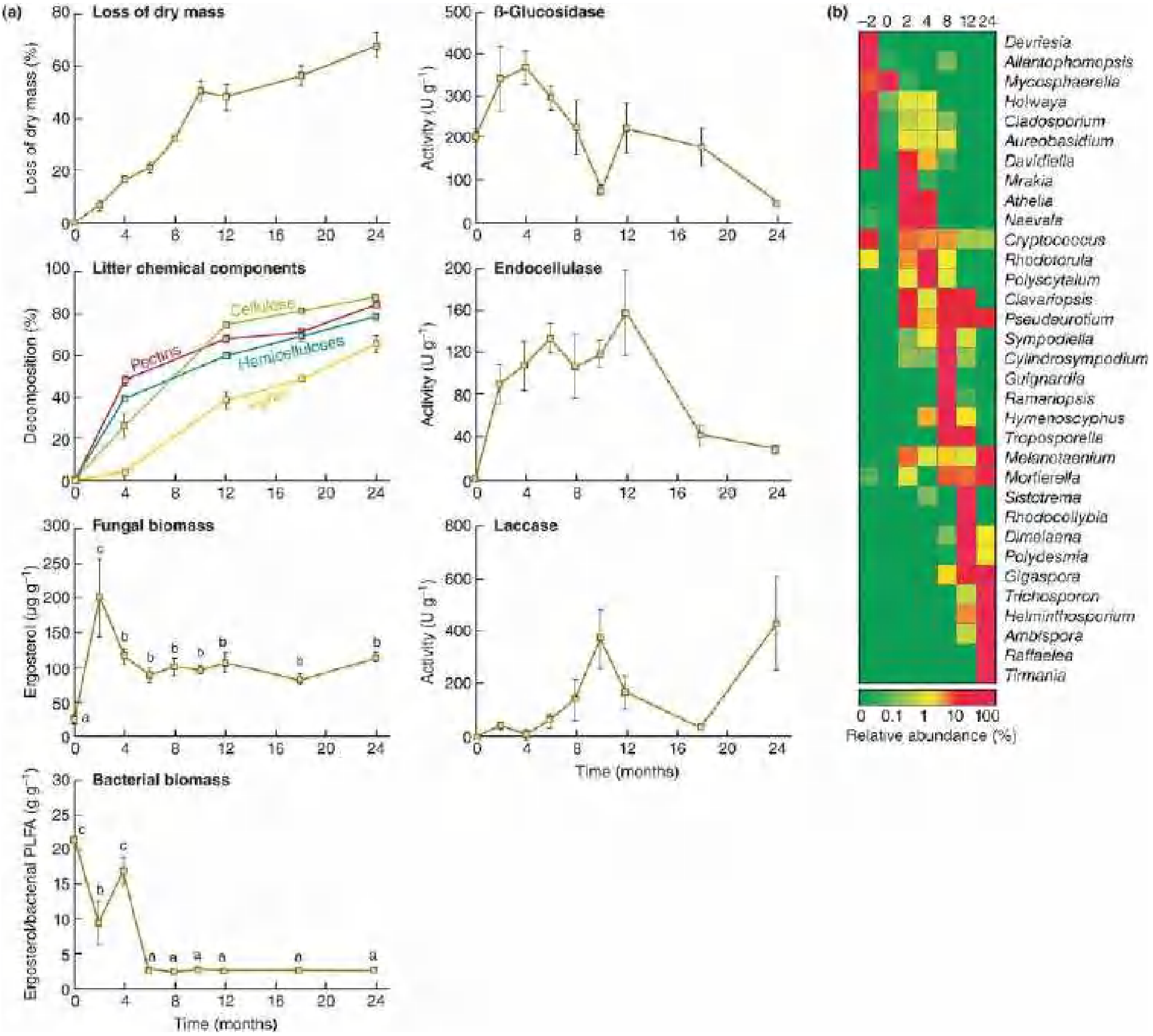


Figure 11.2 Changes in the chemical composition of oak leaf litter and its associated fungal community during 24 months of monitoring. (a) Patterns in leaf dry mass loss, litter chemical components, fungal (ergosterol) and bacterial biomass (bacterial PLFA (phospholipid fatty acids specific to bacteria)), and the activities of three extracellular enzymes, during a two-year study of decomposition of oak leaves. Bars are SEs; bars with different letters are significantly different. (b) Time course of the relative abundance of dominant fungal genera from two months before leaf death (–2) to 24 months afterwards.

Source: After Snajdr *et al.* (2010) and Voriskova & Baldrian (2013).

Some fungi are already present on and in living leaves, but continue to play a role after their death. Phyllosphere fungi such as *Holwaya* spp. were present two months before oak leaf abscission but were still active some months later, presumably taking the opportunity to exploit leaf-derived nutrients rapidly during decomposition (Figure 11.2b). In the first year of decomposition, species in the phylum Ascomycota, generally known to be able to selectively decompose cellulose over lignin, were prominent (e.g. *Clavariopsis*, *Pseudeurotium*). But by the end of the second year, species of Basidiomycota, able to deal with lignin, tended to dominate (e.g. *Melanotaenium*, *Trichosporon*).

Between them, a varied microbiota of bacteria, archaea and fungi can accomplish the complete degradation of dead material of both plants and animals. However, in practice they seldom act alone, and the process would be much slower and, moreover, incomplete, if they did so. The major factor that delays the decomposition of organic residues is the resistance to decomposition of plant cell walls – an invading decomposer meets far fewer barriers in an animal body. The process of plant decomposition is enormously speeded up by any activity that grinds up and fragments the tissues, such as the chewing action of detritivores. This breaks open cells and exposes the contents and the surfaces of cell walls to attack.

11.2.2 Detritivores and specialist microbivores

specialist consumers of microbial organisms: microbivores

Detritivores, as we have noted, consume dead organic matter, but operating alongside them, and often difficult to distinguish from them, are the microbivores: minute animals that specialise at feeding on the microbiota that lives on the dead matter but are able to exclude detritus from their guts. Note that microbivores consume a living resource and are thus unlikely to be subject to donor-controlled dynamics. There are two major groups of microbiota, and their exploitation requires quite different feeding techniques, principally because of differences in growth form. Bacteria and archaea (and yeasts) show a colonial growth form, usually on the surface of small particles. Specialist consumers of these are inevitably very small. They include free-living protozoans such as amoebae, and the terrestrial nematode *Pelodera*, which grazes among sediment particles consuming the bacteria on their surfaces. In contrast, the majority of fungi are filamentous, producing extensively branching hyphae, which in many species are capable of penetrating organic matter. Some specialist consumers of fungi possess piercing, sucking stylets (e.g. the nematode *Ditylenchus*) that they insert into individual fungal hyphae. However, most fungivorous animals graze on the hyphae and consume them whole. In some cases, close mutualistic relationships exist between fungivorous beetles, ants and termites and characteristic species of fungi. These mutualisms are discussed in Section 13.4.3.

In an eight-week study of litter decomposition in a Panama rainforest, Milton and Kaspari (2007) manipulated access of microbivores to the decomposing substrate by using fine-meshed litter bags. They found that microbivore presence was associated initially with lower rates of leaf decomposition, but later microbivores increased decomposition rates. They suggest that during the early colonisation phase of litter, microbivores may keep decomposition in check by recruiting faster than bacteria and fungi. Later, however, microbivores may facilitate decomposition by cropping senescent hyphae.

The larger the animal, the less able it is to distinguish between microbiota as food and the plant or animal detritus on which these are growing. In fact, the majority of detritivores are generalist consumers, of both the detritus itself and the associated microbial populations.

classification of decomposers ... by size in terrestrial environments ...

The protists and invertebrates that take part in the decomposition of dead plant and animal materials are a taxonomically diverse group. In terrestrial environments they are usually classified according to their size. This is not an arbitrary basis for classification, because size is an important feature for organisms that reach their resources by burrowing or crawling among cracks and crevices of litter or soil. The *microfauna* (including the specialist microbivores) includes protozoans, nematode worms and rotifers (Figure 11.3). The principal groups of the *mesofauna* (animals with a body width between 100 μm and 2 mm) are litter mites (Acari), springtails (Collembola) and pot worms (Enchytraeidae). The *macrofauna* (2–20 mm body width) and, lastly, the *megafauna* (>20 mm) include woodlice (Isopoda), millipedes (Diplopoda), earthworms (Megadrili), snails and slugs (Mollusca) and the larvae of certain flies (Diptera) and beetles (Coleoptera). These animals are mainly responsible for the initial shredding of plant remains. By their action, they may bring about a large-scale redistribution of detritus and thus contribute directly to the development of soil structure.

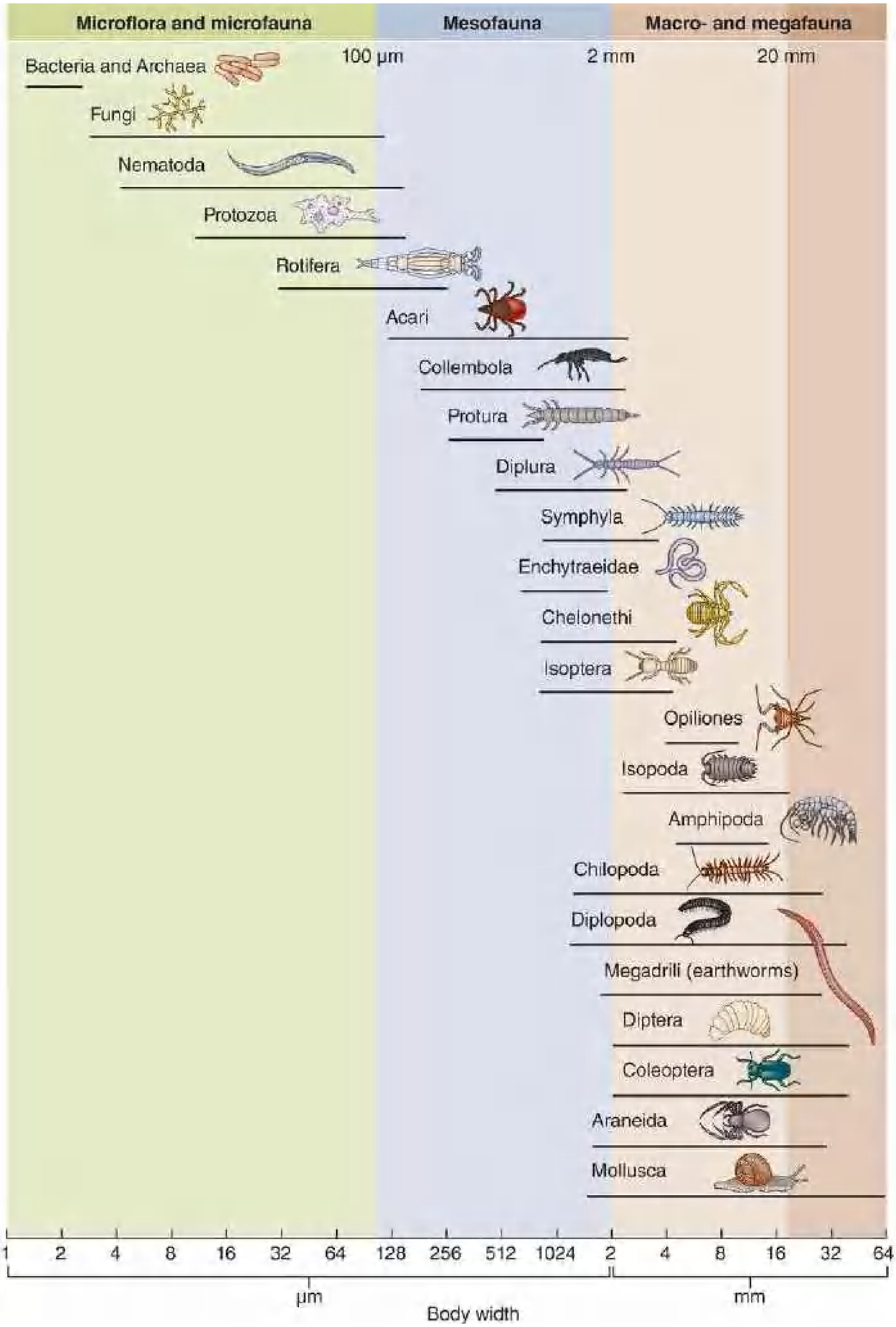

Microflora and microfauna
Mesofauna
Macro- and megafauna
100 µm
2 mm
20 mm
Bacteria and Archaea
Fungi
Nematoda
Protozoa
Rotifera
Acari
Collembola
Protura
Diplura
Symphyla
Enchytraeidae
Chelonethi
Isoptera
Opiliones
Isopoda
Amphipoda
Chilopoda
Diplopoda
Megadrili (earthworms)
Diptera
Coleoptera
Araneida
Mollusca
1
2
4
8
16
32
64
128
256
512
1024
2
4
8
16
32
64
µm
mm
Body width

Figure 11.3 Size classification by body width of organisms in terrestrial decomposer food webs. The following groups are wholly carnivorous: Opiliones (harvest spiders), Chilopoda (centipedes) and Araneida (spiders).

Source: After Swift *et al*. (1979).

APPLICATION 11.1 The importance of earthworms

Long ago, Charles Darwin (1888) estimated that earthworms in some pastures close to his house formed a new layer of soil 18 cm deep in 30 years, bringing about 50 tons ha^{-1} to the soil surface each year as worm casts. Figures of this order of magnitude have since been confirmed on a number of occasions (Feller *et al.*, 2003). Moreover, not all species of earthworm put their casts above ground, so the total amount of soil and organic matter that they move may be much greater than this. Where earthworms are abundant, they bury litter, mix it with the soil (and so expose it to other detritivores and decomposers), create burrows (so increasing soil aeration and drainage) and deposit faeces rich in organic matter. They play a prominent role in soil formation, accelerate the rate of conversion of leaf litter to humus, reduce loss of water in runoff, speed up mineralisation and increase nutrient availability to plants, and may even speed up soil remediation after pollution events (Blouin *et al.*, 2013). Earthworms are generally absent from acid soils and do not occur in deserts or arid grasslands where water is limiting.

In a meta-analysis of 58 studies exploring earthworm effects on crop production, van Groenigen *et al.* (2014) found that on average their presence is associated with a 25% increase in crop yield (Figure 11.4a). The positive effects of earthworms were smaller where rates of application of nitrogen fertiliser were high (Figure 11.4b) and in pastures where legumes were present (these can fix atmospheric nitrogen and so enhance nitrogen availability: Section 13.11). The most positive effects of earthworms occurred where the agricultural practice involved returning large amounts of crop residue to the ground (Figure 11.4c). Van Groenigen *et al.* (2014) suggest that this indicates earthworms stimulate plant growth in particular by releasing nitrogen locked away in residue and soil organic matter.

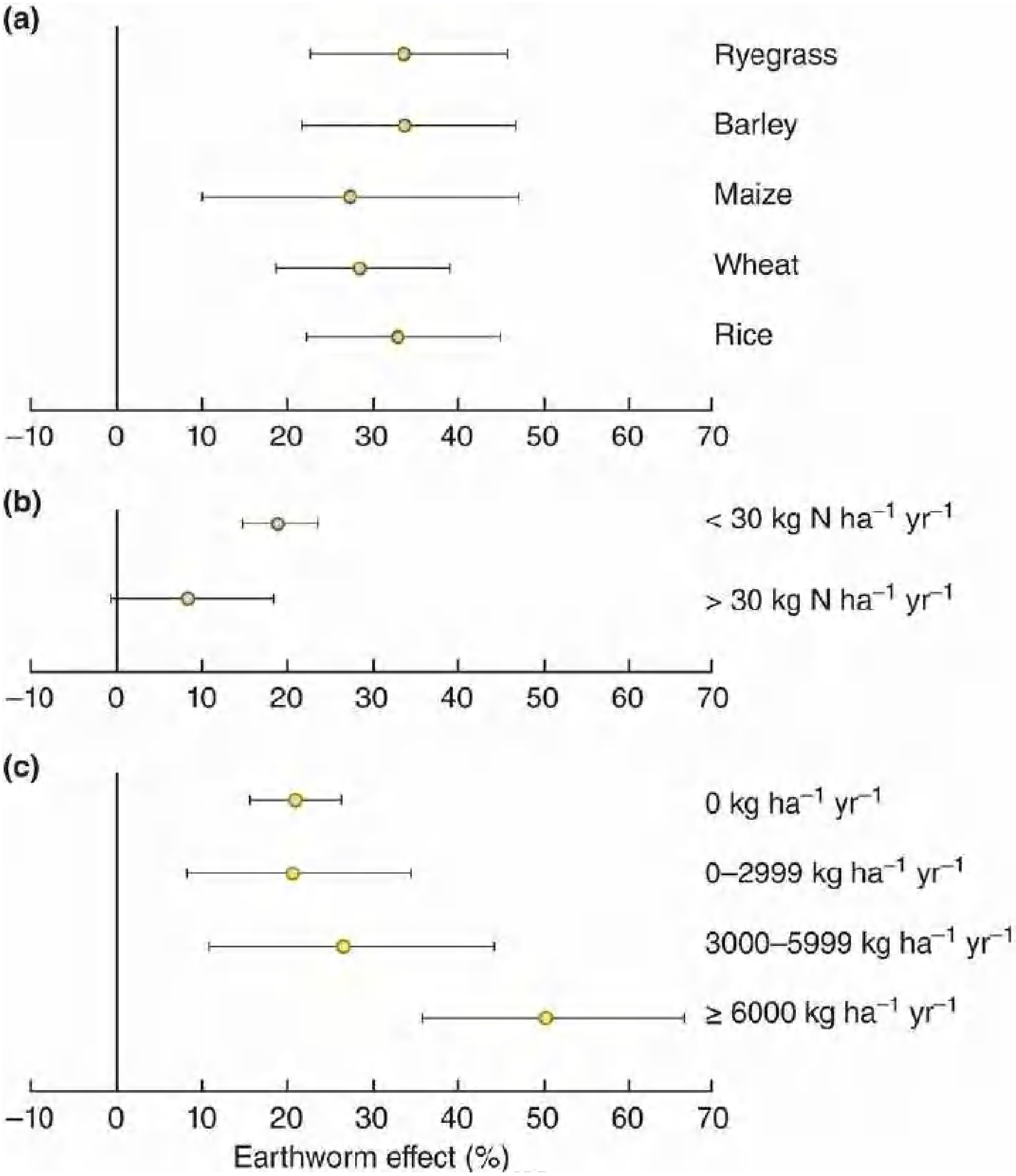


Figure 11.4 The positive effects of earthworms on crop yields. Results of a meta-analysis on the effect of earthworm presence on crop yield in relation to (a) above-ground biomass of individual crops, (b) rate of nitrogen fertiliser application and (c) the rate of application of crop residue. Bars are 95% CIs.

Source: After van Groenigen *et al*. (2014).

It is not surprising that agricultural ecologists become worried about practices that reduce worm populations. On the other hand, earthworm effects are not universally positive. Given the topical question of the need to reduce the release into the atmosphere of greenhouse gases such as CO_2, N_2O and CH_4 it should be noted that earthworms have been found to increase the combined cumulative emissions of CO_2 and N_2O by 7–42% from maize fields in the Netherlands (Lubbers *et al.*, 2015) and of CH_4 by up to 300% from flooded rice paddies in the Philippines (John *et al.*, 2015).

Detritivores occur in all types of terrestrial habitat and are often found at remarkable species richness and in very great numbers. Thus, for example, a mere square metre of temperate woodland soil may contain 1000 species of animals, in populations exceeding 10 million for nematode worms and protozoans, 100 000 for springtails (Collembola) and soil mites (Acari), and 50 000 or so for other invertebrates (Anderson, 1978). Temperature, naturally, has a fundamental role in determining the rate of decomposition, but the thickness of water films on decomposing material also places absolute limits on mobile microfauna (protozoa, nematode worms, rotifers) plus those fungi that have motile stages in their life cycles. In dry soils, such organisms are virtually absent, and a continuum can be recognised from dry conditions through waterlogged soils to true aquatic environments. In the former, the amount of water and thickness of water films are of paramount importance. But as we move along the continuum, conditions change to resemble more and more closely those of the bed of an open-water community, where oxygen shortage, rather than water availability, may dominate the lives of the organisms.

... and by feeding mode in aquatic environments

In freshwater ecology the study of detritivores has been concerned less with the size of the organisms than with the ways in which they obtain their food. Cummins (1974) devised a scheme that recognises four main categories of invertebrate consumer in streams (Figure 11.5). *Shredders* are detritivores that feed on coarse particulate organic matter (particles >2 mm in size), and during feeding these serve to fragment the material. Very often in streams, the shredders, such as cased caddisfly larvae, freshwater shrimps and isopods, feed on tree leaves that fall into the stream. *Collectors* feed on fine particulate organic matter (<2 mm). *Collector–gatherers* obtain dead organic particles from the debris and sediments on the bed of the stream, whereas *collector–filterers* sift small particles from the flowing column of water. *Grazer–scrapers* have mouthparts appropriate for scraping off and consuming the organic layer attached to rocks and stones; this organic layer is comprised of attached algae, bacteria, fungi and dead organic matter adsorbed to the substrate surface. The final invertebrate category is *carnivores*. Figure 11.6 shows the relationships amongst these invertebrate feeding groups and three categories of dead organic matter. This scheme, developed for stream communities, has obvious parallels in terrestrial ecosystems as well as in other aquatic ecosystems. Earthworms are important shredders in soils, while a variety of crustaceans perform the same role on the seabed. On the other hand, filtering is common among marine but not terrestrial organisms.

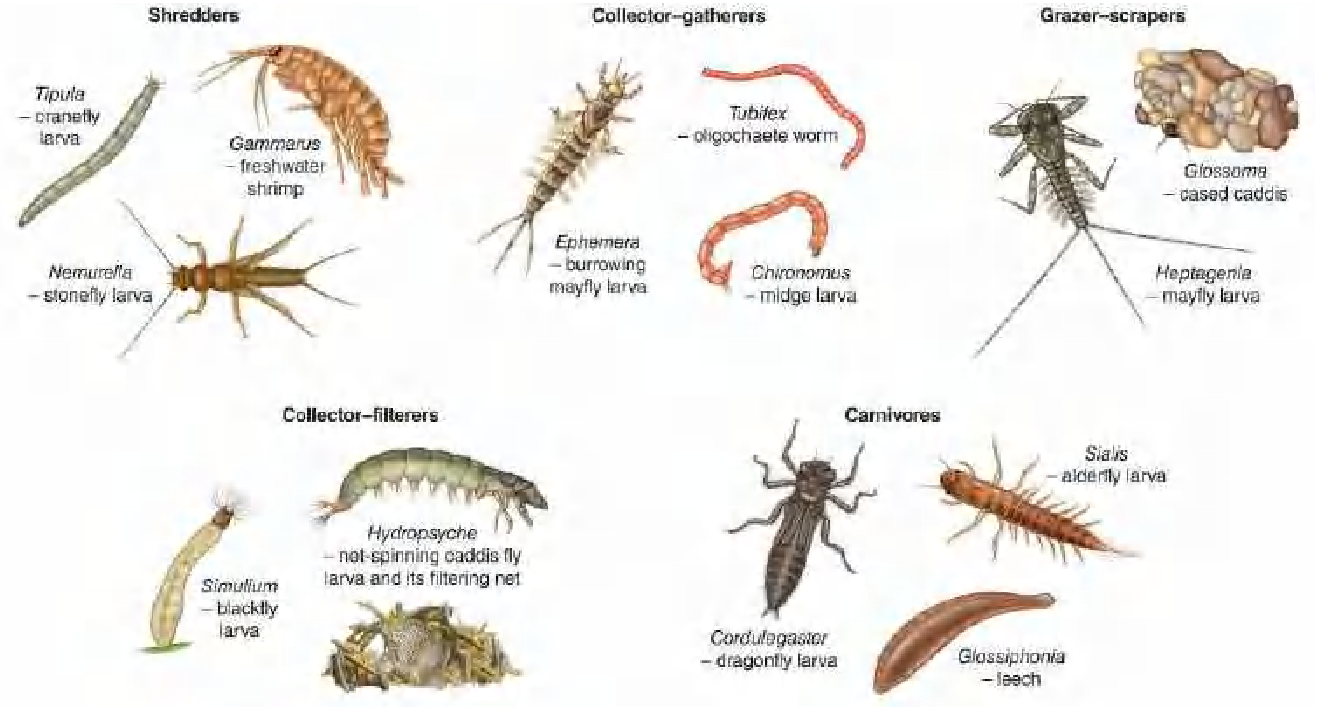


Figure 11.5 Examples of the various categories of invertebrate consumer in freshwater environments.

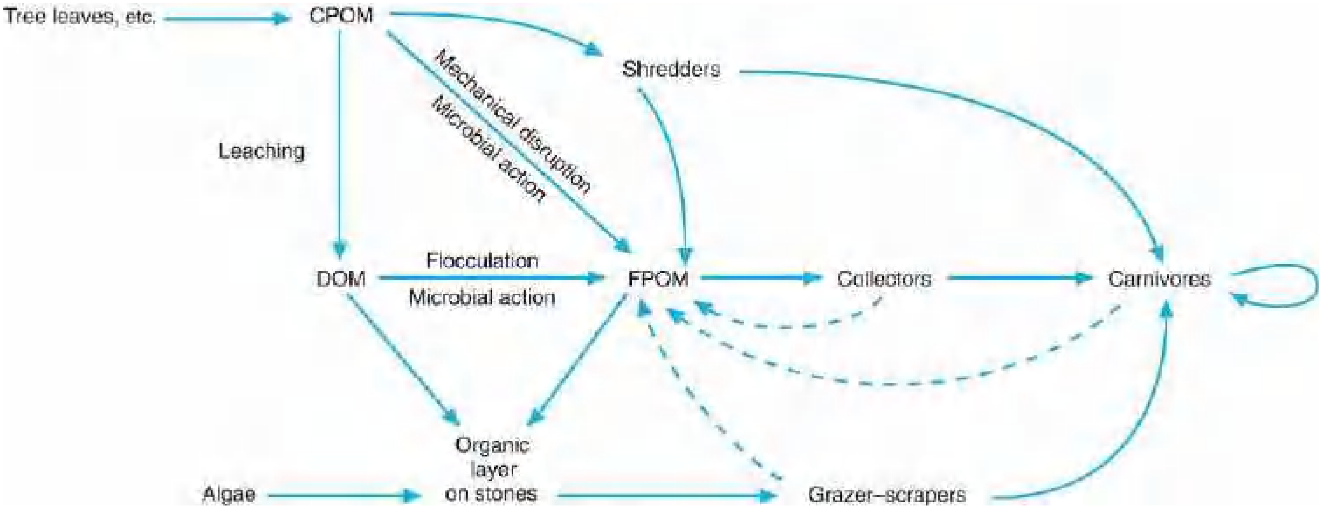


Figure 11.6 **A general model of energy flow in a stream.** A fraction of coarse particulate organic matter (CPOM) is quickly lost to the dissolved organic matter (DOM) compartment by leaching. The remainder is converted by three processes to fine particulate organic matter (FPOM): (i) mechanical disruption by battering; (ii) processing by microorganisms causing gradual break up; and (iii) fragmentation by the shredders. Note also that all animal groups contribute to FPOM by producing faeces (dashed arrows). DOM is also converted into FPOM by a physical process of flocculation or via uptake by microorganisms. The organic layer attached to stones on the stream bed derives from algae, DOM and FPOM adsorbed onto an organic matrix.

The faeces and bodies of aquatic invertebrates are generally processed along with dead organic matter from other sources by shredders and collectors. Even the large faeces of aquatic vertebrates do not appear to possess a characteristic fauna, probably because such faeces are likely to fragment and disperse quickly as a result of water movement. Carrion also lacks a specialised fauna – many aquatic invertebrates are omnivorous, feeding for much of the time on plant detritus and faeces with their associated microorganisms, but ever ready to tackle a piece of dead invertebrate or fish when this is available. This contrasts with the situation in the terrestrial environment, where both faeces and carrion have specialised detritivore faunas (see Sections 11.3.3 and 11.3.4).

detritivore-dominated communities

Some animal communities are composed almost exclusively of detritivores and their predators. This is true not only of the forest floor, but also of shaded streams, the depths of oceans and lakes, and the permanent residents of caves: in short, wherever there is insufficient light for appreciable photosynthesis but nevertheless an input of organic matter from nearby plant communities. The forest floor and shaded streams receive most of their organic matter as dead leaves from trees. The beds of oceans and lakes are subject to a continuous settlement of detritus from above. Caves receive dissolved and particulate organic matter percolating down through soil and rock, together with windblown material and the debris of migrating animals.

11.2.3 The relative roles of decomposers and detritivores

assessing the relative importance of decomposers and detritivores in streams ...

The roles of the decomposers and detritivores in decomposing dead organic matter in streams can be compared in a variety of ways. A comparison of numbers will reveal a predominance of decomposers, particularly bacteria and archaea. This is almost inevitable because we are counting individual cells. A comparison of biomass gives a quite different picture. Figure 11.7a–c shows the relative amounts of biomass represented by different groups involved in the decomposition of tree leaf litter (*Rhododendron maximum*) on the beds of two streams in North Carolina, USA. One stream was naturally poor in inorganic nutrients (nitrogen and phosphorus) while the other had for several years been enriched with nutrients to mimic the effects of agricultural runoff. In both cases the biomass of fungi peaked before that of detritivores (mainly shredding invertebrates) during the 50-day study, while bacterial biomass was much lower. Nutrient enrichment accelerated the colonisation sequence, led to a dramatic increase in shredder biomass in the later stages (mainly larvae of the caddisfly *Pycnopsyche* spp.) and resulted in much faster decomposition overall. Note, though, that biomass is itself a poor measure of the relative importance of decomposers and detritivores. Figure 11.7d provides estimates of the ultimate contributions of fungi, bacteria and detritivores to the decomposition of the leaves after factoring in animal feeding rates and the population growth rates of fungi and bacteria together with their efficiencies of converting organic carbon to biomass. By the end of the experiment, fungi predominated over detritivores in the nutrient-poor stream while shredding by detritivores was much more important in the enriched stream.

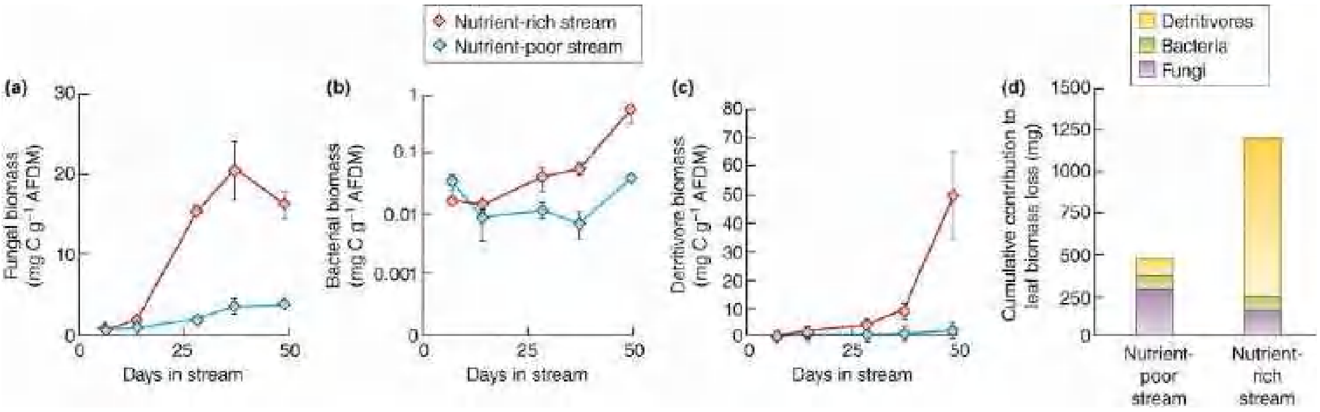


Figure 11.7 Comparing the biomasses of decomposers and detritivores during the decomposition of rhododendron tree leaves. (a) Fungi, (b) bacteria and (c) detritivores in two streams (nutrient-poor and nutrient-enriched) during a 50-day study. (d) The relative contributions of fungi, bacteria and detritivores to leaf biomass loss on the last day of the study. Bars are ± 1 SEs.

Source: Modified from Tant *et al*. (2015).

... in the decomposition of a salt marsh plant, ...

Lillebo *et al*. (1999) attempted to distinguish the relative roles of bacteria, microfauna (e.g. flagellates) and macrofauna (e.g. the snail *Hydrobia ulvae*) in the decomposition of the salt marsh plant, *Spartina maritima*, by creating artificial communities in laboratory microcosms. At the end of the 99-day study, 32% of the biomass of *Spartina* leaves remained in the bacteria treatment, whereas only 8% remained when the microfauna and macrofauna were also present (Figure 11.8a). Separate analyses of the mineralisation of the carbon, nitrogen and phosphorus content of the leaves also revealed that bacteria were responsible for the majority of the mineralisation, but that microfauna and particularly macrofauna enhanced the mineralisation rates in the case of carbon and nitrogen (Figure 11.8b).

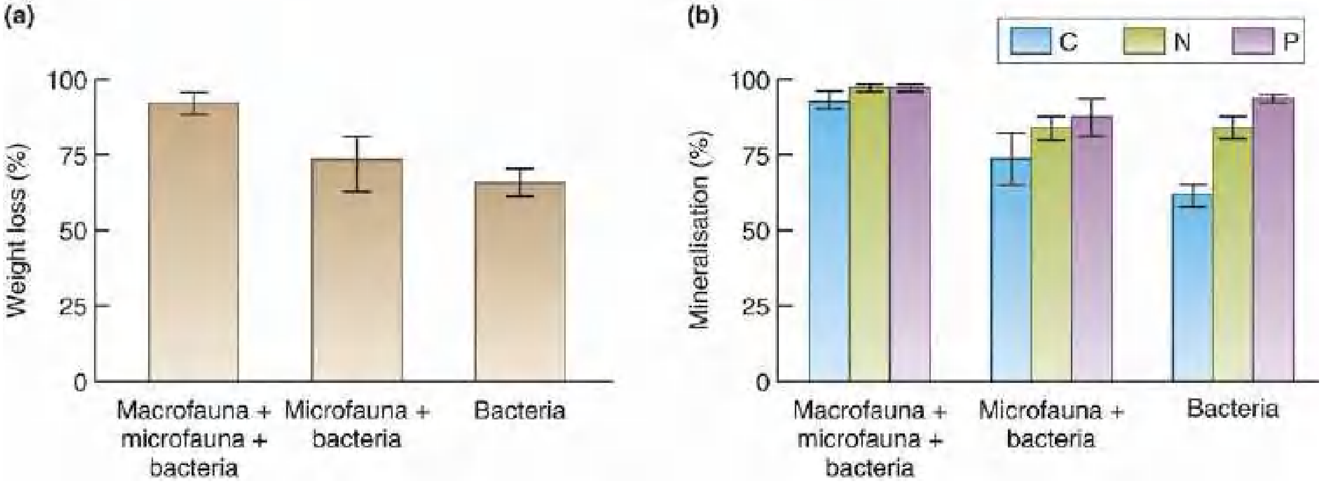


Figure 11.8 Relative roles of decomposers and detritivores in decomposition of a salt marsh plant. (a) Weight loss of *Spartina maritima* leaves during 99 days in the presence of: (i) macrofauna + microfauna + bacteria, (ii) microfauna + bacteria, or (iii) bacteria alone (mean ± SD). (b) Percentage of initial carbon, nitrogen and phosphorus content that was mineralised during 99 days in the three treatments.

Source: After Lillebo *et al*. (1999).

What this emphasises is that the decomposition of dead material is not simply the result of the sum of the activities of microbes and detritivores: it comes largely from the interaction between the two. The shredding action of detritivores usually produces smaller particles with a larger surface area (per unit volume of litter) and thus increases the area of substrate available for microorganism growth. In addition, the activity of fungi may be stimulated by the disruption, through grazing, of competing hyphal networks. Moreover, the activity of both fungi and bacteria may be enhanced by the addition of mineral nutrients in excreta.

... in dead wood ...

Dead wood provides particular challenges to colonisation by microorganisms because of its patchy distribution and tough exterior. Insects can enhance fungal colonisation of dead wood by carrying fungi to their 'target' or by enhancing access of air-disseminated fungal propagules by making holes in the outer bark into the phloem and xylem. Muller *et al.* (2002) distributed standard pieces of spruce wood (*Picea abies*) on a forest floor in Finland. After 2.5 years, the numbers of insect 'marks' (boring and gnawing) proved to be correlated with dry weight loss of the wood (Figure 11.9a). This relationship comes about because of biomass consumption by the insects but also, to an unknown extent, by fungal action that has been enhanced by insect activity. Thus, fungal infection rate was always high when there were more than 400 marks per piece of wood made by the common ambrosia beetle *Trypodendron lineatum* (Figure 11.9b). This species burrows deeply into the sapwood and produces galleries about 1 mm in diameter. Some of the fungal species involved are likely to have been transmitted by the beetle (e.g. *Ceratocystis piceae*) but the invasion of other, air-disseminated types is also promoted by the galleries left by the beetle.

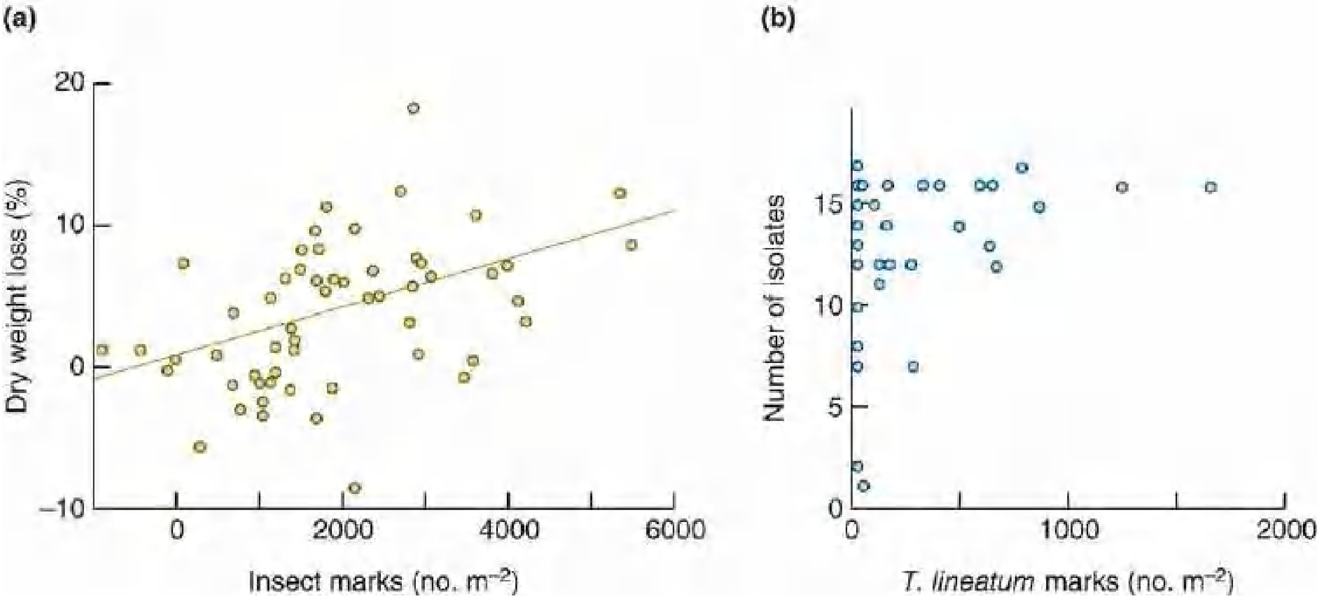


Figure 11.9 **Boring insects facilitate fungal decomposers of wood.** Relationships between (a) the decay of standard pieces of dead spruce wood over a 2.5-year period in Finland and the number of insect marks, and (b) the fungal infection rate (number of fungal isolates per standard piece of wood) and number of marks made by the beetle *Trypodendron lineatum*. Dry weight loss and number of insect marks in (a) were obtained by subtracting the values for each wood sample held in a permanently closed net cage from the corresponding value for its counterpart in a control cage that permitted insect entry. In some cases, the dry weight loss of the counterpart wood sample was lower, so the percentage weight loss was negative. This is possible because the number of insect visits does not explain all the variation in dry weight loss.

Source: After Muller *et al.* (2002).

... and in small mammal carcasses

Paralleling the results for wood decomposition, tunnelling by detritivorous animals can also enhance microbial respiration in small mammal carcasses. We see this when two sets of insect-free rodent bodies were placed individually in chambers for 15 days in an English grassland in autumn – one set repeatedly pierced with a dissecting needle to produce tunnels that simulated the action of blowfly larvae. The rate of production of carbon dioxide, a measure of microbial activity, was markedly faster in the set with tunnels, which enhanced dissemination of decomposers and increased aeration (Putman, 1978).

11.2.4 Are local communities predisposed to deal effectively with local litter?

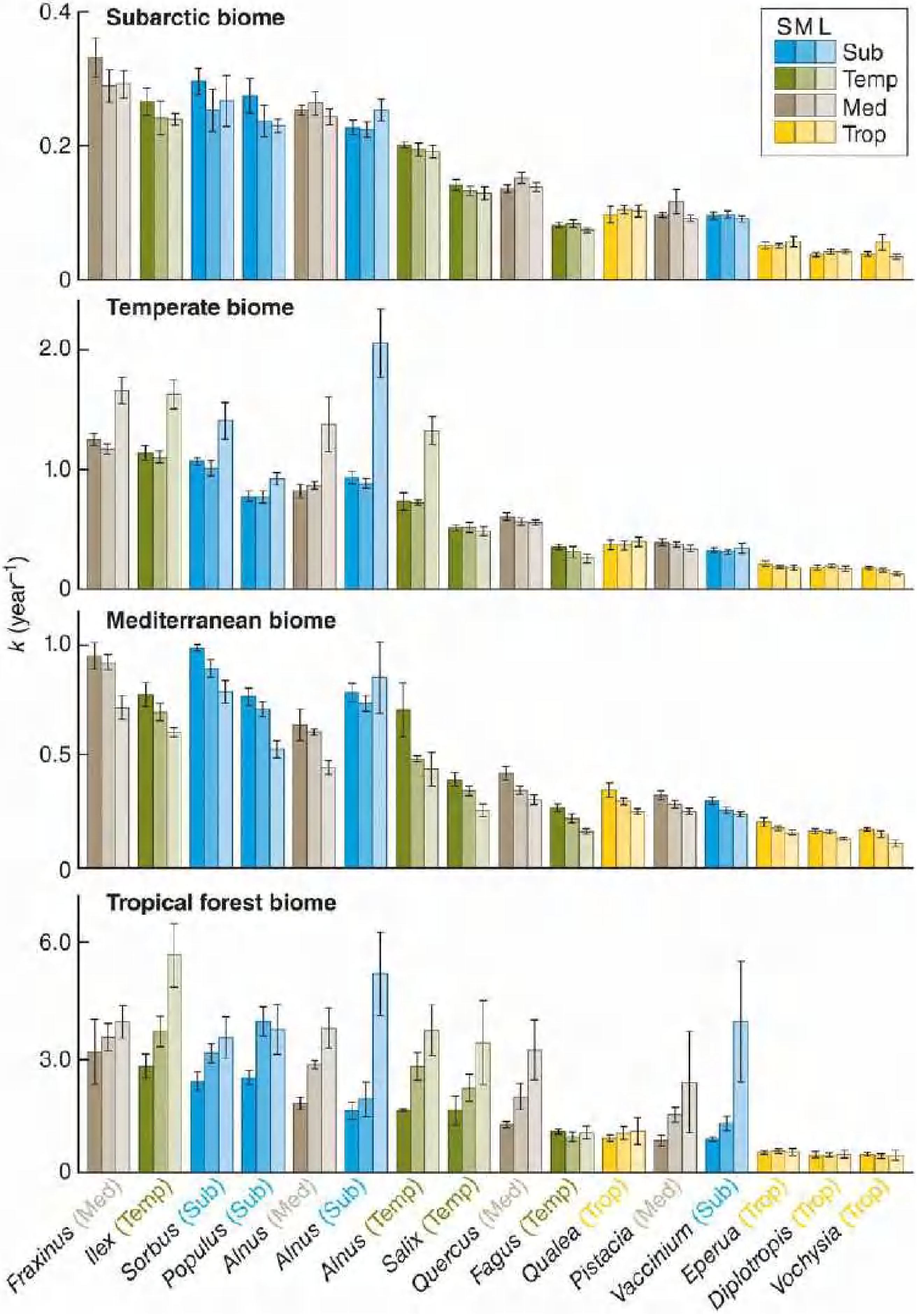
Given the great variation in physical and chemical make-up of plants, we should not be surprised if decomposer/detritivore communities include species able to deal effectively with their local litter. Veen *et al.* (2015) performed a meta-analysis of 125 published reciprocal litter transplant experiments worldwide to address the idea that decomposition rate of plant litter is faster in the vicinity of the plant(s) from which it originates (home) than away from the plant.

a ‘home-field’ advantage at a local scale

They found an average home-field advantage, with the rate of litter mass loss 7.5% faster at home than away. However, there was considerable variation, and so Veen *et al.* (2015) tested the hypothesis that the home-field advantage would increase with the magnitude of dissimilarity between the plant communities in the home and away study situations, presuming that the more different the plant communities, the more different would be their decomposer/detritivore assemblages. They found that when an ‘away’ site contained both different plant species and different functional groups (trees, shrubs, grasses, forbs) (e.g. transplanting litter from forest to grassland or vice versa), the home field advantage increased to about 28%.

faunal contributions to litter decomposition seem to depend on climatic conditions ...

Even where it occurs, however, home field advantage accounts for a relatively small amount of the variation in decomposition rates. Makkonen *et al.* (2012) investigated the relative importance of other influential factors in a global experiment involving four forest biomes. From each, they selected four representative native woody species and performed reciprocal transplants among all four biomes. Every transplant consisted of 5 g of dried litter in litterbags with meshes of three sizes that admitted microorganisms + microfauna (S), or microorganisms + microfauna + mesofauna (M), or the complete assemblage including macrofauna (L) (Figure 11.10). The composition of the decomposer/detritivore assemblage played a relatively modest role, with evidence of accelerated decomposition in the presence of macrofauna mainly restricted to the tropical and temperate biomes. The authors suggest that larger detritivores play their greatest role only under the most favourable climatic conditions. In another global decomposition experiment across 39 sites in six continents, Wall *et al.* (2008) also reported that the mesofauna increased decomposition rate in temperate and wet tropical climates, but had neutral effects where temperature or moisture constrains biological activity (Figure 11.11), highlighting the importance of climatic conditions in determining the contribution to decomposition of these soil animals.



Subarctic biome
0.4
0.2
0
Temperate biome
2.0
1.0
0
Mediterranean biome
1.0
0.5
0
Tropical forest biome
6.0
3.0
0
k (year⁻¹)
S M L
Sub
Temp
Med
Trop
Fraxinus (Med)
Ilex (Temp)
Sorbus (Sub)
Populus (Sub)
Alnus (Med)
Alnus (Sub)
Alnus (Temp)
Salix (Temp)
Quercus (Med)
Fagus (Temp)
Qualea (Trop)
Pistacia (Med)
Vaccinium (Sub)
Eperua (Trop)
Diplotropis (Trop)
Vochysia (Trop)

Figure 11.10 **Leaf decomposition compared in different biomes.** Average decomposition rate constants (*k* ± SE; *k* = –ln[final mass/initial mass]/time) in reciprocal transplant experiments involving 16 plant species, four from each of subarctic (Sweden), temperate (the Netherlands), Mediterranean (France) and tropical forest biomes (French Guiana). Bar shade indicates composition of the detritivore community as influenced by mesh size (small, medium, large). Note the different scales on the vertical axes.

Source: From Makkonen *et al.* (2012).

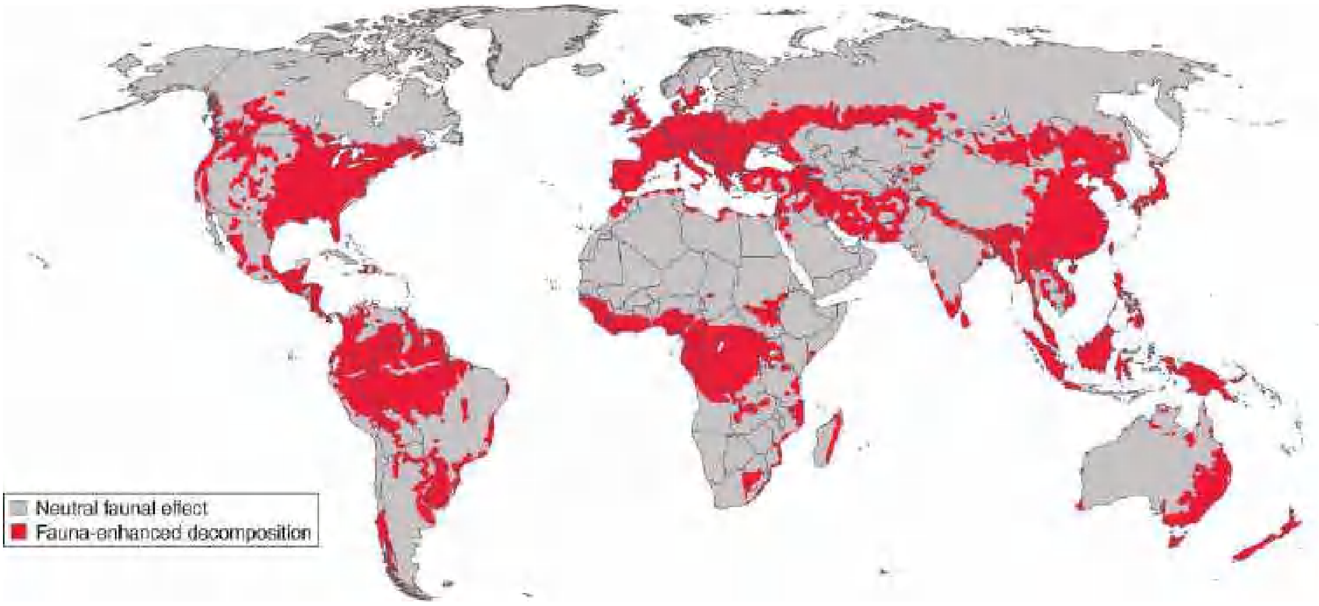


Figure 11.11 **Distribution of climatic regions where soil animals can be expected to enhance decomposition rates.** This is based on the results of litter bag experiments (2 mm mesh, containing dried *Agropyron cristatum* grass) where half the bags had been treated with naphthalene (‘mothballs’) to suppress soil animals (including isopods, collembolans, insects and insect larvae, millipedes, mites, harvest spiders, spiders, annelids and snails).

Source: After Wall *et al.* (2008).

… but litter quality plays the dominant role in determining decomposition rate

The most striking finding of Makkonen *et al.* (2012) was that by far the strongest drivers of litter decomposition rate were a few specific aspects of litter quality (water saturation capacity, magnesium and tannins). The capacity of litter to absorb water probably controls the microclimate for decomposers, while magnesium is an essential element for detritivores and tannins can be toxic and bind to dietary proteins and digestive enzymes (Makkonen *et al.*, 2012). Overall, decomposition rates of all litter species were highest in the tropical biome, followed by the temperate, Mediterranean and finally subarctic biomes. However, in all biomes, tropical litter species were decomposed more slowly than species from other biomes (even in the tropical biome), testifying to the poor quality of their litter.

11.2.5 Ecological stoichiometry and the chemical composition of decomposers, detritivores and their resources

‘ecological stoichiometry’ and relations between resources and consumers

Ecological stoichiometry, defined by Elser and Urabe (1999) as the analysis of constraints and consequences in ecological interactions of the mass balance of multiple chemical elements (particularly the ratios of carbon to nitrogen, carbon to phosphorus and nitrogen to phosphorus), is an approach that can shed light on the relations between resources and consumers (see Section 3.7.1). Many studies have focused on plant–herbivore relations but the approach is also important when considering decomposers, detritivores and their resources.

There is a great contrast between the chemical composition of dead plant tissue and that of the tissues of the heterotrophic organisms that consume and decompose it. While the major components of plant tissues, particularly cell walls, are structural polysaccharides, these are only of minor significance in the bodies of microorganisms and detritivores. However, being harder to digest than storage carbohydrates and protein, the structural chemicals still form a significant component of detritivore faeces. Hence, detritivore faeces and plant tissue have much in common chemically, but the protein and lipid contents of detritivores and decomposers themselves are significantly higher than those of plants and faeces.

decomposition rate depends on ... biochemical composition ...

The rate at which dead organic matter decomposes is strongly dependent on its biochemical composition. This is because microbial tissue has very high nitrogen and phosphorus contents, indicative of high requirements for these nutrients. Roughly speaking, the stoichiometric ratio of carbon : nitrogen : phosphorus in soil microbial biomass is 60 : 7 : 1 (Cleveland & Liptzin, 2006), meaning that a microbial biomass of 68 g can only develop if there is 7 g of nitrogen and 1 g of phosphorus available. Terrestrial plant material has much higher ratios of about 3000 : 46 : 1 (Reich & Oleksyn, 2004). Consequently, this material can support only a limited biomass of decomposer organisms and the whole pace of the decomposition process will itself be limited by nutrient availability. Marine and freshwater plants and algae tend to have ratios more similar to the decomposers, and their rates of decomposition are correspondingly faster (Figure 11.12a). Figure 11.12b and c illustrate the strong relationships between initial nitrogen and phosphorus concentration in plant tissue and its decomposition rate for a wide range of plant detritus from terrestrial, freshwater and marine species.

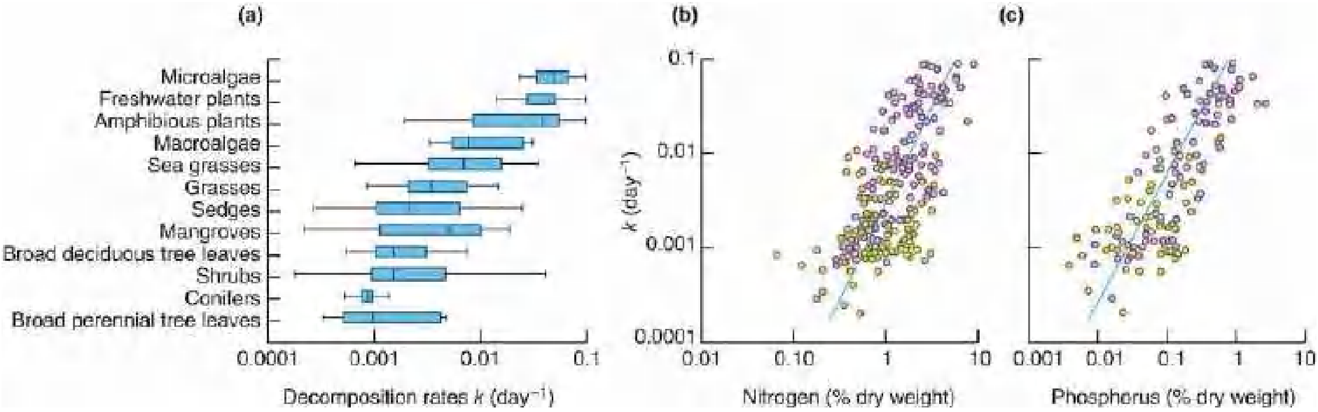


Figure 11.12 Decomposition rates vary with habitat type and detrital nutrient concentrations. (a) Box plots showing the recorded decomposition rates of detritus from different sources. The decomposition rate is expressed as *k* (per day). Boxes encompass the 25% and 75% quartiles of all data from the literature for each plant type. The central line represents the median and bars extend to the 95% confidence limits. The relationships between decomposition rate and the initial concentrations in the tissues (% dry weight) of (b) nitrogen and (c) phosphorus are also shown. Solid blues lines represent fitted regression lines and green and purple circles represent detritus decomposing on land and submersed, respectively.

Source: After Enriquez *et al*. (1993).

... and mineral nutrients in the environment

The rate at which dead organic matter decomposes is also influenced by the amounts of ‘extraneous’ inorganic nutrients, especially nitrogen (as ammonium or nitrate), that are available from the environment. Thus, greater microbial biomass can be supported, and decomposition proceeds faster, if nitrogen is absorbed from outside. The extraneous nitrogen most often derives from human activities (e.g. agricultural fertilisation, combustion of fossil fuel), but the guano of seabirds deposited in their burrows in coastal forest can produce a similar outcome. In a comparison of forest soils in New Zealand with and without petrel colonies (*Procellaria westlandica*), a paucity of leaf litter in seabird-inhabited forest resulted from the enhanced processing of a limited pool of terrestrial soil carbon by decomposers primed by the nitrogen-rich guano (Hawke & Vallance, 2015).

complex relationships between decomposers and living plants ...

One consequence of the capacity of decomposers to use inorganic nutrients is that after plant material is added to soil, the level of soil nitrogen tends to fall rapidly as it is incorporated into microbial biomass and subsequently into organic matter pools. The effect is particularly evident in agriculture, where the ploughing in of stubble can result in nitrogen deficiency of the subsequent crop. In other words, the decomposers compete with the plants for inorganic nitrogen. This raises a significant and somewhat paradoxical issue. We have noted that plants and decomposers are linked by an indirect *mutualism* mediated by nutrient recycling – plants provide energy and nutrients in organic form that are used by decomposers, and decomposers mineralise the organic material back to an inorganic form that can again be used by plants. However, stoichiometric constraints on carbon and nutrients also lead to *competition* between the plants and decomposers (usually for nitrogen in terrestrial communities, often for phosphorus in freshwater communities, and either nitrogen or phosphorus in marine communities).

... competition and mutualism

Daufresne and Loreau (2001) developed a model that incorporates both mutualistic and competitive relationships and posed the question ‘what conditions must be met for plants and decomposers to coexist and for the ecosystem as a whole to persist?’ Their model showed that the plant–decomposer system is generally persistent (both plant and decomposer compartments reach a stable positive steady state) only if decomposer growth is limited by the availability of carbon in the detritus – and this condition can only be achieved if the competitive ability of the decomposers for a limiting nutrient (e.g. nitrogen) is great enough, compared with that of plants, to maintain them in a state of carbon limitation. Experimental studies have shown that bacteria can, in fact, outcompete plants for nutrients (e.g. Hitchcock & Mitrovik, 2013).

In contrast to terrestrial plants, the bodies of animals have nutrient ratios that are of the same order as those of microbial biomass; thus their decomposition is not limited by the availability of nutrients, and animal bodies tend to decompose much faster than plant material.

When dead organisms or their parts decompose in or on soil, they begin to acquire the C : N ratio of the decomposers. On the whole, if material with a nitrogen content of less than 1.2–1.3% is added to soil, any available ammonium ions are absorbed. If the material has a nitrogen content greater than 1.8%, ammonium ions tend to be released. One consequence is that the C : N ratios of soils tend to be rather constant around values of 14 (Cleveland & Liptzin, 2006); the decomposer system is in general remarkably homeostatic. However, in extreme situations, where the soil is very acid or waterlogged, the ratio may be higher (an indication that decomposition is slow).

It should not be thought that the only activity of the microbial decomposers of dead material is to respire away the carbon and mineralise the remainder. A major consequence of microbial growth is the accumulation of microbial by-products, particularly fungal cellulose and microbial polysaccharides, which may themselves be slow to decompose and contribute to maintaining soil structure.

11.3 Detritivore–resource interactions

11.3.1 Consumption of plant detritus

Two of the major organic components of dead leaves and wood are cellulose and lignin. These pose considerable digestive problems for animal consumers, most of which are not capable of manufacturing the enzymatic machinery to deal with them. Cellulose catabolism (cellulolysis) requires *cellulase* enzymes. Without these, detritivores are unable to digest the cellulose component of detritus, and so cannot derive from it either energy to do work or the simpler chemical modules to use in their own tissue synthesis. Cellulases of animal origin have been definitely identified in rather few species, including some cockroaches, termites, beetles, sea squirts and molluscs (Davison & Blaxter, 2005). In these organisms, cellulolysis poses no special problems.

most detritivores rely on microbial cellulases – they do not have their own

The majority of detritivores, lacking their own cellulases, rely on the production of cellulases by associated decomposers or, in some cases, protozoa. The interactions range from *obligate mutualism* between a detritivore and a specific and permanent gut microbiota or microfauna, through *facultative mutualism*, where the animals make use of cellulases produced by a microbiota that is ingested with detritus as it passes through an unspecialised gut (Figure 11.13). A final class consists of a wide range of detritivores that ingest the metabolic products of external cellulase-producing microbiota associated with decomposing plant remains or faeces. Because microbes and plant detritus are often very intimately associated, many animals simply cannot manage to take a mouthful of one without the other. The invertebrates then consume the partially digested plant detritus along with its associated bacteria and fungi, no doubt obtaining a significant proportion of the necessary energy and nutrients by digesting the microbiota itself. These animals, such as the isopod *Porcellio scaber* and the springtail *Tomocerus*, can be said to be making use of an ‘external rumen’ in the provision of assimilable materials from indigestible plant remains (Figure 11.13).

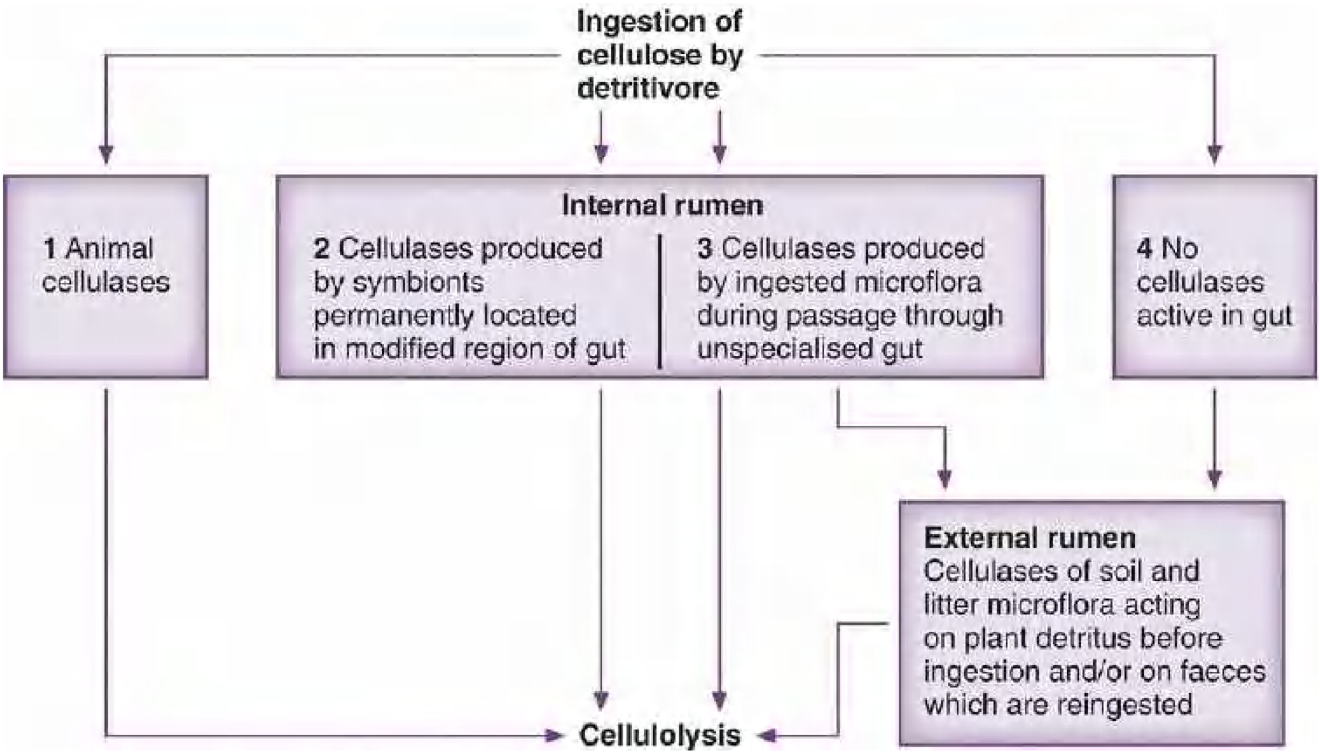


Figure 11.13 The range of mechanisms that detritivores adopt for digesting cellulose (cellulolysis).

Source: After Swift *et al.* (1979).

cockroaches and termites generally rely on bacteria and protozoa

Clear examples of obligate mutualism are found amongst certain species of cockroach and termite that rely on symbiotic bacteria or protozoa for the digestion of structural plant polysaccharides (discussed further in Section 13.6.3). Indeed, Nalepa *et al.* (2001) describe the evolution of digestive mutualisms among the Dictyoptera (cockroaches and termites) from cockroach-like ancestors in the Upper Carboniferous that fed on rotting vegetation and relied on an 'external rumen' through to the specialised internal rumens of many modern species. In lower termites, such as *Eutermes*, for example, symbiotic protozoa may make up more than 60% of the insect's body weight. The protozoa are located in the hindgut, which is dilated to form a rectal pouch. They ingest fine particles of wood, and are responsible for extensive cellulolytic activity, though bacteria are also implicated. Note that most termites probably depend mainly on symbiotic gut bacteria. Termites feeding on wood generally show effective digestion of cellulose but not of lignin, except for *Reticulitermes*, which has been reported to digest 80% or more of the lignin present in its food.

why no animal cellulases?

Given the versatility apparent in the evolutionary process, it may seem surprising that so few animals that consume plants can produce their own cellulase enzymes. Janzen (1981) has argued that cellulose is the master construction material of plants 'for the same reason that we construct houses of concrete in areas of high termite activity'. He views the use of cellulose, therefore, as a defence against attack, since higher organisms can rarely digest it unaided. From a different perspective, it has been suggested that cellulolytic capacity is uncommon simply because it is a trait that is rarely advantageous for animals to possess (Martin, 1991). For one thing, diverse bacterial communities are commonly found in hindguts and this may have facilitated the evolution of symbiont-mediated cellulolysis. For another, the diets of plant-eaters generally suffer from a limited supply of critical nutrients, such as nitrogen and phosphorus, rather than of energy, which cellulolysis would release. This imposes the need for processing large volumes of material to extract the required quantities of nutrients, rather than extracting energy efficiently from small volumes of material.

fruit flies and rotten fruit

Of course, not all plant detritus is so difficult for detritivores to digest. Fallen fruit, for example, is readily exploited by many kinds of opportunist feeders, including insects, birds and mammals. However, like all detritus, decaying fruits have associated with them a microbiota, in this case mainly dominated by yeasts, fermenting fruit sugars to alcohol, which is normally toxic. Fruit flies (*Drosophila* spp.) specialise at feeding on these yeasts and their by-products. *D. melanogaster* tolerates such high levels of alcohol because it produces large quantities of alcohol dehydrogenase (ADH), an enzyme that breaks down ethanol to harmless metabolites. In fact, ovipositing females of *D. melanogaster* show a preference for high alcohol levels, a behaviour that might serve to reduce attacks on their eggs by parasitoids (Zhu & Fry, 2015).

11.3.2 Feeding on invertebrate faeces

A large proportion of dead organic matter in soils and aquatic sediments may consist of invertebrate faeces, which generalist detritivores often include in their diets. Some of the faeces derive from grazing insects. In the laboratory, the faeces of caterpillars of *Operophthera fagata* that had grazed leaves of beech (*Fagus sylvatica*) decomposed faster than leaf litter itself; however, the decomposition rate was much enhanced when detritivorous isopods (*Porcellio scaber* and *Oniscus asellus*) fed on the faeces (Figure 11.14). Thus, rates of decomposition and nutrient release into the soil from grazer faeces can be increased through the feeding activity of coprophagous (that is, faeces-eating) detritivores.

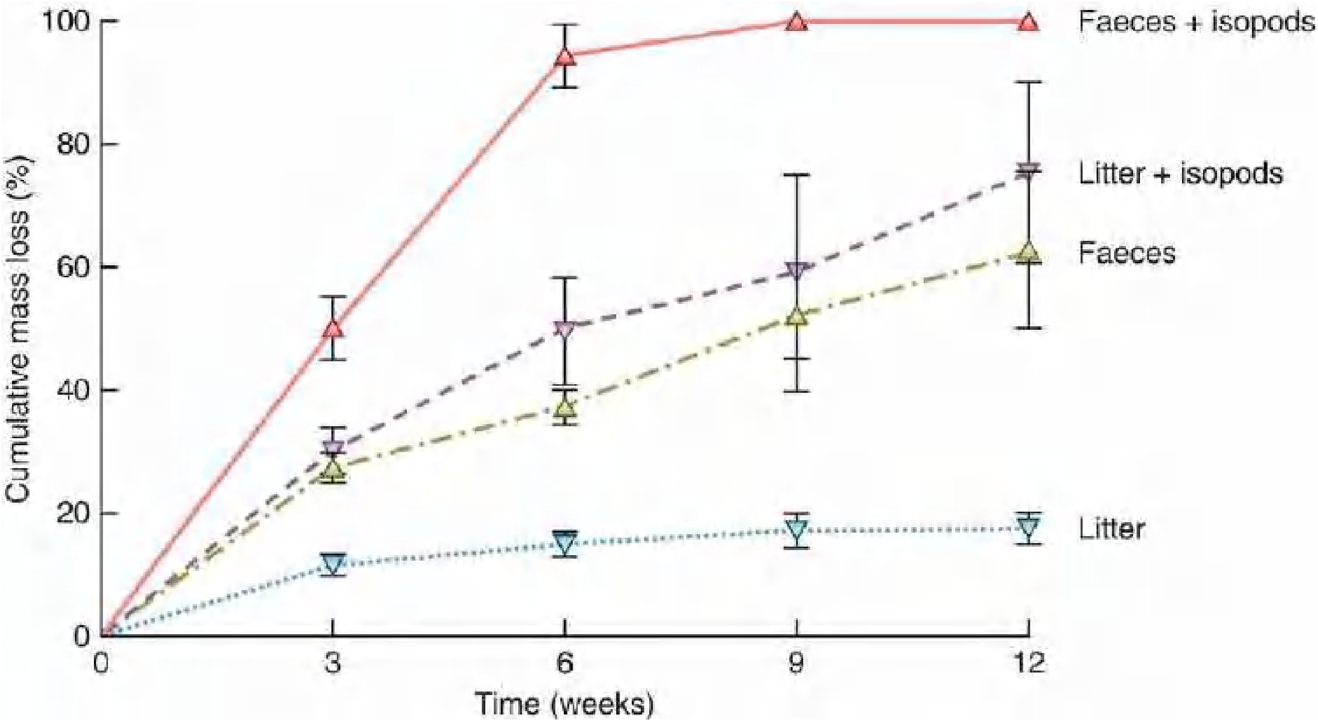


Figure 11.14 Isopods enhance the decomposition of both leaf litter and the faeces of caterpillars that grazed the living leaves. The cumulative mass loss of beech leaf litter and faeces of grazing caterpillars (*Operophthera fagata*) in the presence and absence of feeding by isopods. Bars are SEs.

Source: After Zimmer & Topp (2002).

'coprophagy' may be more valuable when detrital quality is low

Faeces of detritivores are common in many environments. In some cases, reingestion of faeces by the species, or even the individual, producing them may be critically important, by providing essential micronutrients or highly assimilable resources. In most cases, however, there are probably not marked nutritive benefits of feeding on faeces compared with the detritus from which the faeces were derived. Thus, the isopod *Porcellio scaber* gained no more from feeding on its faeces, even when these were experimentally inoculated with microbes, than from feeding directly on the leaf litter of alder (*Alnus glutinosa*) (Kautz *et al.*, 2002). On the other hand, in the case of the less nutritionally preferred leaves of oak (*Quercus robur*), faeces provided a small but significant increase in growth rate compared with the parent oak leaf material. Coprophagy may be more valuable when detrital quality is poor or, alternatively, in limited supply. Thus, a behavioural study of the minute deep-sea amphipod *Parvipalpus major* reported the animal bringing its head to the anal opening and gently extracting faeces from the hindgut (Corbari *et al.*, 2005). Coprophagy may be of particular importance in the deep sea where food is scarce and episodic.

11.3.3 Feeding on vertebrate faeces

carnivore dung is attacked mainly by bacteria and fungi

The dung of carnivorous vertebrates is relatively poor-quality stuff. Carnivores assimilate their food with high efficiency (usually 80% or more is digested) and their faeces retain only the least digestible components. In addition, carnivores are necessarily much less common than herbivores, and their dung is probably not sufficiently abundant to support a diverse fauna of specialist detritivores. Bacteria and fungi play the dominant role in the decomposition of carnivore faeces.

‘autocoprophagy’ among mammalian herbivores

In contrast, herbivore faeces still contain an abundance of organic matter. Autocoprophagy (reingesting one’s own faeces) is quite a widespread habit among small to medium-sized mammalian herbivores, being reported from rabbits and hares, rodents, marsupials and some primates (Hirakawa, 2001). Many species produce both soft and hard faeces, and it is the soft faeces that are usually reingested (directly from the anus), being rich in vitamins and microbial protein. If prevented from reingestion, many of these animals exhibit symptoms of malnutrition and grow more slowly.

herbivore dung supports its own characteristic detritivores

Herbivore dung is also sufficiently thickly spread in the environment to support its own characteristic fauna, consisting of many occasional visitors but with several specific dung-feeders. A wide range of animals are involved, including earthworms, termites and, in particular, flies and beetles. Flies (*Scatophaga stercoraria* and *Calliphora vicina*) and/or beetles (mainly *Aphodius* spp. and *Canthon* spp.) were experimentally excluded by mesh bags from standardised yak (*Bos grunniens*) dung pats in an alpine pasture on the Tibetan Plateau (Figure 11.15). The adult beetles feed on dung liquid and small particles while their larvae specialise on solid fibrous material. The flies lay their eggs on the dung and their larvae generally ingest fluids and aggregated bacteria-sized particles. When both animal groups were excluded, dung weight loss during the 32-day experiment was negligible (9.5% dry weight), compared to 70% lost when both groups had access. Dung loss was greater when beetles acted alone compared to flies acting alone, the latter having lower consumption rates and shorter residence times.

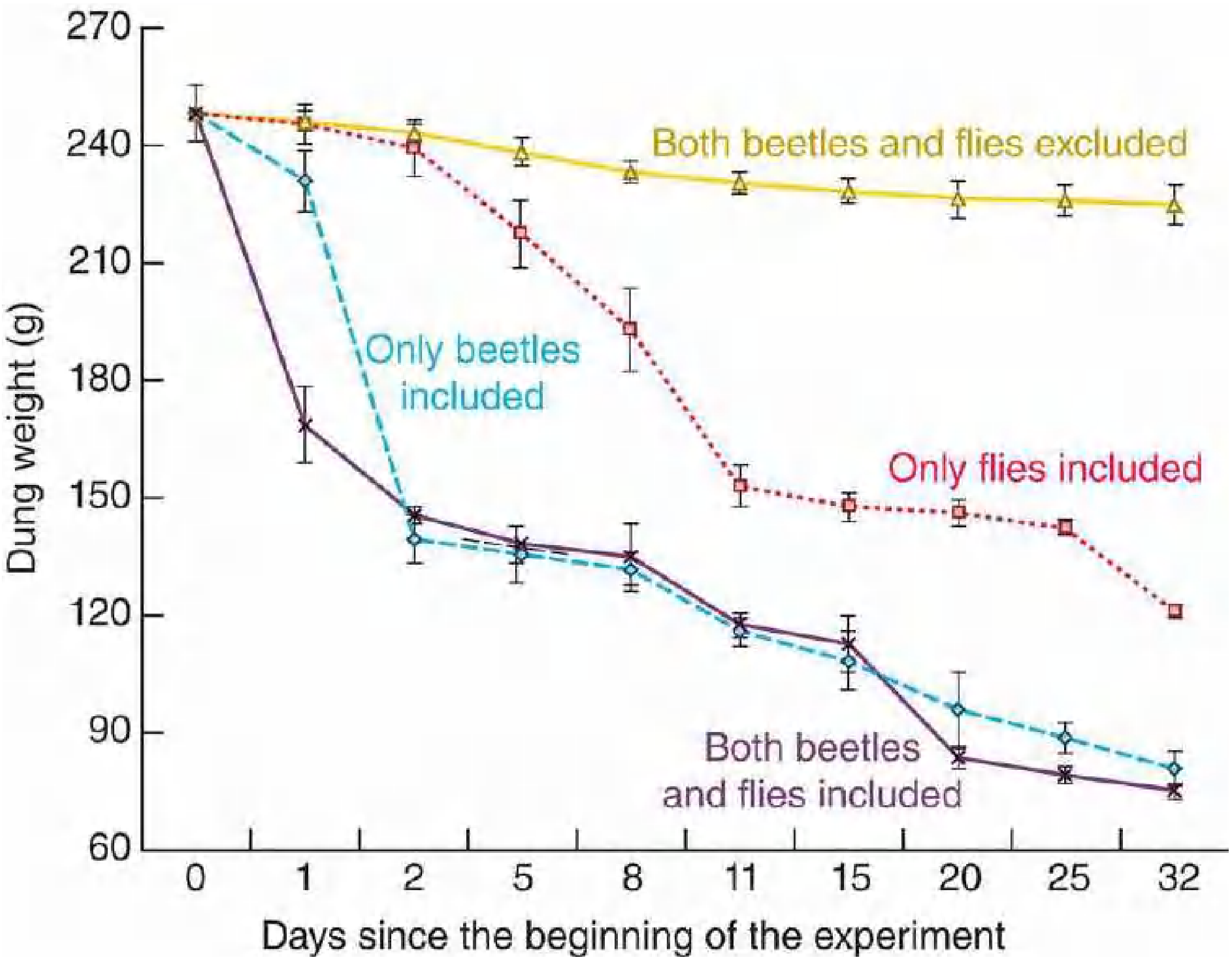


Figure 11.15 **The action of flies and particularly dung beetles accelerates decomposition of bovine dung.** Loss in weight of yak dung (dry weight ± SE) during a 32-day experiment in four treatments differing in whether dung beetles and/or flies were excluded.

Source: From Wu & Sun (2010).

a diversity of dung beetles

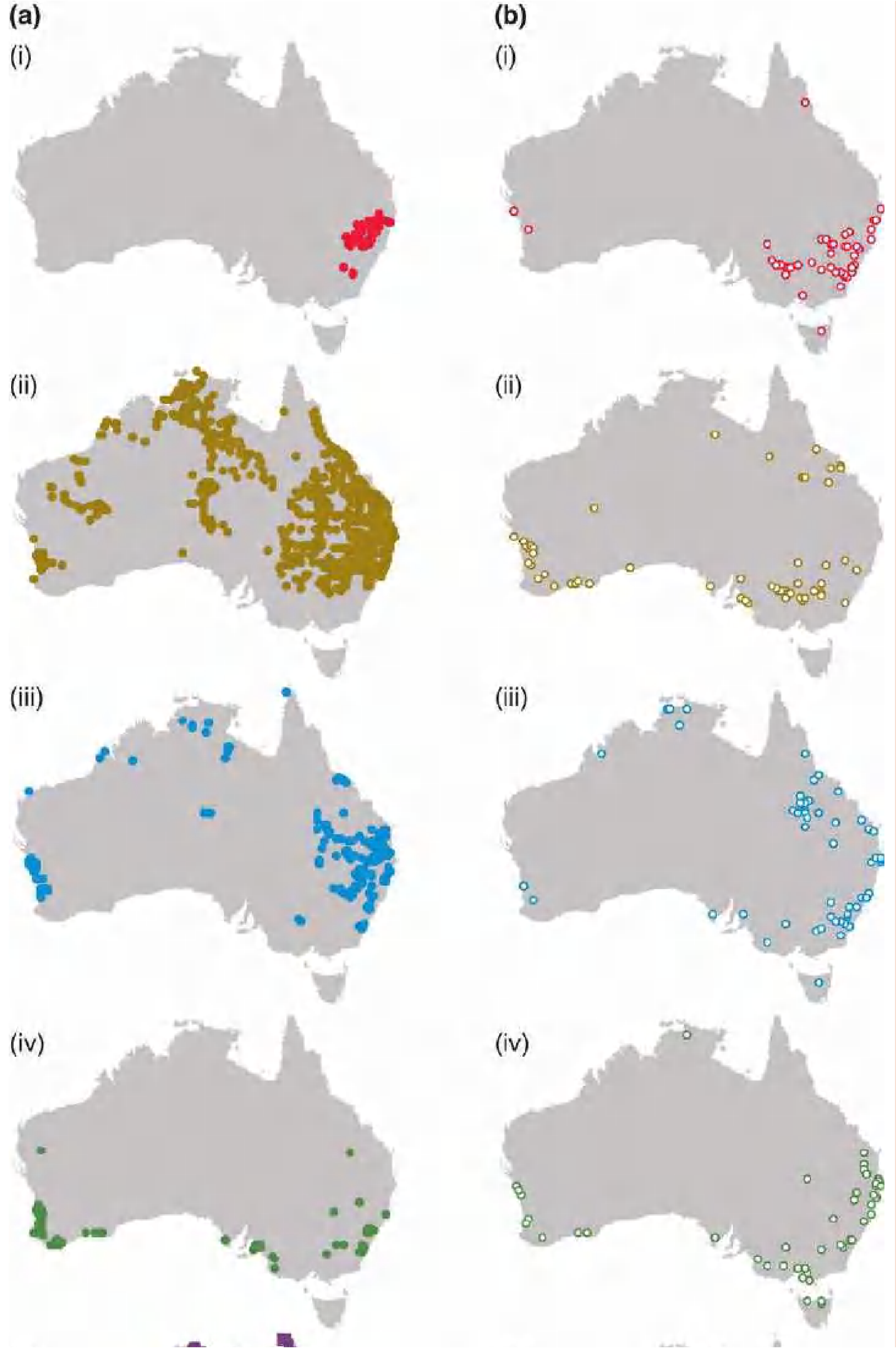
The full range of tropical dung beetles in the family Scarabaeidae vary in size from a few millimetres in length up to the 6 cm long *Heliocopris*. Some remove dung and bury it at a distance from the dung pile. Others excavate their nests at various depths immediately below the pile, while yet others build nest chambers within the dung pile itself. Beetles in other families do not construct chambers but simply lay their eggs in the dung, and their larvae feed and grow within the dung mass until fully developed, when they move away to pupate in the soil.

The beetles associated with elephant dung may remove 100% of the dung pile in the wet season. Any left may be processed by other detritivores such as flies and termites, as well as by decomposers. Dung that is deposited in the dry season is colonised by relatively few beetles (adults emerge only in the rains). Some microbial activity is evident, but this soon declines as the faeces dry out. Rewetting during the rains stimulates more microbial activity but beetles do not exploit old dung. In fact, a dung pile deposited in the dry season may persist for more than two years, compared with 24 hours or less for one deposited during the rains.

APPLICATION 11.2 The value of dung beetles to agriculture

helping to solve the problem of cow dung in Australia and New Zealand

The persistence of dung can be problematic. Each day cattle in Australia deposit some 300 million dung pats, covering as much as 2.4 million hectares per year with dung. Deposition of bovine dung poses no particular problem in most parts of the world, where bovines have existed for millions of years and have an associated fauna that exploits the faecal resources. However, the largest herbivorous animals in Australia, until European colonisation, were marsupials such as kangaroos. The native detritivores that deal with the dry, fibrous dung pellets of marsupials cannot cope with cow dung, and the loss of pasture under dung has imposed a huge economic burden on Australian agriculture. Adding to the problem, Australia is plagued by native bushflies (*Musca vetustissima*); their larvae feed on dung but fail to survive in dung that has been buried by beetles. The decision was therefore made in 1963 to establish beetles, mainly of South African origin, able to dispose of bovine dung in the most important places and under the most prevalent conditions where cattle are raised (Figure 11.16). Using suitable quarantine protocols, more than 50 species have been introduced (Duncan *et al*., 2009), of which about half established successfully and had positive effects on dung burial, pasture health and pest fly control. Bovine dung has also proved a problem in New Zealand, whose native fauna included no land mammals at all, except for three species of bat; dung beetle introductions commenced in New Zealand in 2013.


(a)
(b)
(i)
(i)
(ii)
(ii)
(iii)
(iii)
(iv)
(iv)

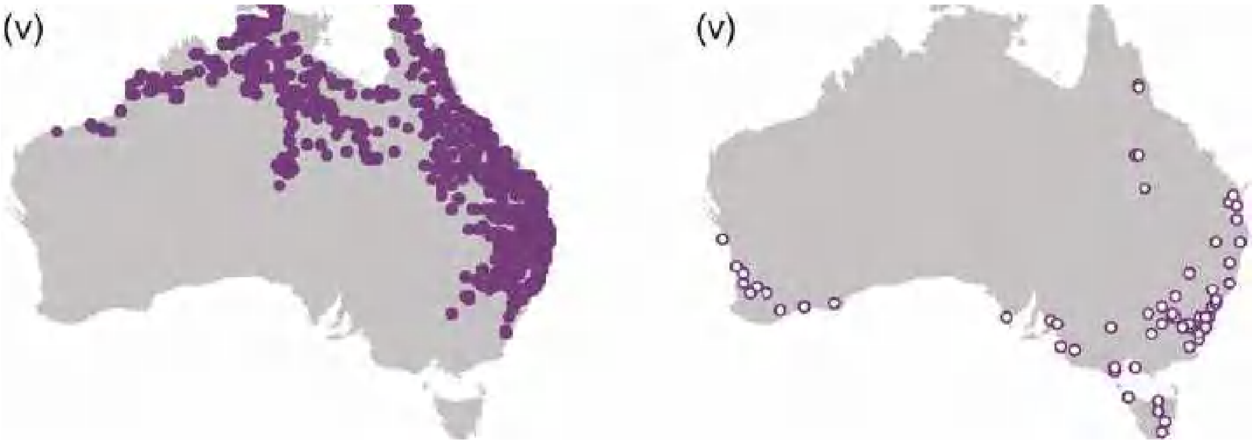


Figure 11.16 Release locations in Australia for five species of dung beetle where the populations (a) persisted and (b) failed to persist. (i) *Euoniticellus africanus*, (ii) *E. intermedius*, (iii) *Onitis alexis*, (iv) *Onthophagus binodis* and (v) *Onthophagus gazella*.

Source: From Duncan *et al.* (2009).

the economic importance of conserving dung beetles in the UK

If the introduction of appropriate dung beetles can improve agricultural productivity and reduce pest problems in Australasia, it is worth asking: what is the monetary value of dung beetles in areas where they are endemic? This is particularly pertinent if any agricultural practice puts dung beetles at risk, such as aspects of intensive agriculture (deep tillage, high stocking and fertiliser rates) or the treatment of cattle with certain anthelmintics (to counter parasitic worms). Dung beetles in cattle-grazed pasture provide ecosystem functions that benefit humans. These 'ecosystem services' include reduced pasture fouling and increased nitrogen cycling (enhancing herbage and livestock productivity), as well as reduced pest fly populations and reduced transmission of livestock gastrointestinal parasites (boosting cattle weight gains and milk production). Using a standard environmental economics framework, Beynon *et al.* (2015) estimated that dung beetles are worth £367 million per year to the UK cattle industry. They concluded that the adoption of agri-environment schemes designed to protect dung beetles could save the industry £40.2 million per year. England's Entry Level Stewardship Scheme, for example, attracts a government subsidy in returnfor the adoption of sympathetic pasture management to minimise erosion, reduce fertiliser and pesticide application and provide refuges for wildlife including beetles. The impact of anthelmintic treatment of cattle on dung beetle activity was quantified in a mesocosm experiment with and without the excretion of anthelmintic in quantities sufficient to affect the beetles. The authors concluded that cessation of the, largely unnecessary, treatment of adult cattle with anthelmintics could save the industry an extra £6.2 million per year.

11.3.4 Consumption of carrion

three categories of organisms attack dead bodies: decomposers, ...

When considering the decomposition of dead bodies, it is helpful to distinguish three categories of organisms that attack carcasses. Again, both decomposers and invertebrate detritivores have a role to play but scavenging vertebrates are often also of considerable importance. Mice carcasses placed on soil with intact communities of bacteria, archaea and fungi decompose two to three times faster than their counterparts on sterilised soil (Lauber *et al.*, 2014). This testifies to the important role that decomposers can play, but only if the carcass is not found quickly by detritivores.

> ... invertebrate detritivores ...

Certain components of animal corpses are particularly resistant to attack and are the slowest to disappear. However, some invertebrate detritivores possess the enzymes to deal with them. For example, the blowfly larvae of *Lucilia* species produce a collagenase that can digest the collagen and elastin present in tendons and soft bones. The chief constituent of hair and feathers, keratin, forms the basis of the diet of species characteristic of the later stages of carrion decomposition, in particular tineid moths and dermestid beetles. The midgut of these insects secretes strong reducing agents that break the resistant covalent links binding together peptide chains in the keratin. Hydrolytic enzymes then deal with the residues. Among the decomposers, fungi in the family Onygenaceae are specialist consumers of horn and feathers.

APPLICATION 11.3 Forensic entomology and microbiology

Some insect species associated with decomposing vertebrate corpses (including human bodies) act out their life histories in a sufficiently predictable way to act as a biological clock of time of death. Thus, calliphorid blowflies, among the first colonisers, can provide forensic scientists with useful information about time of death for two to three weeks postmortem; in comparison, medical examiner's estimates are limited to a day or two (Amendt *et al*., 2004). Forensic examination involves estimating time to reach the insect life stage observed in a corpse, which depends on environmental conditions, particularly temperature (Figure 11.17). It is important to incorporate region-specific information because different species develop at different rates and the suite of species involved varies regionally and seasonally.

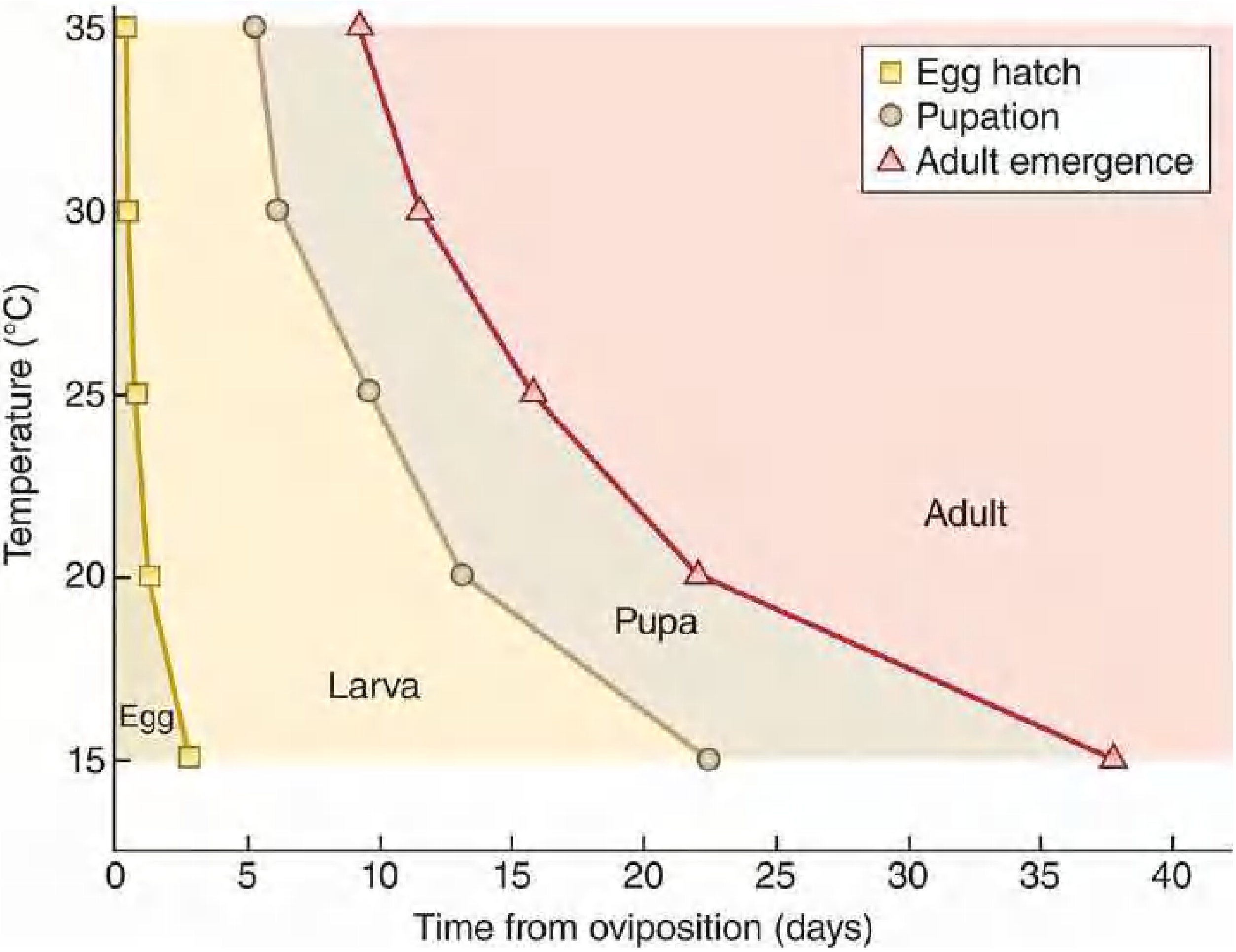


Figure 11.17 Growth curves for the forensically important holarctic blowfly *Protophormia terraenovae*. Shown is the time required to reach the larval, pupal and adult stages at 15, 20, 25, 30 and 35°C.

Source: From Grassberger and Reiter (2002).

A succession of insect species arrive on a corpse in a reasonably predictable fashion, providing forensic scientists with a further opportunity to estimate the minimum postmortem interval. In one study, pig carcasses, which have figured prominently in forensic entomology experiments, were set out in an Australian forest near the site of a murder. The researchers recorded for 125 days the timing of appearance of insect species (and their separate life stages) and their subsequent disappearance from 10 pig carcasses (Table 11.1). Using these summary statistics of taxon arrival and departure, Archer (2014) determined that the period 16–34 days postmortem most closely matched the actual pattern of occurrence of insects on the human corpse. This evidence was presented at the murder trial, where eyewitness testimony also indicated 21 days from death to body discovery. The suspects were found guilty.

Table 11.1 Summary statistics for date of first appearance of insects on a 'corpse' and of their subsequent disappearance. Shown are dates for life stages of selected insect taxa on 10 pig carcasses in an Australian forest in summer.

Source: After Archer (2014).

Taxon	**Appearance day**		**Disappearance day**	
	Mean (± SD)	**Range**	**Mean (± SD)**	**Range**
Diptera				
Calliphora augur instar 3	4.0 (±1.3)	3–7	8.3 (±2.1)	5–11
Calliphora stygia instar 3	3.9 (±1.2)	3–6	14.2 (±6.8)	7–25
Australophyra rostrata instar 1	8.1 (3.3)	4–15	24.9 (±15.7)	4–55
Australophyra rostrata instar 2	9.1 (±3.6)	4–15	40.8 (±28.9)	15.83
Australophyra rostrata instar 3	11.7 (±4.6)	6–20	66.5 (±29)	26–112
Australophyra rostrata adult	47.1 (±23.5)	17–77	64.4 (±25.5)	25–91
Chrysomya varipes pupa	14.9 (±3.3)	11–20	46 (±7.8)	36–62
Coleoptera				
Ptomaphila lacrymosa larvae	14.5 (±7.5)	7–29	33.6 (±5.8)	23–40
Saprinus spp adult	5.6 (±2.4)	3–10	70.8 (±23.7)	21–105
Dermestes maculatus adult	11.6 (±7.7)	6–32	46.7 (±16.3)	21–70
Hymenoptera				
Chalcidae parasite of *Ch. varipes*	28.1 (±7.8)	16–42	42.1 (±9.2)	25–59

Certain bacteria and fungi are also particularly associated with decay of human corpses and may provide further useful information in forensic investigations. For example, the fungi *Dichotomomyces cejpii* and *Talaromyces* spp. are characteristic of soil associated with a human corpse, and may have potential as above-ground indicators of the presence of a clandestine grave (Tranchida *et al.*, 2014).

... and vertebrate detritivores

Many carcasses of a size to make a single meal for a vertebrate scavenging detritivore will be removed completely within a very short time of death, leaving nothing for bacteria, fungi or invertebrates. This role is played, for example, by arctic foxes and skuas in polar regions, by crows, gluttons and badgers in temperate areas, and in the tropics by a wide variety of birds and mammals, including kites, jackals and hyenas. Individual scavenger species may take only part of the carcass. Salmon (*Oncorhynchus* spp.) accumulate almost all their body mass at sea and after returning to spawn in their natal streams in North America their energy-rich carcasses are preyed upon or scavenged by a variety of species: bears (*Ursus* spp.) characteristically consume brain, muscle or ovary tissue, wolves (*Canis lupus*) generally take just the head, while bald eagles

(*Haliaeetus leucocephalus*), gulls and corvids often scavenge the leavings of bears and wolves (Field & Reynolds, 2013).

The chemical composition of the diet of carrion-feeders is quite distinct from that of other detritivores, and this is reflected in their complement of enzymes. Carbohydrase activity is weak or absent, but protease and lipase activity is vigorous. Carrion-feeding detritivores possess basically the same enzymatic machinery as carnivores, reflecting the chemical identity of their food. In fact, many species of carnivore (such as lions, *Panthera leo*) are also opportunistic carrion-feeders while classic carrion-feeders such as hyenas (*Crocuta crocuta*) sometimes operate as carnivores.

APPLICATION 11.4 Ecosystem services provided by vultures

Declines in the populations of vultures have brought into sharp focus the importance of the waste-disposal ecosystem services they normally provide. On the Indian subcontinent, for example, the white-backed vulture (*Gyps bengalensis*), and related species, declined by 97% or more between 1990 and 2000 (Galligan *et al.* 2014). The resulting proliferation of rotting dead buffalo posed obvious health risks. But the consequences were more dire for a particular religious group, the Parsees, because their funeral rite – placing bodies naked in a special tower to be completely stripped by vultures – was no longer viable. Vultures feeding on domesticated ungulates that have been treated shortly before death with diclofenac, a non-steroidal anti-inflammatory drug, die of visceral gout and kidney failure. The banning of diclofenac in 2006 seems to have halted the decline in Asia but vulture abundance may now be threatened in southern Europe where approval has recently been granted for veterinary diclofenac use (Margalida *et al.*, 2014).

Chapter 12
Parasitism and Disease

12.1 Introduction: parasites, pathogens, infection and disease

Previously, in Chapter 9, we defined a parasitic organism as one that obtains its nutrients from one or a very few host individuals, normally causing harm but not causing death immediately. We must follow this now with some more definitions, since there are a number of related terms that are often misused.

When parasites colonise a host, that host is said to harbour an *infection*. Only if that infection gives rise to symptoms that are clearly harmful to the host should the host be said to have a *disease*. With many parasites, there is a presumption that the host can be harmed, but no specific symptoms have as yet been identified, and hence there is no disease. '*Pathogen*' is a term that may be applied to any parasite that gives rise to a disease (i.e. is 'pathogenic'). Thus, measles and tuberculosis are infectious diseases (combinations of symptoms resulting from infections). Measles is the result of a measles virus infection; tuberculosis is the result of a bacterial (*Mycobacterium tuberculosis*) infection. The measles virus and *M. tuberculosis* are pathogens. But measles is not a pathogen, and there is no such thing as a tuberculosis infection.

Parasites are an influential group of organisms in the most direct sense. Millions of people are killed each year by various types of infection, and many millions more are debilitated or deformed. There are more than 250 million cases of elephantiasis at present, over 200 million cases of schistosomiasis, and the list goes on. When the effects of parasites on domesticated animals and crops are included, the cost in terms of human misery and economic loss becomes immense. As we prepare this text for publication, we are living through the first months of the COVID-19 global pandemic.

Parasites are also important simply by virtue of their pervasiveness. An organism in a natural environment that does not harbour several species of parasite is a rarity. Moreover, many parasites and pathogens are host-specific or at least have a limited range of hosts. Thus, the conclusion seems unavoidable that more than 50% of the species on earth, and many more than 50% of individuals, are parasites. But do parasites affect their animal and plant hosts more generally in the way we see for medical and veterinary effects? This is one of the key questions we will address in this chapter.

12.2 The diversity of parasites

The language and jargon used by plant pathologists and animal parasitologists are often very different, and there are important differences in the ways in which animals and plants serve as habitats for parasites, and in the way they respond to infection. But for the ecologist, the differences are less striking than the similarities, and we therefore deal with the two together. One distinction that is useful, though, is that between microparasites and macroparasites (May & Anderson, 1979).

micro- and macroparasites

Microparasites are small and often intracellular, and they multiply directly within their host, where they are often extremely numerous. Hence, it is generally difficult, and usually inappropriate, to estimate precisely the number of microparasites in a host. The number of infected hosts, rather than the number of parasites, is the parameter usually studied. For example, a study of a measles epidemic will involve counting the number of cases of the disease, rather than the number of particles of the measles virus.

Macroparasites have a quite different biology: they grow but do not multiply in their host, and then produce specialised infective stages that are released to infect new hosts. The macroparasites of animals mostly live on the body or in the body cavities (e.g. the gut), rather than within the host cells. In plants, they are generally intercellular. It is usually possible to count or at least estimate the numbers of macroparasites in or on a host (e.g. worms in an intestine or lesions on a leaf), so the numbers of parasites as well as the numbers of infected hosts can be studied by the epidemiologist.

direct and indirect life cycles: vectors

Cutting across the distinction between micro- and macroparasites, parasites can also be subdivided into those that are transmitted directly from host to host and those that require a vector or intermediate host for transmission and therefore have an indirect life cycle. The term ‘vector’ signifies an animal carrying a parasite from host to host, and some vectors play no other role than as a carrier; but many vectors are also intermediate hosts within which the parasite grows and/or multiplies. Indeed, parasites with indirect life cycles may elude the simple micro/macro distinction. For example, schistosome parasites spend part of their life cycle in a snail and part in a vertebrate (in some cases a human). In the snail, the parasite multiplies and so behaves as a microparasite, but in an infected human the parasite grows and produces eggs but does not itself multiply, and so behaves as a macroparasite.

biotrophic and necrotrophic parasites

We can also draw a distinction between parasites that kill hosts and then continue to live on them (necrotrophic parasites) and those for which the host must be alive (biotrophic parasites). Strictly speaking, necrotrophic parasites could equally well be thought of as predators insofar as host death is often inevitable and sometimes quite rapid, and once the host is dead they become pioneer detritivores, one step ahead of their competitors. Examples include the blowfly of sheep, *Lucilia cuprina*, that lays eggs on the living host that hatch into larvae (maggots) that eat into the sheep’s flesh, possibly killing it and continuing to exploit the carcass after death; and *Botrytis fabi*, a typical fungal necroparasite of plants that develops in the leaves of the bean, *Vicia faba*, forming blotches of dead tissues and continuing to develop as a decomposer, producing dispersive spores. For as long as the host is alive, however, necroparasites share many features with other types of parasite. Thus, this is one example – and there are many – where the organisms themselves snub their noses at our attempts to fit them all into neat categories.

12.2.1 Microparasites

Probably the most obvious microparasites are the bacteria and viruses that infect animals (such as the measles virus and the typhoid bacterium) and plants (e.g. the yellow net viruses of beet and tomato and the bacterial crown gall disease). The other major group of microparasites affecting animals is the protozoa (e.g. the trypanosomes that cause sleeping sickness and the *Plasmodium* species that cause malaria; Figure 12.1a). In plant hosts some of the simpler fungi behave as microparasites.

Figure 12.1 A range of parasites. (a) *Plasmodium falciparum* in a human blood smear with red blood cells. (b) *Onchocerca volvulus* – colour-enhanced electron micrograph. (c) Lesions of black stem rust on wheat. (d) Dodder (*Cuscuta* spp.) engulfing a tree in Florida, USA.

The transmission of a microparasite from one host to another can be in some cases almost instantaneous, as in venereal disease and the short-lived infective agents carried in the water droplets of coughs and sneezes (influenza, measles, etc.). In other species the parasite may spend an extended dormant period ‘waiting’ for its new host. This is the case with the ingestion of food or water contaminated with the protozoan *Entamoeba histolytica*, which causes amoebic dysentery, and with the plant parasite *Plasmodiophora brassicae*, which causes ‘club root disease’ of crucifers.

Alternatively, a microparasite may depend on a vector for its spread. The two most economically important groups of vector-transmitted protozoan parasites of animals are the trypanosomes, transmitted by various vectors including tsetse flies (*Glossina* spp.) and causing sleeping sickness in humans and nagana in domesticated (and wild) mammals, and the various species of *Plasmodium*, transmitted by anopheline mosquitoes and causing malaria. In both these cases, the flies act also as intermediate hosts, the parasite multiplying within them.

Many plant viruses are transmitted by aphids. In some 'non-persistent' species (e.g. cauliflower mosaic virus), the virus is only viable in the vector for one hour or so and is often borne only on the aphid's mouthparts. In other 'circulative' species (e.g. lettuce necrotic yellow virus), the virus passes from the aphid's gut to its circulatory system and thence to its salivary glands. Here, there is a latent period before the vector becomes infective, but it then remains infective for an extended period. Finally, there are 'propagative' viruses (e.g. the potato leaf roll virus) that multiply within the aphid. Nematode worms are also widespread vectors of plant viruses.

12.2.2 Macroparasites

The parasitic helminth worms are major macroparasites of animals, including humans. In 2008, it was estimated that around one-third of the approximately 3 billion people living on less than two US dollars per day in the developing world had a helminth infection (Hotez *et al*., 2008). The intestinal nematodes of humans, for example, all of which are transmitted directly, are perhaps the most important human intestinal parasites, both in terms of the number of people infected and their potential for causing ill health. There are also many types of medically important animal macroparasites with indirect life cycles. For example, the tapeworms are intestinal parasites as adults, absorbing host nutrients directly across their body wall and proliferating eggs that are voided in the host's faeces. The larval stages then proceed through one or two intermediate hosts before the definitive host (in these cases, the human) is reinfected. The schistosomes, as we have seen, infect snails and vertebrates alternately. Human schistosomiasis affects the gut wall where eggs become lodged, and also affects the blood vessels of the liver and lungs when eggs become trapped there. Filarial nematodes are another group of long-lived parasites of humans. These all require a period of larval development in a blood-sucking insect. One, *Wuchereria bancrofti*, does its damage (Bancroftian filariasis) by the accumulation of adults in the lymphatic system (classically, but only rarely, leading to elephantiasis). Larvae (microfilariae) are released into the blood and are ingested and may ultimately be transmitted into another host by mosquitoes. Another filarial nematode, *Onchocerca volvulus*, which causes 'river blindness', is transmitted by adult blackflies (the larvae of which live in rivers, hence the name of the disease). Here, though, it is the microfilariae that do the major damage when they are released into the skin tissue and reach the eyes (Figure 12.1b).

In addition, there are lice, fleas, ticks and mites and some fungi that attack animals. Lice spend all stages of their life cycle on their mammalian or avian host, and transmission is usually by direct physical contact between host individuals, often between mother and offspring. Fleas, by contrast, lay their eggs and spend their larval lives in the 'home' (usually the nest) of their host (again, a mammal or a bird). The emerging adult then actively locates a new host individual, often jumping or walking considerable distances in order to do so.

Plant macroparasites include the higher fungi that give rise to the mildews, rusts and smuts (each lesion being equivalent to a worm in an animal, giving rise, in this case, to infective spores), as well as the gall-forming and mining insects, and some flowering plants that are themselves parasitic on other plants. Direct transmission is common amongst the fungal macroparasites of plants. For example, in the development of mildew on a crop of wheat, infection involves contact between a spore (usually wind-dispersed) and a leaf surface, where the spore begins to grow, eventually becoming apparent as a lesion of altered host tissue, which subsequently matures and starts to produce new spores. By contrast, indirect transmission of plant macroparasites via an intermediate host is common amongst the rust fungi. For example, in black stem rust (Figure

12.1c), infection is transmitted from an annual grass host (especially the cultivated cereals such as wheat) to the barberry shrub (*Berberis vulgaris*) and from the barberry back to wheat. Infections on the cereal are polycyclic – within a season, spores may infect and form lesions that release spores that infect further cereal plants – but the barberry is a long-lived shrub and the rust is persistent within it. Infected barberry plants may therefore serve as persistent foci for the spread of the rust into cereal crops.

holo- and hemiparasitic plants

Plants in a number of families have become specialised as parasites on other flowering plants. These are of two quite distinct types. Holoparasites, such as dodder (*Cuscuta* spp.; Figure 12.1d), lack chlorophyll and are wholly dependent on the host plant for their supply of water, nutrients and fixed carbon. Hemiparasites, on the other hand, such as the mistletoes (*Phoraradendron* spp.), are photosynthetic but have poorly developed root systems of their own, or none at all. They form connections with the roots or stems of other species and draw most or all of their water and mineral nutrients from the host.

brood and social parasitism

Brood parasites are organisms that rely on others to rear their offspring. The most widely recognised example is the cuckoo. Usually, a host and its parasite come from very distant systematic groups (e.g. mammals and bacteria, fish and tapeworms, plants and viruses). In contrast, brood parasitism usually occurs between closely related species and even between members of the same species. Yet the phenomenon falls clearly within the definition of parasitism, since a brood parasite does indeed ‘obtain its nutrients from one or a few host individuals, normally causing harm but not causing death immediately’. Brood parasitism is well developed in social insects (sometimes then called social parasitism), where the parasites use workers of another, usually very closely related species to rear their progeny, but as noted already, the phenomenon is best known in birds.

Bird brood parasites lay their eggs in the nests of other birds, which then incubate and rear them. They usually depress the nesting success of the host. Amongst ducks, *intra*specific brood parasitism appears to be most common. Most brood parasitism, however, is *inter*specific. About 1% of all bird species are brood parasites. They usually lay only a single egg in the host’s nest and may adjust the host’s clutch size by removing one of its eggs. The developing parasite may evict the host’s eggs or nestlings and harm any survivors by monopolising parental care. There is therefore the potential for brood parasites to have profound effects on the population dynamics of the host species. However, the frequency of parasitised nests is usually very low (less than 3%), and some time ago Lack (1963) concluded that ‘the cuckoo is an almost negligible cause of egg and nestling losses amongst English breeding birds’.

12.3 Hosts as habitats

The essential difference between the ecology of parasites and that of free-living organisms is that the habitats of parasites, the hosts, are themselves alive. Hosts are reactive. A host can respond to the presence of a parasite by changing its nature, developing immune reactions to the parasite, or by otherwise killing the parasite. Hosts are also able to evolve. And in the case of many animal parasites, hosts are mobile and have patterns of movement that dramatically affect dispersal (transmission) from one habitable host to another.

12.3.1 The distribution of parasites within host populations: aggregation

The distribution of parasites among their host patches – individual hosts within the host population – are rarely random. For any particular species of parasite it is usual for many hosts to harbour few or no parasites, and a few hosts to harbour many, i.e. the distributions are usually aggregated or clumped (Figure 12.2). The hosts with many parasites are in turn likely to be the most important in the onward transmission of the parasites to other hosts. We return to these 'superspreaders' in Section 12.5.2.

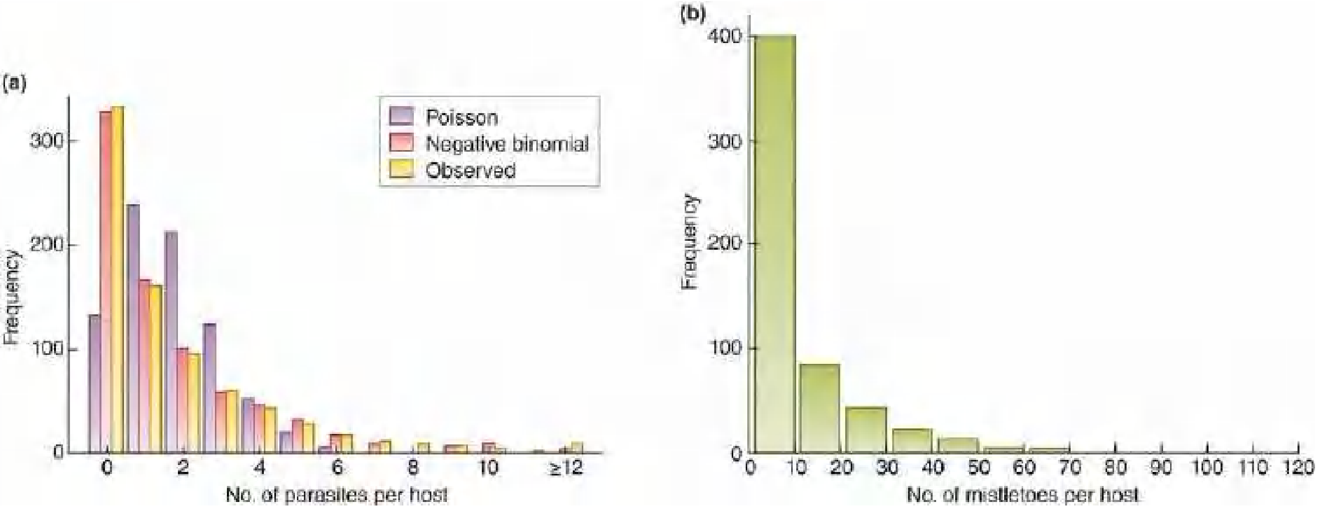


Figure 12.2 Aggregated distributions of parasite numbers per host. (a) Crayfish, *Orconectes rusticus*, infected with the flatworm *Paragonimus kellicotti*. The distribution is significantly different from Poisson (random) ($X^2 = 723$, $P < 0.001$) but conforms well with a 'negative binomial', which is good at describing aggregated distributions ($X^2 = 12$, $P \approx 0.4$). (b) Distribution of parasitic mistletoe plants, *Tristerix aphyllus*, on their cactus hosts, *Echinopsis chilensis*, in Chile. Again the distribution conforms well to a negative binomial ($X^2 = 2.7$, $P \approx 0.6$), but not to a Poisson ($X^2 = 172$, $P < 0.001$).

Source: (a) After Stromberg *et al.* (1978) and Shaw & Dobson (1995). (b) After Medel *et al.* (2004).

prevalence, intensity and mean intensity

In such populations, the mean density of parasites (mean number per host) may have little meaning. In a human population in which only one person is infected with anthrax, the mean density of the causative agent, *Bacillus anthracis*, is a particularly useless piece of information. A more useful statistic, especially for microparasites, is the prevalence of infection: the proportion or percentage of a host population that is infected. On the other hand, especially for macroparasites, infection may often vary in severity between individuals, related to the number of parasites that they harbour. The number of parasites in or on a particular host is referred to as the intensity of infection. The mean intensity of infection is then the mean number of parasites per host in a population (including those hosts that are not infected).

Aggregations of parasites within hosts may arise because individual hosts vary in their susceptibility to infection (whether due to genetic, behavioural or environmental factors), or because individuals vary in their exposure to parasites (Wilson *et al.*, 2002). The latter is especially likely to arise because of the local nature of transmission, and especially when hosts are relatively immobile. Infection then tends to be concentrated, at least initially, close to an original source of infection, and to be absent in individuals in areas that infection has yet to reach, or where it was previously but the hosts have recovered. Medel *et al.* (2004) discuss how the aggregation of mistletoe in Figure 12.2b results from a balance between the behaviour of the mockingbird vector of the mistletoe seeds, *Mimus thenca*, and the resistance traits of the host cactus.

12.3.2 Host specificity: host ranges and zoonoses

We saw in the chapters on the interactions between predators and their prey that there is often a high degree of specialisation of a particular predator species on one or a few species of prey. The specialisation of parasites is even more striking. Some examples are shown in Figure 12.3. Many are only able to live on one particular host species, and even for those with wider host ranges, the acceptable hosts are a tiny subset of those available. The overwhelming majority of other organisms are quite unable to serve as hosts. Often, we do not know why.

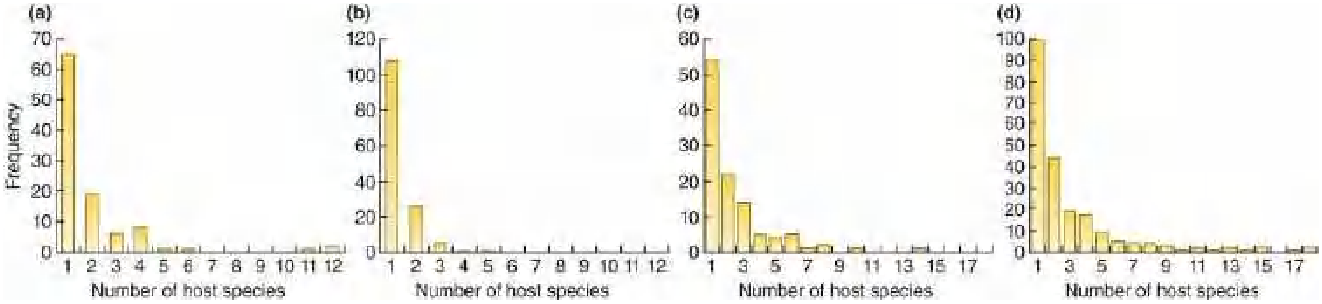


Figure 12.3 Many parasites are specialists, often attacking just one host species. Frequency distributions of the numbers of host species parasitised by (a) chewing lice of the order Amblyocera parasitising rodents worldwide, (b) chewing lice of the order Ischnocera parasitising rodents worldwide, (c) cestode worms parasitising birds in Azerbaijan, and (d) trematode worms parasitising birds in Azerbaijan.

Source: After Poulin & Keeney (2007).

There are some patterns amongst the variations in host specificity, but none that could be described as general or universal. Among the virus and helminth parasites of primates, for example, host specificity is lower for those transmitted by a vector, often a biting insect, than for those that require close contact between infectious and susceptible hosts for transmission to occur (Figure 12.4a, c). This makes sense, insofar as close or direct contact implies a sharing of habitat that is most likely among similar individuals and hence, often, individuals of the same species. A vector, by contrast, can much more readily acquire an infection from a host individual of one species and then cross habitat boundaries and pass it on to another species. On the other hand, for protozoa, where vector-borne infections are particularly common, the pattern, if anything, is the opposite (Figure 12.4b). This serves simply to illustrate the absence of simple answers, and the range of factors that combine to determine host specificity in particular cases.

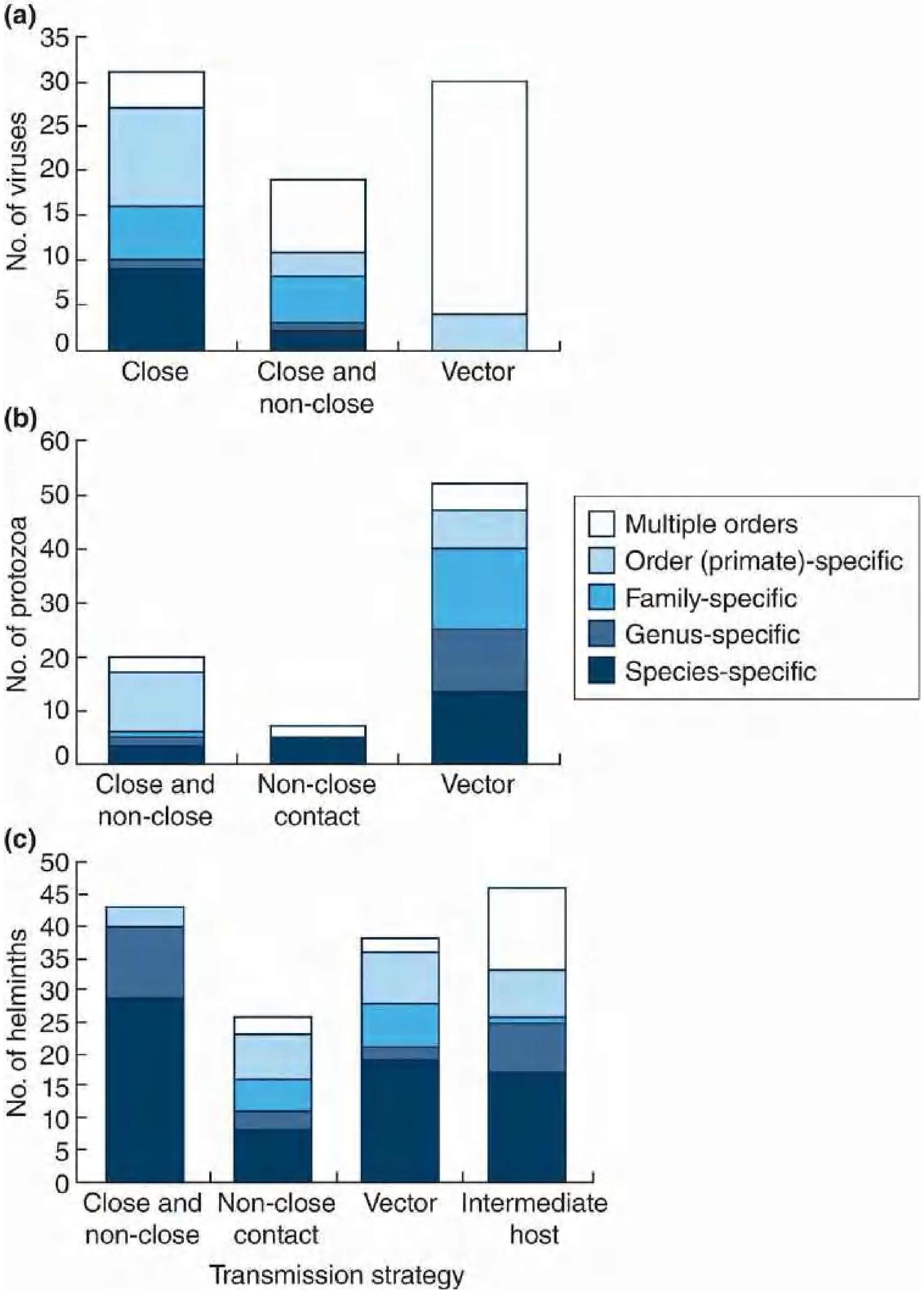
(a)
No. of viruses
0
5
10
15
20
25
30
35
Close
Close and non-close
Vector
(b)
No. of protozoa
0
10
20
30
40
50
60
Close and non-close
Non-close contact
Vector
Multiple orders
Order (primate)-specific
Family-specific
Genus-specific
Species-specific
(c)
No. of helminths
0
5
10
15
20
25
30
35
40
45
50
Close and non-close
Non-close contact
Vector
Intermediate host
Transmission strategy

Figure 12.4 Patterns in the effect of transmission strategy on host specificity in three groups of parasites of wild primates: (a) viruses (*n* = 80), (b) protozoa (*n* = 79), and (c) helminths (*n* = 153). Some parasites were transmitted by both close and non-close contact.

Source: After Pedersen *et al.* (2005).

natural and dead-end hosts

It is also important to realise that within a species' host range, not all host species are equal. Species outside the host range are relatively easily characterised: the parasite cannot establish an infection within them. But for those inside the host range, the response may vary from a serious pathology and certain death, on the one hand, to an infection with no overt symptoms on the other. And cutting across these distinctions is another continuum, ranging from species in which infected hosts shed masses of infectious stages capable of infecting further hosts, to other host species that are 'dead-ends' – susceptible to infection themselves, and perhaps liable to become diseased, but incapable of passing the infection on. What is more, it is often the 'natural' host of a parasite, the one with which it has coevolved (see Section 12.4), in which infection is asymptomatic, whereas in dead-end hosts, where there can have been no coevolution, infection may often give rise to a fatal pathology.

APPLICATION 12.1 Zoonotic infections

The issues of parasites crossing from natural hosts to other host species take on not just parasitological but also medical importance in the case of *zoonotic infections*. These are infections that circulate naturally, and have coevolved, in wildlife or domesticated animal species, but also have a pathological effect on humans. In these cases, we refer to the animal species as the 'reservoir' of infection and often say that the infection 'spills over' into humans, where it often reaches a dead end (see Viana *et al.*, 2014). A classic, long-recognised example is rabies (a viral infection). Rabies circulates naturally in a wide range of species, notably wild and domestic dogs, several bat species, and, in the USA, raccoons (*Procyon lotor*). Rabies may then be passed, say, to humans (where it is very often fatal, killing more than 69 000 people each year) if they are bitten by a rabid dog or bat. Vaccination of domestic dogs in Africa, for example, seems to be the most effective means not only of saving the lives of the dogs themselves and humans, but also of eliminating the infection from local wildlife species (Lankester *et al.*, 2014). The domestic dogs, in other words, are the main reservoir species.

However, humans are not always dead-ends. Onward human-to-human transmission may be possible and even powerful, as we have seen recently with COVID-19. Another well-known example is plague (infection with the bacterium *Yersinia pestis*). Plague is best known for having devastated the population of Europe in Mediaeval times, but it is still present in many parts of the world, circulating in a variety of wild rodents, though, mercifully, killing many fewer people than it did in the past (Stenseth *et al.*, 2008). In fact, plague also provides us with a good example of an infection crossing the species barrier and causing a conservation problem. *Y. pestis* circulates naturally within populations of a number of species of wild rodent in south-western USA in which there are few if any symptoms of infection. There are, however, other species, for example prairie dogs, *Cynomys* spp., also rodents, into which infection spills over, and which are so susceptible that populations are regularly annihilated by epidemics of plague. What is more, the black-footed ferret, *Mustela nigripes*, a highly endangered predator in the USA, is placed in double jeopardy by plague: it feeds almost entirely on prairie dogs but is itself also highly susceptible to plague (Salkeld *et al.*, 2016) habitat specificity within hosts

Most parasites are also specialised to live only in particular parts of their host. Malarial parasites live in the red blood cells of vertebrates. *Theileria* parasites of cattle, sheep and goats live in the lymphocytes of the mammal, and in the epithelial cells and later in the salivary gland cells of the tick that is the disease vector, and so on. Parasites may move to their target habitats. Indeed, when nematode worms (*Nippostrongylus brasiliensis*) were transplanted from the jejunum into the anterior and posterior parts of the small intestine of rats, they migrated back to their original habitat (Alphey, 1970). In other cases, habitat search may involve growth rather than bodily movement. For example, the fungal pathogen loose smut of wheat, *Ustilago tritici*, infects the exposed stigmas of wheat flowers and then grows as an extending filamentous system into the young embryo. Growth continues in the seedling, and the fungus mycelium keeps pace with the growth of the shoot. Ultimately, the fungus grows rapidly into the developing flowers and converts them into masses of spores.

12.3.3 Hosts as resources and reactors

We have seen in previous chapters that more or less all species are preyed upon by other species in the trophic levels above them, and that they compete with other individuals from their own and other species for resources in the trophic levels below them. For parasites, however, the same host species – indeed, the same host individual – typically acts as resource *and* predator. The idea is illustrated for a typical vertebrate host in Figure 12.5. Note that parasites within hosts sit at the heart of a web of interactions that is reminiscent of the predatory and competitive interactions we have previously discussed, but is different from them in important ways. In particular, in a community of free-living organisms, we might think of the levels in Figure 12.5, reading up, as plant resources at the bottom (water, nitrogen, phosphorus and so on), then, at the next level up, the plant species themselves that compete for those resources, followed, at the top, by the herbivores that prey upon the plants. The plant resources and the herbivores are, of course, completely separate entities. But if that middle represents not plants, but parasites within a host, then the resources for which they compete and the ‘predators’ that attack them, far from being separate, are all interlinked parts of the same host. The parasites themselves can be subject to exploitation, interference and apparent competition (see Chapter 8), all generated by the same host individual. These are three trophic levels – but not as we normally know them.

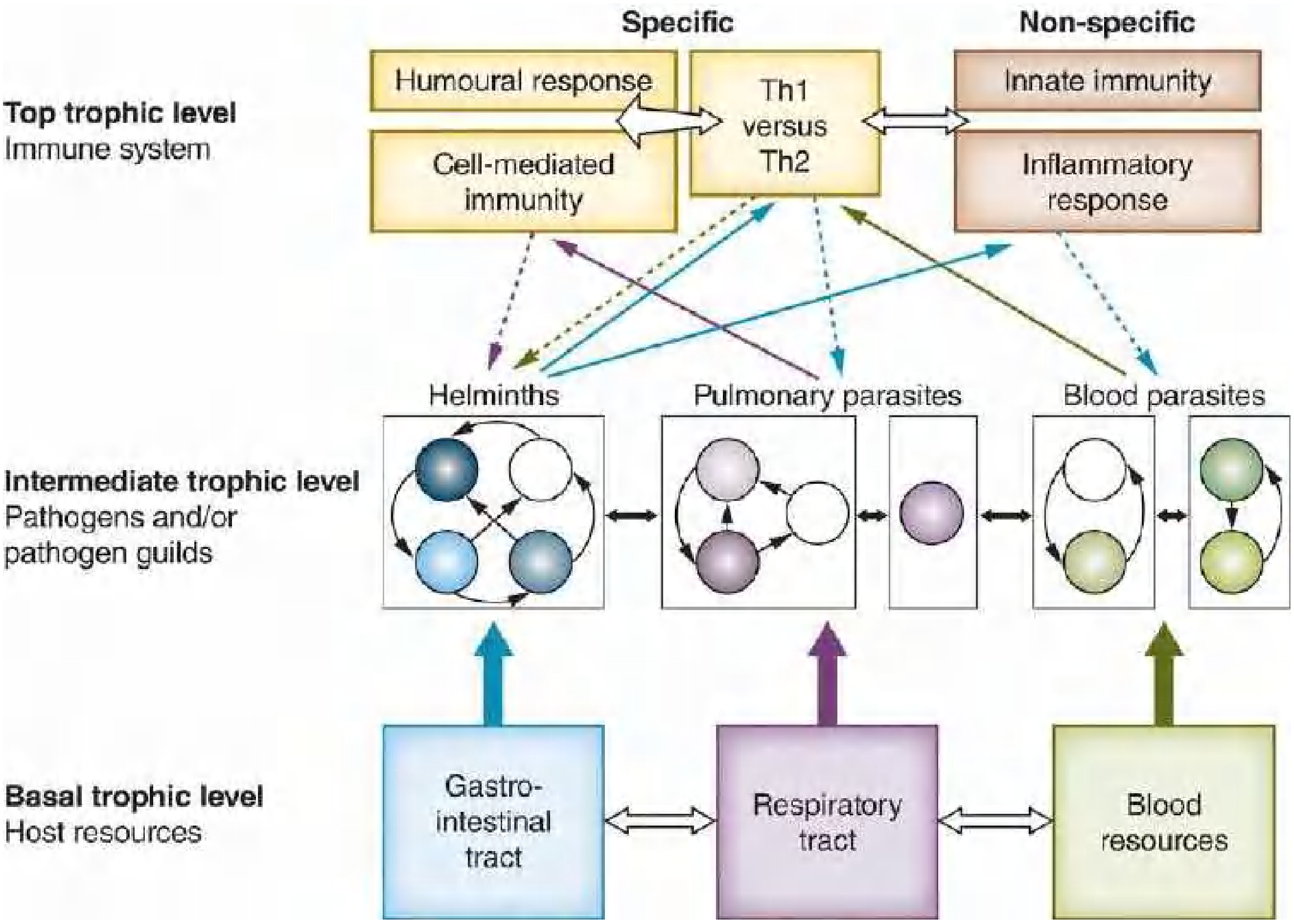


Figure 12.5 **Parasites within hosts: simultaneously both predator and prey.** A schematic representation of the way parasites occupy an intermediate trophic level within their hosts (in this case, vertebrates), consuming and competing for host resources (some of which are identified), while being attacked (preyed upon) by the various elements of the host's immune system. These elements include non-specific phagocytosing cells and chemicals that are active against infections in general, and specific responses that are effective against the parasites that induced them. These too include cellular and chemical (humoural) elements. Put very simply, responses are 'orchestrated' by T-helper (Th) cells, giving rise to two main types of response: the Th1 response targeting mostly intracellular viral and bacterial infections, and the Th2 response targeting helminths and smaller extracellular parasites. The horizontal open arrows in the top and bottom trophic levels acknowledge the linkages that exist because all elements are part of the same host. The vertical arrows represent the flow of energy from the host to the parasites. The middle trophic level comprises several species of parasite, arranged in guilds, with arrows between them indicating possible direct interactions. The arrows between the top and middle trophic levels represent the effects of the host immune system on the parasites (dashed arrows) and the effects of the parasites in inducing those responses (solid arrows).

Source: After Pedersen & Fenton (2006).

12.3.4 Hosts as reactors: resistance and recovery

The detailed mechanisms through which hosts repel or destroy the parasites that attack them, or tolerate those that they cannot escape – the 'natural history' of host defence – are massive topics in their own right. Broadly, the defences of vertebrates (see Pedersen & Fenton, 2006) are more sophisticated than those of either invertebrates (see Siva-Jothy *et al.*, 2005) or plants (see Dodds & Rathjen, 2010), though as we learn more about the capacities of each group, the gap between them appears to be diminishing. However, these will not be reviewed here. From an ecological point of view, what are more important are some key features, described here and in the next two subsections, that either unite or divide these different types of host defences.

acquired immunity: S-I-R not S-I-S

First, a host, in responding to and perhaps fighting off an infection, may acquire an immunity to subsequent reinfection. This is especially well developed in vertebrates, where the pathways of cell-mediated immunity (Figure 12.5) give rise to specific populations of cells able to attack the pathogens or parasites that provoked their production within the immune system. A host that has been infected once thus has a 'memory' of that infection, rendering it resistant to that pathogen to a degree that a naïve, susceptible host is not. Even in some plants that survive a pathogenic attack, 'systematic acquired resistance' to subsequent attacks may be elicited. For example, a tobacco plant infected on one leaf with tobacco mosaic virus can produce local lesions that restrict the virus infection locally, but the plant then also becomes resistant to new infections not only by the same virus but to other parasites as well. In some cases the process involves the production of 'elicitins', which have been purified and shown to induce vigorous defense responses by the host (Yu, 1995). In all these cases, three classes of host can be defined: susceptible to infection, denoted *S*, infected (and typically infectious), *I*, and recovered from but also resistant to infection, *R*. Hence, mathematical models that aim to help us understand the population dynamics of this whole class of infections (see below) are referred to as S-I-R models; whereas in those cases where hosts that recover from infection acquire no resistance and so are once again susceptible to infection, we refer to S-I-S models.

12.3.5 Hosts as reactors: the cost of resistance

trade-offs

To understand host defensive responses from an ecological point of view, we must remember that these responses are often costly to the host. Thus they may be subject to trade-offs – where resources invested in one component of fitness are thereby diverted from some other component (see Section 7.3). We see in Figure 12.6a, for example, that in fruit-flies, *Drosophila melanogaster*, increases in male reproductive activity are achieved at the cost of a reduced ability to fight off infection by the bacterium *Escherichia coli*. These flies also illustrate a more fundamental resistance–reproduction trade-off in that those flies genetically best able to resist infection by another bacterium *Providencia rettgeri*, have a lower intrinsic fecundity (Figure 12.6b) – though interestingly, this trade-off is only apparent under poor nutritional conditions, where resources are in short supply.

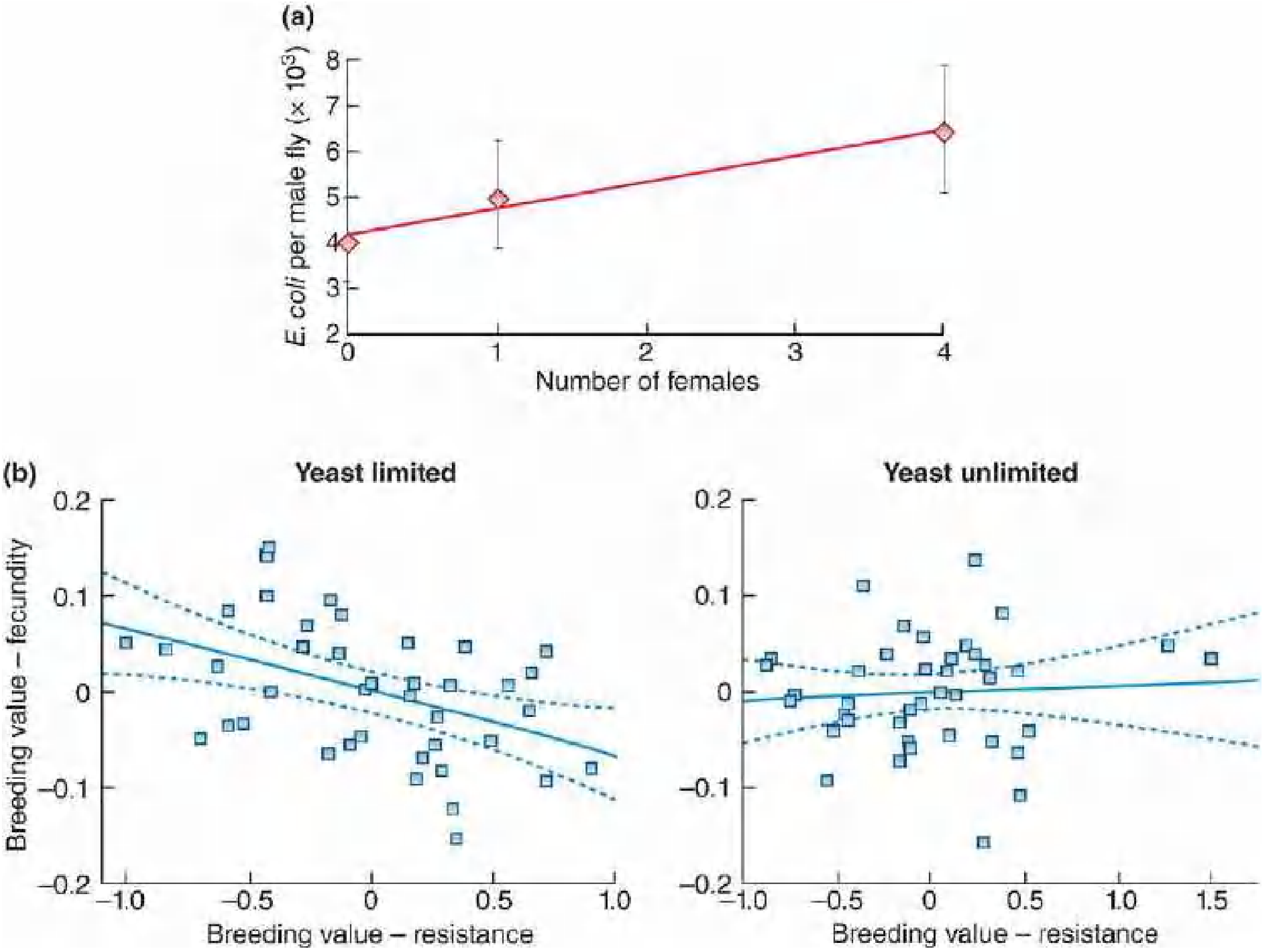


Figure 12.6 **Trade-offs reveal the costs of defending against parasites.** (a) The intensity of infection (number of *Escherichia coli* found) in fruit-fly males, *Drosophila melanogaster,* when they are maintained alone or with one or four virgin females that are replaced daily. (b) The genetic correlation among different strains of *D. melanogaster* between their fecundity (in the absence of infection) and their resistance to the bacterium *Providencia rettgeri*. This is shown both where yeast (food resources) is limited and a trade-off (negative correlation) is apparent (r = –0.44, P = 0.004), and where food is unlimited and there is no apparent trade-off (r = 0.069, P = 0.67). The 'breeding values' for fecundity and resistance are the deviations between strain values and the population means. The dotted lines are 95% CIs.

Source: (a) After McKean & Nunney (2001). (b) After McKean *et al.* (2008).

As far as trade-offs *within* an overall host response are concerned, the one that has received most attention is that between the so-called T-helper 1 (Th1) and Th2 arms of the vertebrate immune response (see Figure 12.5). Recall from that figure that the Th1 response targets mostly intracellular viral and bacterial infections, while the Th2 response primarily targets helminths and smaller extracellular parasites. An example is shown in Figure 12.7. African buffalo, *Syncerus caffer*, are infected by gastrointestinal nematodes, inducing a predominantly Th2 response, and the bacterium *Mycobacterium bovis* (causing bovine tuberculosis (BTB)) which induces a predominantly Th1 response. The buffalo exhibit a clear negative correlation between their infections with the two parasites (Figure 12.7a): individuals with worms are much less likely to have BTB, and indeed, buffalo herds with a higher prevalence of BTB tend to have a lower prevalence of worm infection. That this pattern is a reflection of an underlying Th1–Th2 trade-off is suggested, first, by the fact that worm-free buffalo have significantly stronger worm-protective Th2 immunity, as measured by the concentration of eosinophils (a type of white blood cell) in the blood. But, crucially, those with stronger Th2 immunity tend to have weaker Th1 immunity, as measured by the concentration in the blood of interferon-γ (Figure 12.7b) – though interestingly, again, this was apparent only when resources were in short supply: in the dry but not in the rainy season.

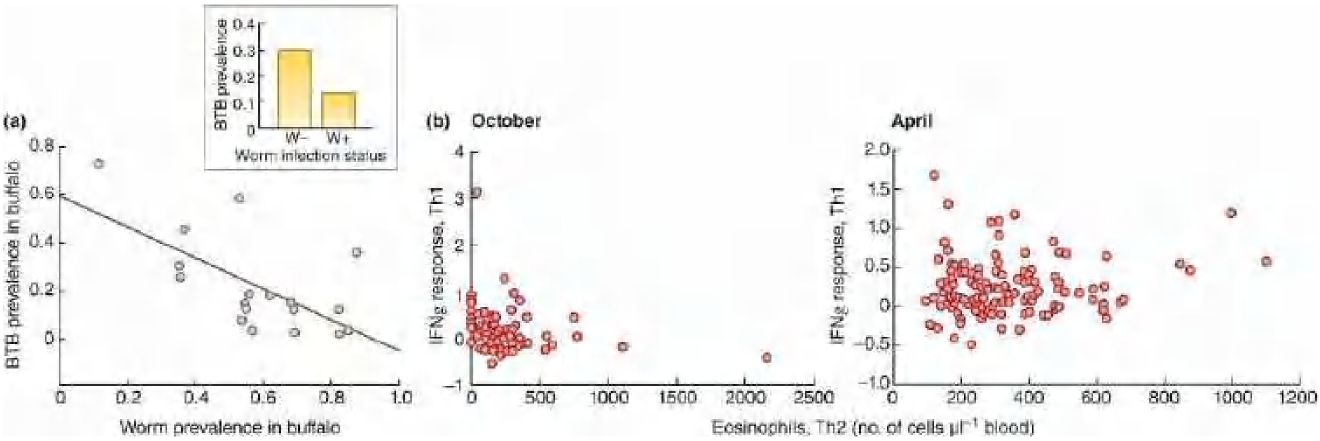


Figure 12.7 **Th1 and Th2 trade-offs in buffalo.** (a) A negative correlation (trade-off) is seen between bovine tuberculosis (BTB) and worm infection prevalence in different herds of African buffalo, *Syncerus caffer*, in South Africa ($r = -0.64$, $P < 0.0025$). The inset shows the equivalent trade-off between individuals across the entire population ($r = -0.21$, $P < 0.001$). (b) An apparent trade-off is seen in the buffalo between a Th1 and Th2 response in October, when resources are limited ($F = 5.7$, $P = 0.019$), but there is no apparent trade-off in April, when resources are plentiful ($F = 1.6$, $P = 0.20$). The Th1 response was measured as the log-transformed proportionate change when hosts were challenged with *Mycobacterium bovis* antigen. The Th2 response was the concentration of eosinophils in the blood. IFN, interferon.

Source: After Jolles *et al.* (2008).

APPLICATION 12.2 Indirect effects of therapeutic treatments

All of this reminds us that the repertoire of host responses is large, especially in vertebrates, and that the different elements of that response cannot be considered in isolation. This in turn suggests that in treating a host, or a host population, for one infection, possible knock-on effects on other infections need to be taken into account. This was examined in the case of the buffalo. Treating them with an anthelmintic drug to control their worms led to an increase in their Th1 (interferon-γ) immunity, suggesting that the treatment was releasing additional resources. However, this increase did not lead to any decrease in the chances of the buffalo acquiring BTB (Figure 12.8a). Their chances of *surviving* BTB infection, on the other hand, were significantly improved (Figure 12.8b). This is obviously beneficial to the individual host, but also increases both the number of hosts shedding infective bacteria, and the period of time over which they do so, enhancing the spread of disease and potentially doing more harm than good overall. Clearly at a time when anthelmintic treatment is being considered as a way of helping patients cope better with chronic microbial diseases like HIV/AIDS and tuberculosis, such indirect effects cannot be ignored (Ezenwa & Jolles, 2015).

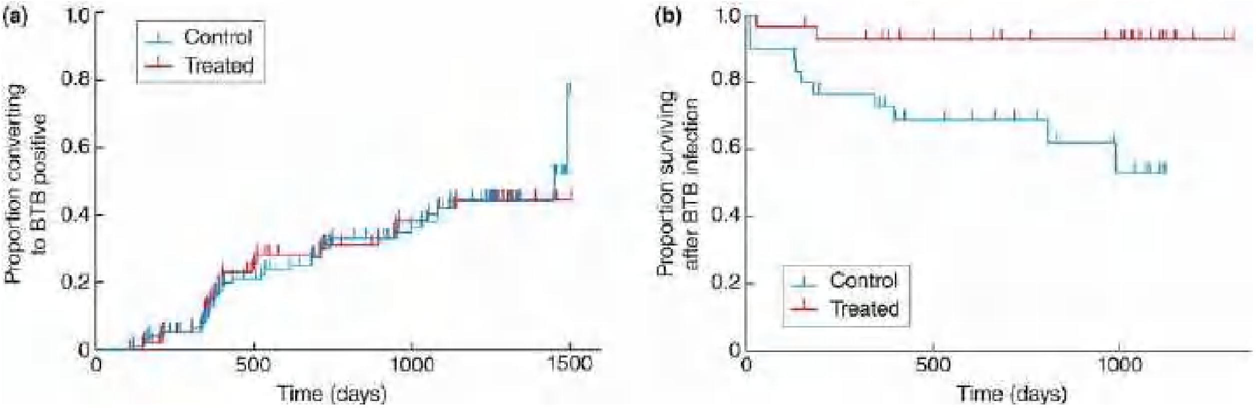
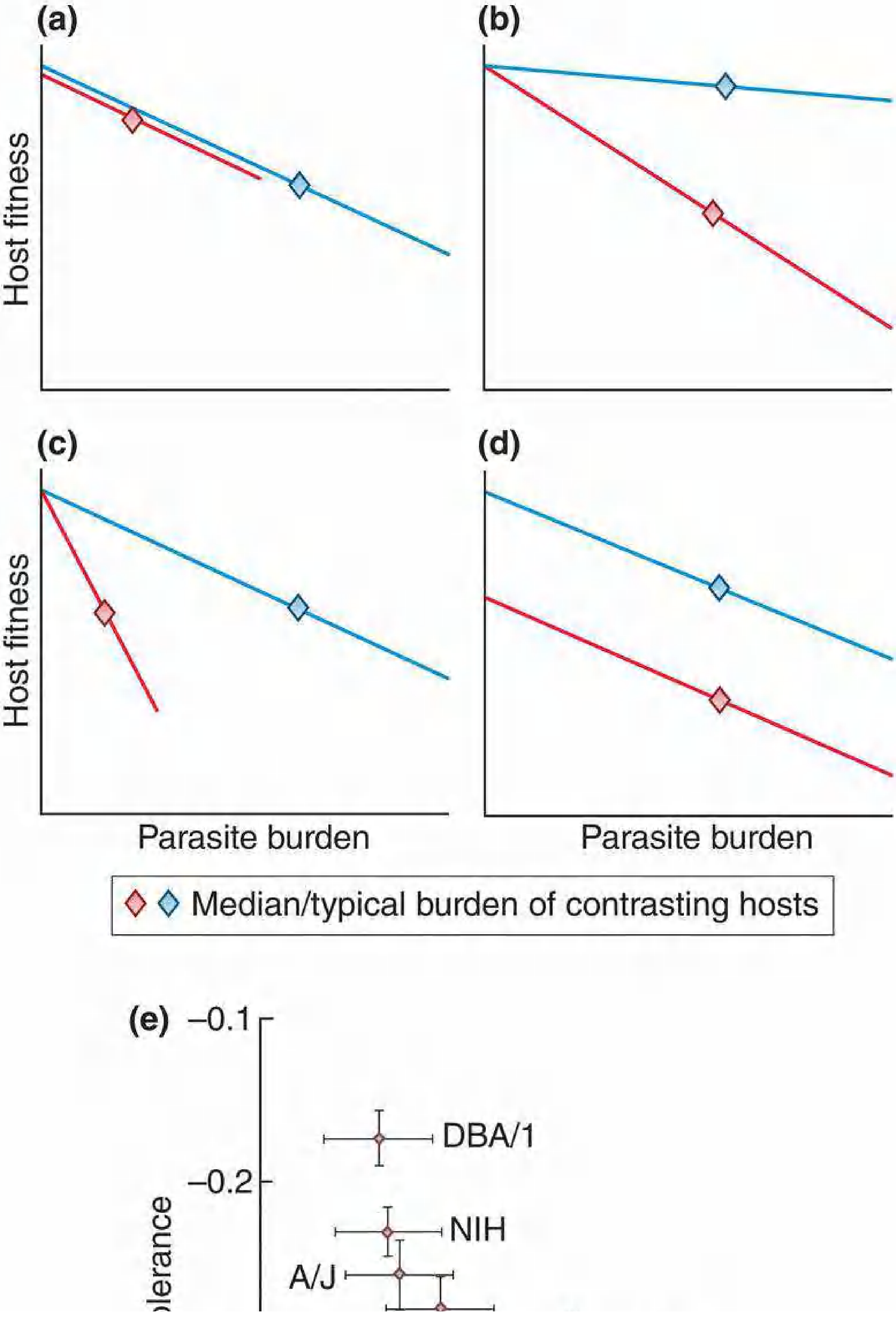


Figure 12.8 Contrasting effects at the individual and population level of treating African buffalo, *Syncerus caffer*, in South Africa with an anthelmintic drug. (a) Treated animals were no less likely to get bovine tuberculosis (BTB) than untreated controls. (b) However, the treated animals survived for much longer and so the proportion of their population with BTB was itself higher.

Source: After Ezenwa & Jolles (2015).

12.3.6 Hosts as reactors: resistance, tolerance and virulence

There is, perhaps, a tendency to assume that hosts will have evolved either to avoid parasitic infection in the first place or to rid themselves of infection once it has been acquired. Both are, in a sense, 'resistance', based on the simple idea that the fewer parasites a host is infected with, the greater its health or fitness will be. Thus, in Figure 12.9a we see two hosts differing in their resistance, in that both are indeed fitter the lower their parasite burden, but one is better able to keep its parasite burden down and so typically carries a lower burden. Its resistance is greater, its typical loss of fitness is lower, or, looking at things more from the parasite's perspective, we can say that the parasite is less virulent in this host. (We discuss the evolution of virulence, below, in Section 12.4, but note immediately that virulence is a property of neither the parasite nor the host alone, but rather of the interaction between the two.) A host, though, might also limit its loss of fitness, and the virulence of its parasites, by tolerating them – that is, responding in a way that minimises the harm they do without necessarily eliminating them. Thus, in Figure 12.9b we see two hosts differing in their tolerance. The more tolerant host suffers a much less severe loss of fitness as parasite burden increases, such that even though both typically carry similar parasite burdens, the more tolerant host is much less affected. Figure 12.9c then shows hosts differing in tolerance *and* resistance: the tolerant host carries the greater burden, but its loss of fitness is similar to its more resistant counterpart. And finally Figure 12.9d shows hosts differing in neither resistance nor tolerance, but with one having a higher intrinsic fitness than the other, even in the absence of parasites.


(a)
(b)
(c)
(d)
Host fitness
Host fitness
Parasite burden
Parasite burden
Median/typical burden of contrasting hosts
(e)
−0.1
−0.2
olerance
DBA/1
NIH
A/J

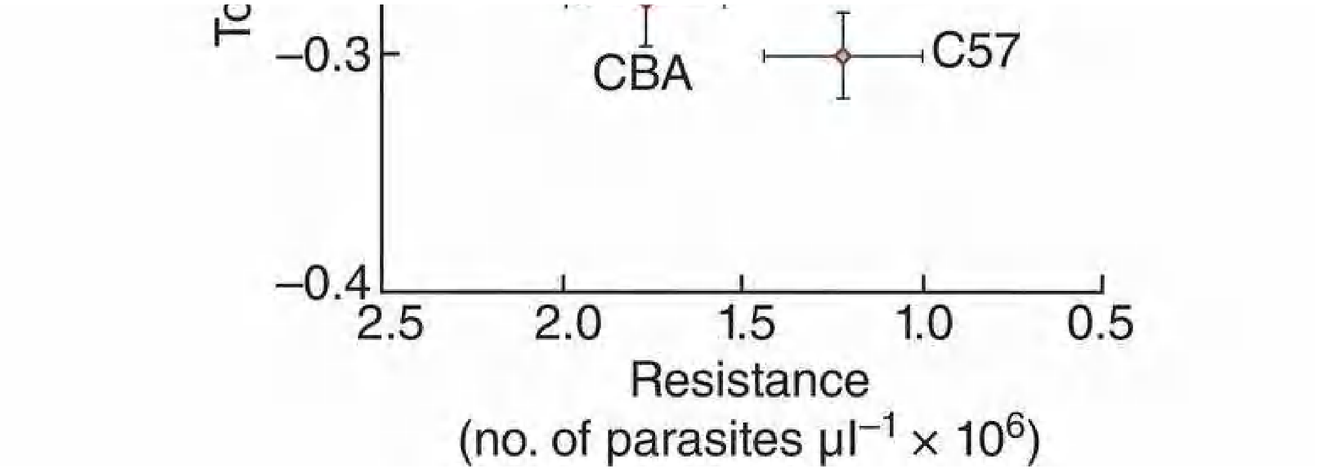


Figure 12.9 **Resistance, tolerance and virulence,** showing the effects of parasite burden on host fitness. The lengths of lines in (a) to (d) indicate the range of parasite burdens observed in a particular case, with typical burdens also indicated. Hosts with greater resistance are those that keep typical and maximum parasite burdens to lower levels. Hosts with greater tolerance are those with shallower slopes of declining host fitness as parasite burden increases. Virulence is the reduction in host fitness at typical parasite burdens compared with the absence of parasites, and hence reflects the interaction between parasite and host. (a) Two hosts differing in resistance, generating a difference in virulence. (b) Two hosts differing in tolerance, generating a difference in virulence. (c) Two hosts differing in resistance and tolerance, leading (in this case) to the same virulence. (d) Two hosts differing in intrinsic fitness but in neither resistance nor tolerance, leading to the same virulence. (e) The trade-off between tolerance (as measured by the slope of anaemia, minimum red blood cell concentration, against peak parasite density) and resistance (as measured by the inverse of peak parasite density) in five laboratory mouse strains, as indicated, when infected with the malaria parasite, *Plasmodium chabaudi*. SEs are shown for each strain.

Source: Adapted from Raberg *et al.* (2007).

The idea that tolerance may be as important as resistance has been longer established among plant pathologists than among animal parasitologists, but acknowledgment is now widespread. Nor should it be imagined that tolerance means passive indifference. A host may need to invest as much in being tolerant as in being resistant, for example in upregulating their wound healing or producing chemicals that reduce the harm parasites do rather than eliminating them. Thus, for example, we see in Figure 12.9e another example of a trade-off within the immune system – in this case, between the tolerance and resistance of different laboratory mouse strains to the malaria parasite, *Plasmodium chabaudi*.

Note, too, that since a tolerant host supports a larger population of parasites within its body, possibly for longer, tolerance may promote rather than inhibit the onward transmission of the parasite, and ultimately its prevalence within the population. This seems indeed to have been the case for the buffalo and BTB in Figure 12.8.

A rare example illustrating all these points is shown in Figure 12.10. In forest patches in Santa Fe province, Argentina, larvae of the parasitic fly, *Philornis torquans*, attack nestlings of many species of birds but attack three most frequently: the great kiskadee, *Pitangus sulphuratus*, which is the preferred host of the fly and in which prevalence is highest, the greater thornbird, *Phacellodomus ruber*, and the little thornbird, *Ph. sibilatrix*. In a series of observations and experiments, the great kiskadee was the most tolerant of the three host species (Figure 12.10a). It suffered no reduction in nestling survival as a result of fly infestation, which often reached high intensities, and it made little or no investment in resisting the infestations (Figure 12.10b). Thus, virulence in this host was low (no reduction in survival and little reduction in growth) but parasite productivity was high (10.2 larvae completing development per host originally available).

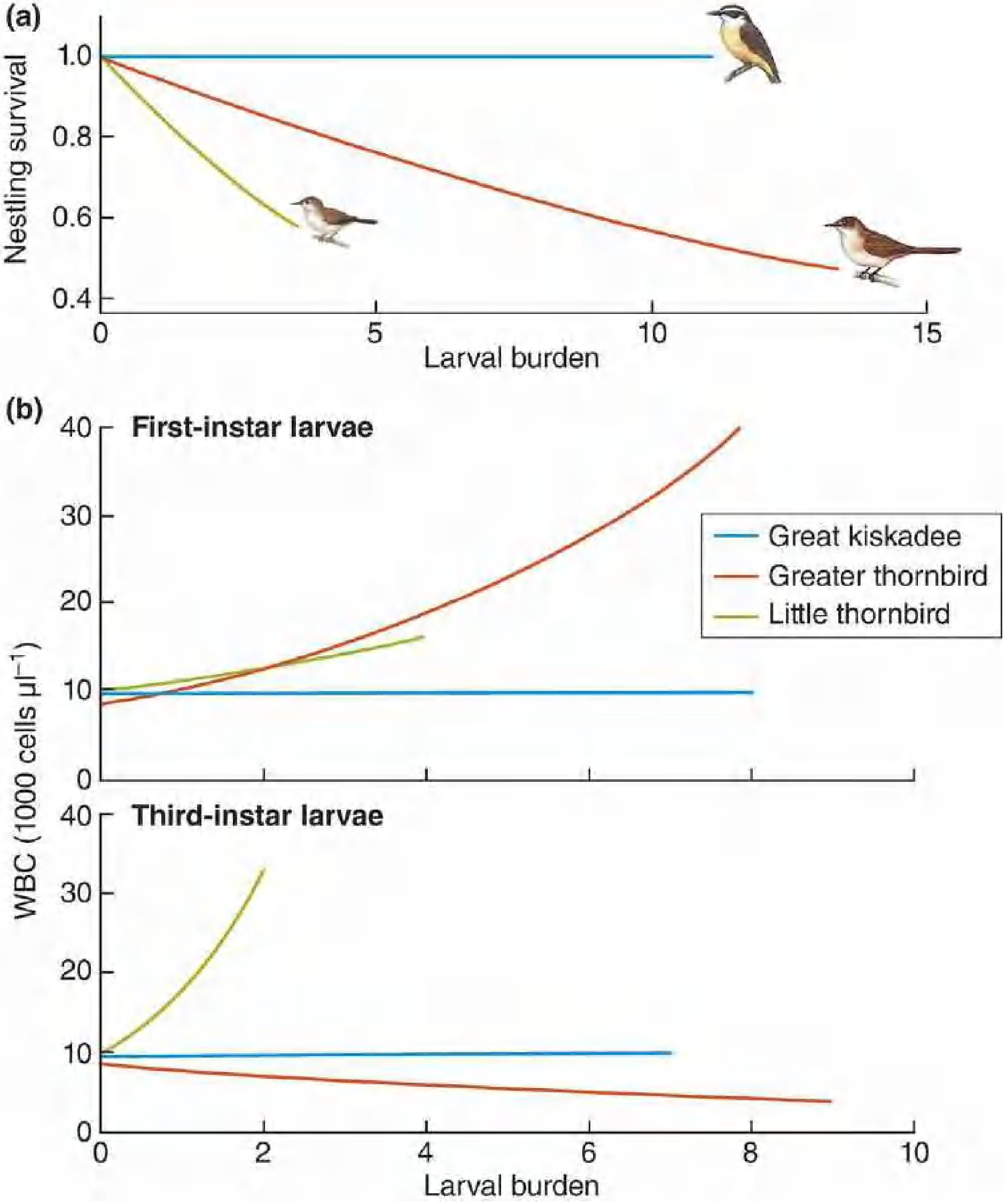

(a)
Nestling survival
1.0
0.8
0.6
0.4
0
5
10
15
Larval burden
(b)
First-instar larvae
40
30
20
10
0
Great kiskadee
Greater thornbird
Little thornbird
Third-instar larvae
40
30
20
10
0
WBC (1000 cells μl^{-1})
0
2
4
6
8
10
Larval burden

Figure 12.10 Argentinian bird host species show contrasting combinations of tolerance and resistance to infestations of parasitic fly larvae, with contrasting consequences for the hosts and parasites. (a) Nestling survival for the great kiskadee, greater thornbird and little thornbird (relative to uninfected hosts) in relation to levels of infestation with larvae of the parasitic fly *Philornis torquans*. The lines represent the results of statistical analyses (a mixed Cox proportional hazard model) that were adjusted for effects of brood size on survival. The length of each line extends to the third quartile of the burden recorded for each species and hence reflects typical levels of infestation in each case. (b) The investment of each species in resistance (an inflammatory response) in relation to levels of infestation, as measured by the concentrations of white blood cells (WBC) in the blood for first-instar and third-instar larvae, as indicated. The lines represent the results of statistical analyses (a generalised linear mixed model) that were adjusted for effects of age on survival. The length of each line extends to the third quartile of the burden recorded for each species and hence reflects typical levels of infestation in each case.

Source: After Manzoli *et al*. (2018).

By contrast, tolerance in the little thornbird was very low – it suffered rapid reductions in nestling survival as larval intensity increased (Figure 12.10a) – but it made a large investment in resisting infestation, which it sustained even as the larvae developed (Figure 12.10b). As a result, although virulence of the parasite was higher here than in the great kiskadee (a 42% reduction in survival and also small reductions in growth), intense infestations were uncommon (its resistance was effective) and parasite productivity was therefore low (0.7 larvae per host). By contrast again, tolerance in the greater thornbird was also low, though higher than in the little thornbird (Figure 12.10a), but its investment in resistance, while considerable initially, was only short-lived and ineffective (Figure 12.10b). As a result, relatively intense infestations were observed, and the virulence of the fly overall was highest here of the three hosts (a 46% reduction in survival and also moderate reductions in growth). Parasite productivity (1.3 larvae per host) was thus somewhat higher than in the little thornbird. Overall, therefore, this example illustrates not only the contrasting combinations of tolerance and resistance that hosts can exhibit, but also the consequences of these for both the hosts and the parasites,

12.3.7 Competition among parasites for host resources

constant final yield?

Since hosts are the habitat patches for their parasites, and since even within hosts, most parasites specialise on particular organs or tissues, it is not surprising that intra- and interspecific competition, observed in other species in other habitats, can also be observed in parasites within their hosts. There are many examples of the fitness of individual parasites decreasing within a host with increasing overall parasite abundance (Figure 12.11a), and of the overall output of parasites from a host reaching a saturation level (Figure 12.11b) reminiscent of the ‘constant final yield’ found in many plant monocultures subject to intraspecific competition (see Section 5.2.2).

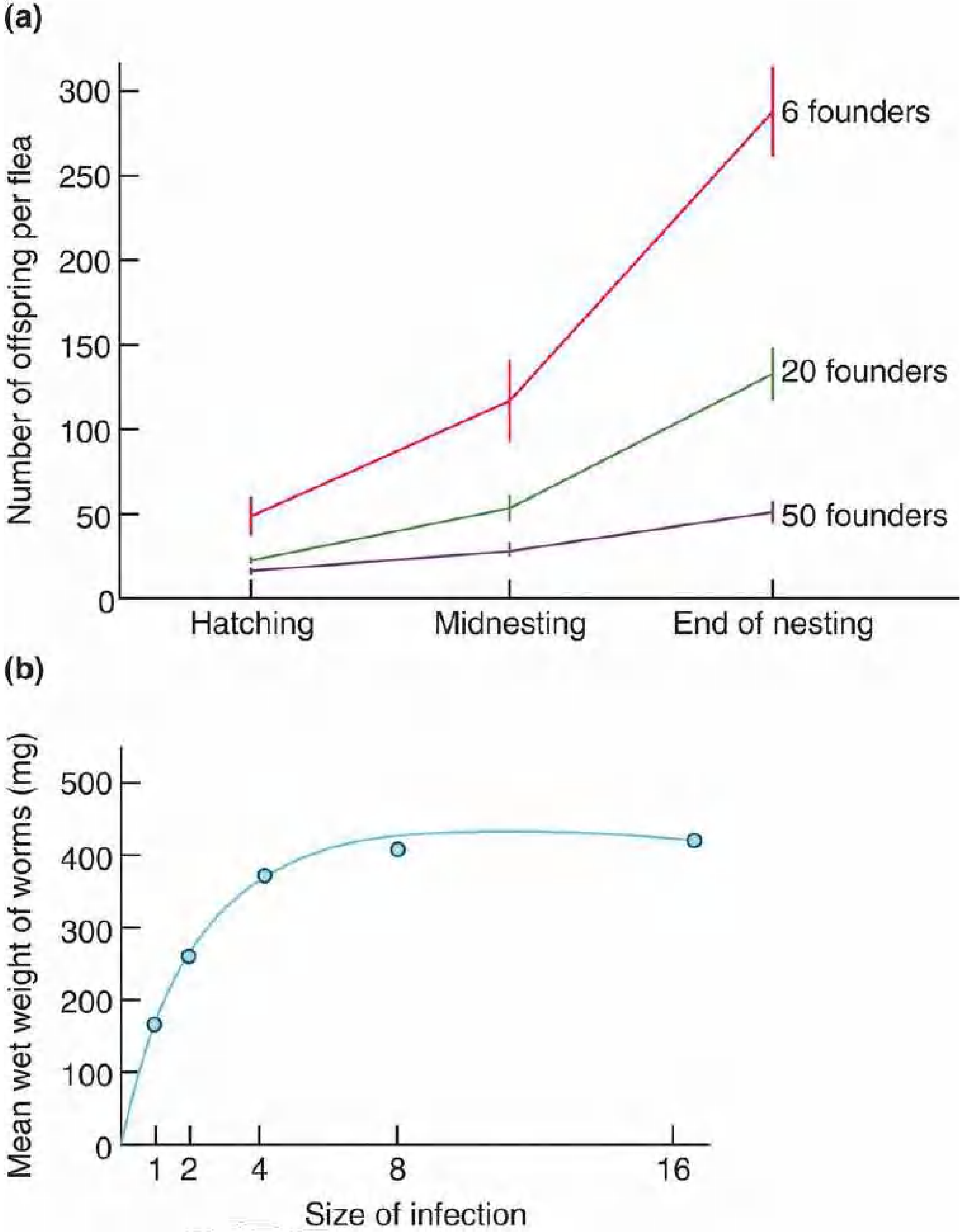

(a)
Number of offspring per flea
300
250
200
150
100
50
0
6 founders
20 founders
50 founders
Hatching
Midnesting
End of nesting
(b)
Mean wet weight of worms (mg)
500
400
300
200
100
0
1
2
4
8
16
Size of infection

Figure 12.11 Density-dependent responses of parasites within their hosts. (a) The relationship between the number of fleas *Ceratophyllus gallinae* ('founders') added to the nests of blue tits and the number of offspring per flea (mean ± SE). The greater the density, the lower the reproductive rate of the fleas. (b) The mean weight of worms per infected mouse reaches a 'constant final yield' after deliberate infection at a range of levels with the tapeworm *Hymenolepis microstoma*.

Source: (a) After Tripet & Richner (1999). (b) After Moss (1971).

competition or the immune response?

However, in vertebrates at least, we need to be cautious in interpreting such results simply as a consequence of intraspecific competition for limited resources, since the intensity of the immune reaction elicited from a host itself typically depends on the abundance of parasites. A rare attempt to disentangle these two effects utilised the availability of mutant rats lacking an effective immune response (Paterson & Viney, 2002). These and normal, control rats were subjected to experimental infection with a nematode, *Strongyloides ratti*, at a range of doses. Any reduction in parasite fitness with dose in the normal rats could be due to intraspecific competition and/or an immune response that itself increases with dose; but clearly, in the mutant rats only the first of these is possible. In fact, there was no observable response in the mutant rats (Figure 12.12), indicating that at these doses, which were themselves similar to those observed naturally, there was no evidence of intraspecific competition, and that the pattern observed in the normal rats is entirely the result of a density-dependent immune response. Of course, this does not mean that there is never intraspecific competition amongst parasites within hosts, but it does emphasise the particular subtleties that arise when an organism's habitat is its reactive host.

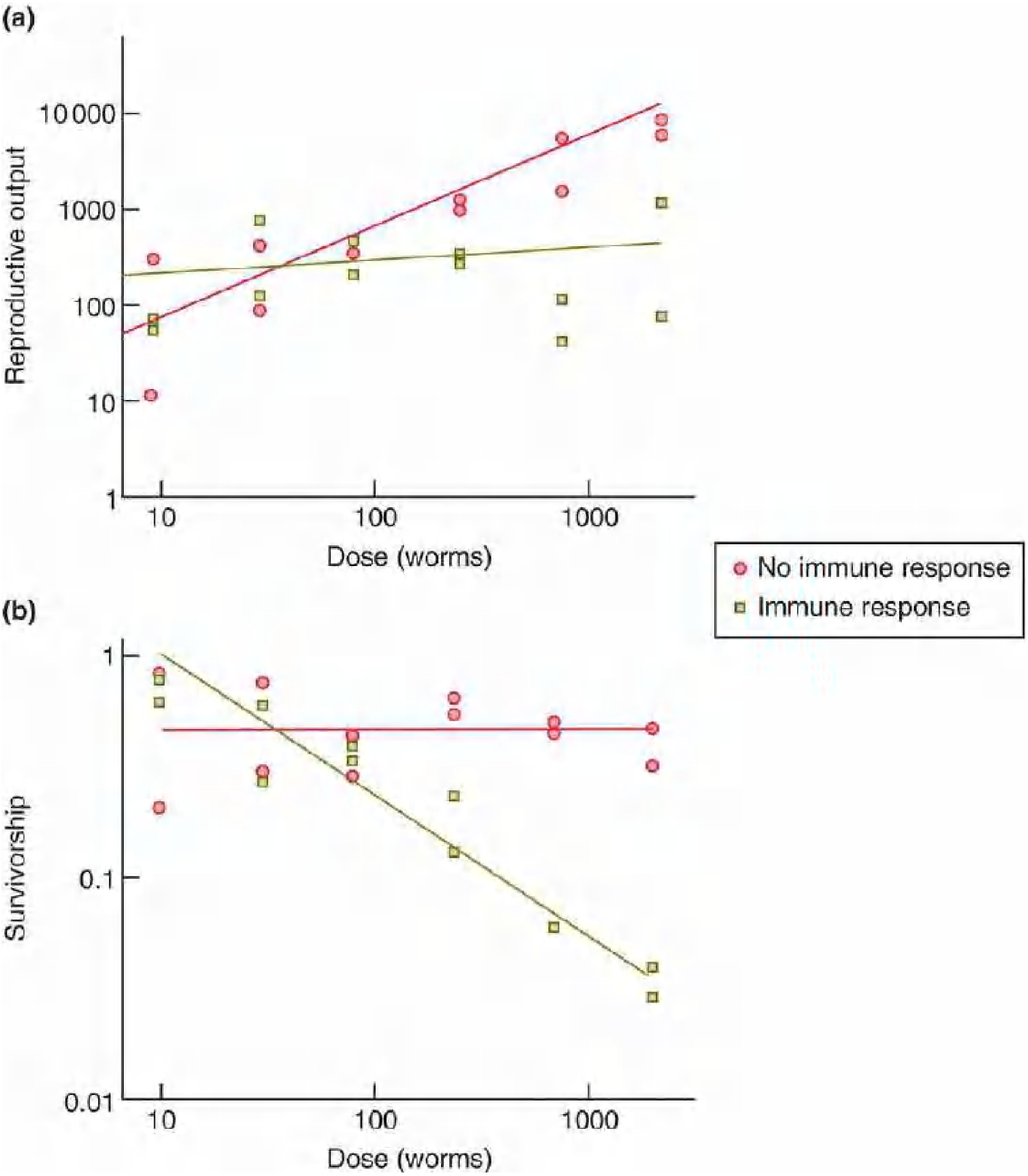


Figure 12.12 Host immune responses are necessary for density dependence in infections of the rat with the nematode *Strongyloides ratti*. (a) Overall reproductive output of the parasite increases in line with the initial dose in mutant rats without an immune response (slope not significantly different from 1), but with an immune response it is roughly independent of initial dose, i.e. it is regulated (slope = 0.15, significantly less than 1, $P < 0.001$). (b) Parasite survivorship is independent of the initial dose in mutant rats without an immune response (slope not significantly different from 0), but with an immune response it declines (slope = −0.62, significantly less than 0, $P < 0.001$).

Source: After Paterson & Viney (2002).

interspecific competition amongst parasites for hosts

We know from Chapter 8 that niche differentiation, and especially species having more effect on their own populations than on those of potential competitors, lies at the heart of our understanding of competitor coexistence. Clearly, parasite species can compete with one another – and show niche differentiation – at two levels: they may compete for, and specialise on, different host species or different resources (tissues, organs) within a host.

An example of the former comes from two species of parasitic nematode worms, *Howardula aoronymphium* and *Parasitylenchus nearcticus*, that infect the fruit-fly *Drosophila recens* (Perlman & Jaenike, 2001). Of these, *P. nearcticus* is a specialist, being found only in *D. recens*, whereas *H. aoronymphium* is more of a generalist, capable of infecting a range of *Drosophila* species. *P. nearcticus* has the more profound effect on its host, typically sterilising females, whereas *H. aoronymphium* seems to reduce host fecundity by only around 25% (though this itself represents a drastic reduction in host fitness). It is also apparent that whereas *H. aoronymphium* is profoundly affected by *P. nearcticus* when the two coexist within the same host in experimental infections (Figure 12.13a), this effect is not reciprocated (Figure 12.13b). Overall, therefore, competition is strongly asymmetric between the two parasites (as interspecific competition frequently is): the specialist *P. nearcticus* is both a more powerful exploiter of its host (reducing it to lower densities through its effect on fecundity) and stronger in interference competition. Coexistence between the species is possible, however, because the fly host provides the whole of the fundamental and the realised niche of *P. nearcticus*, whereas it is only part of the realised niche of *H. aoronymphium*.

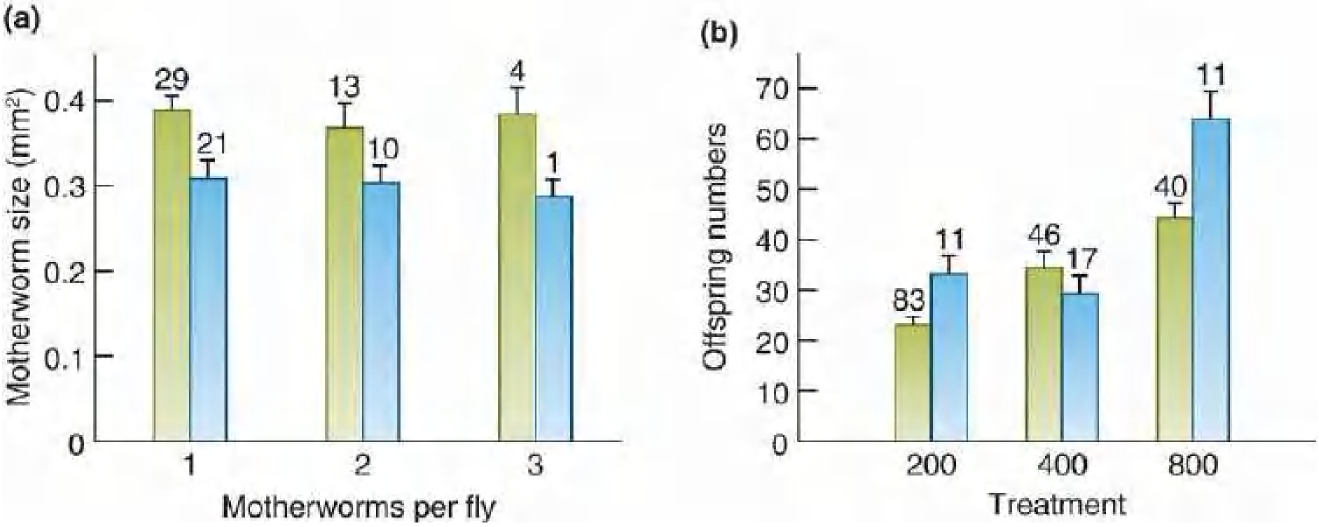


Figure 12.13 Competition between two worm species for a fruit-fly host. (a) Mean size ± SE (mm^2, longitudinal section area) of *Howardula aoronymphium* motherworms in one-week-old hosts, *Drosophila recens*, in single and mixed infections. Size is a good index of fecundity in *H. aoronymphium*. The hosts contained either one, two or three *H. aoronymphium* motherworms, having been reared on a diet contaminated with either *H. aoronymphium* (green bars) or mixed infections (*H. aoronymphium* and *Parasitylenchus nearcticus* – blue bars). Size (fecundity) was consistently lower in mixed infections. (b) Number of *P. nearcticus* offspring (i.e. fecundity) ± SE, in single (green bars) and mixed (blue bars) infections. Numbers above the bars indicate sample sizes of flies; treatment numbers refer to the numbers of nematodes added to the diet. Fecundity was not reduced in mixed infections.

Source: After Perlman & Jaenike (2001).

interspecific competition amongst parasites for host resources

We can get a broad view of the consequences of parasites competing for resources within hosts from a meta-analysis of studies in which laboratory mice were simultaneously infected – 'coinfected' – with helminth worms and microparasites (Figure 12.14). For some pairs of species, the helminths induce anaemia and the microparasite requires red blood cells for its reproduction. In these cases, the species are likely to compete for red blood cells, and it is therefore notable that helminth coinfection reduced the density of the competitor microparasites. But where there was no such potential for competition between the species, helminth coinfection actually led to an *in*crease in microparasite density on average, probably due to a decreased ability of coinfected hosts to resist the microparasite.

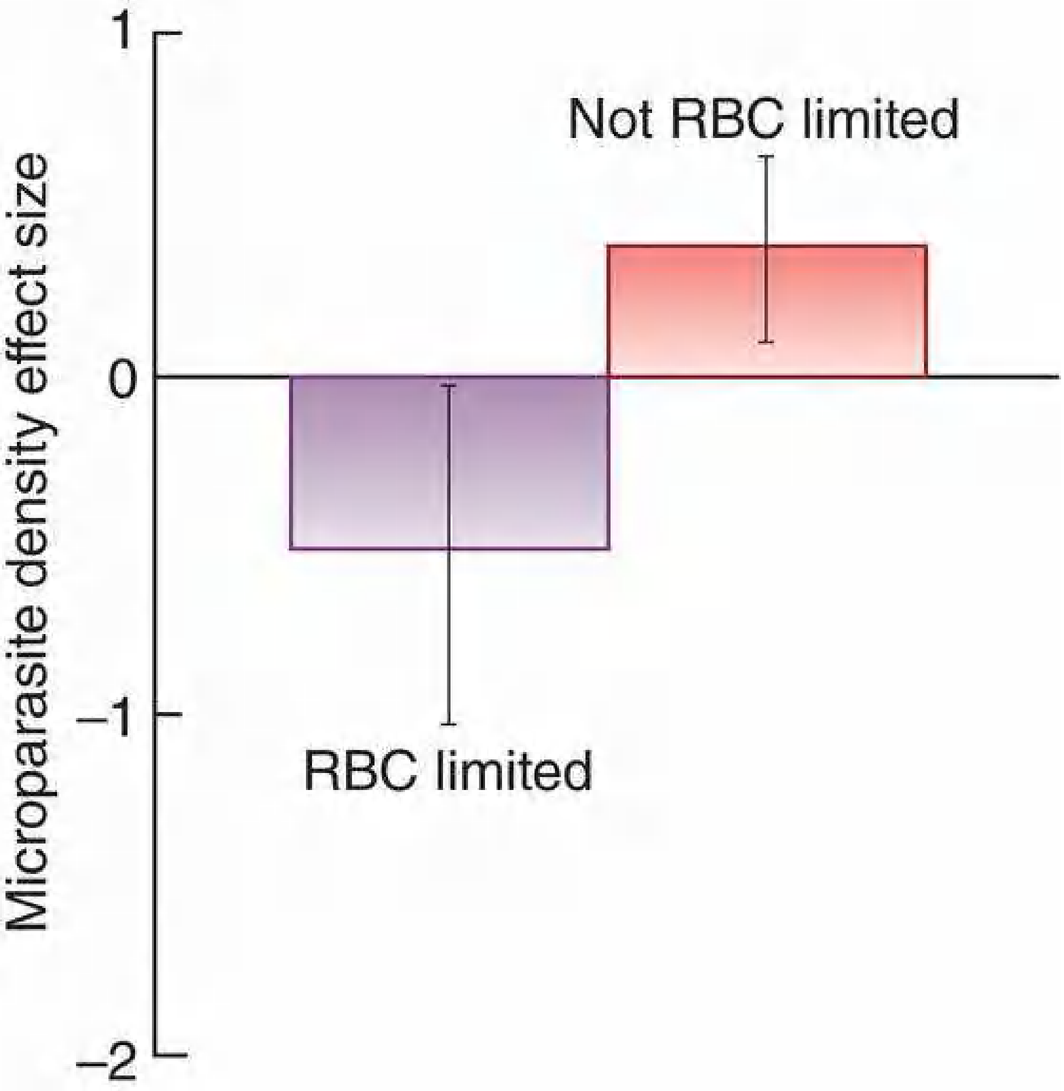


Figure 12.14 Effects on microparasites of helminth coinfection in mice. The mean effect size on microparasite density in the blood was examined from a meta-analysis of 54 studies of laboratory mice in which they were coinfected with helminths. The studies have been divided into those that are likely to have competed for red blood cells (RBCs) as a resource – where microparasite density tended to be driven down by coinfection – and those in which there was no such likelihood of competition, where microparasite density tended to be higher in coinfected hosts. The effect sizes are the log-transformed differences between densities in singly infected and coinfected hosts. Bars are 95% CIs.

Source: After Graham (2008).

12.3.8 The power of coinfection

In fact, although parasite ecologists have for practical reasons typically studied one host–one parasite interactions, coinfection is undoubtedly the norm in natural environments, and hence something that it is important for us to understand. As we saw in Figure 12.5, coinfecting parasites may be simultaneously competing for host resources and interacting via the host's immune system, and it will usually be difficult to disentangle any separate effects. The meta-analysis in Figure 12.14 was successful in doing this, because it was possible to separate species-pairs that were likely or unlikely to compete for an identifiable resource, red blood cells. But this is not usually possible. Whatever the mechanism of interaction, though, it is important that we understand the net effect that coinfecting parasites can have on one another, and we can see an example of this in Figure 12.15. Interactions were studied among four coinfecting microparasites in a natural population of field voles, *Microtus agrestis*, in the UK: cowpox virus, the bacteria *Anaplasma* and *Bartonella*, and the protozoan *Babesia microti*. The voles' infection status was checked each month, and the question being asked of each of the possible pair-wise interactions was: 'How does being infected with the first parasite *this* month affect the probability of becoming infected with the second parasite by the next month?' Figure 12.15 shows that these effects were significant – in some cases positive (the first infection increased the probability of acquiring the second), and in others negative. Indeed, they were not only significant statistically but also biologically. Even for a parasite with a marked seasonal cycle in its prevalence, like *Bartonella*, the probability of acquiring infection was affected more by its coinfecting parasites (between 15% and 65%; Figure 12.15b) than it was by which season of the year it was (between 15% and 55%; Telfer *et al.*, 2010). Clearly, if we wish to gauge a host's chances of becoming infected with any particular parasite, we need to know what else it is infected with.

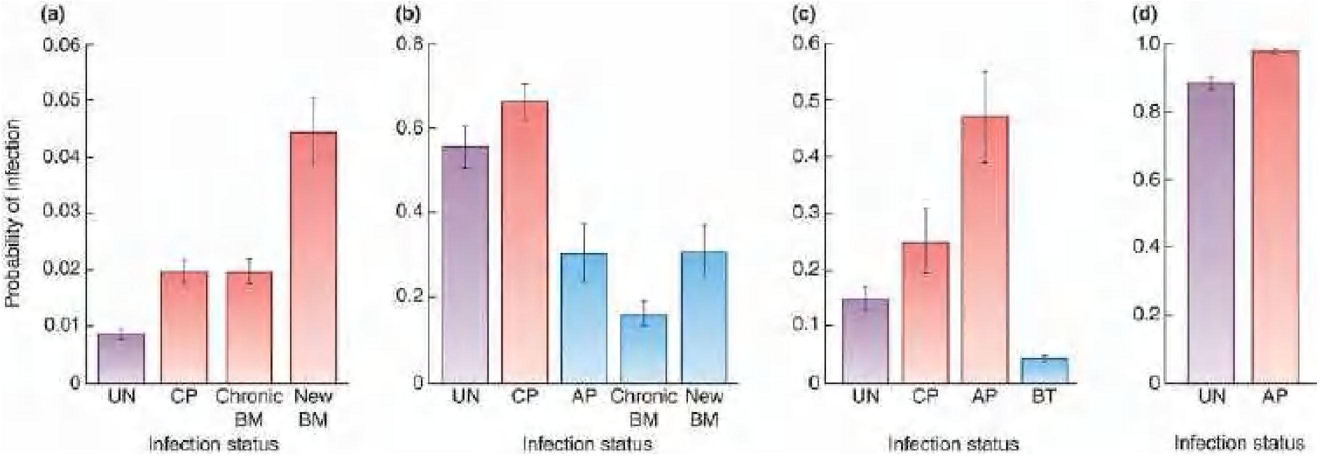


Figure 12.15 Positive and negative effects of coinfection in field voles. The probability of field voles, *Microtus agrestis,* acquiring an infection with four microparasites, depending on their infection status with those same microparasites. The four parts show field voles already infected with (a) *Anaplasma phagocytophilum* (AP), (b) *Bartonella* spp. (BT), (c) *Babesia microti* (BM), and (d) cowpox virus (CP). BM infections may have been acquired either before the preceding month (chronic) or newly acquired that month. Only significant effects are shown, contrasted with voles not infected with any of these (UN). Coinfections increasing the probability are coloured red, those decreasing it are coloured blue. Bars are 95% CIs.

Source: After Telfer *et al.* (2010).

12.4 Coevolution of parasites and their hosts

We have seen earlier how the resistance and tolerance of hosts and the virulence of parasites combine to determine the outcome, in fitness terms, of host–parasite interactions. It may seem obvious that parasitic infection selects for the evolution of more resistant or tolerant hosts, and that increases in resistance drive the evolution of more infective parasites, and so on, in a coevolutionary arms race. In fact, the process is not necessarily so straightforward, although there are certainly examples where a host and parasite drive one another’s evolution.

APPLICATION 12.3 Myxomatosis

A classic, now almost ancient, example of host–parasite coevolution involves the rabbit and the myxoma virus, which causes myxomatosis. The virus originated in the South American jungle rabbit *Sylvilagus brasiliensis*, where it causes a mild disease that only rarely kills the host. The virus, however, is usually fatal when it infects the European rabbit *Oryctolagus cuniculus*. In one of the greatest examples of the biological control of a pest, the myxoma virus was introduced into Australia in the 1950s to control the European rabbit, which had become a pest of grazing lands. The disease spread rapidly in 1950–51, and rabbit populations were greatly reduced. At the same time, the virus was introduced to England and France, and there too it resulted in huge reductions in the rabbit populations. The evolutionary changes that then occurred in Australia were followed in detail by Fenner and his associates (Fenner & Ratcliffe, 1965 ; Fenner, 1983) who had the brilliant research foresight to establish baseline genetic strains of both the rabbits and the virus. They used these to measure subsequent changes in the virulence of the virus and the resistance of the host as they evolved in the field.

When the disease was first introduced to Australia it killed more than 99% of infected rabbits. This 'case mortality' fell to 90% within 1 year and then declined further (Fenner & Ratcliffe, 1965). The virulence of isolates of the virus sampled from the field was graded according to the survival time and the case mortality of control rabbits. The original, highly virulent virus (1950–51) was grade I, which killed 99% of infected laboratory rabbits, but even by 1952, most of the virus isolates from the field were the less virulent grade III (Figure 12.16). At the same time, the rabbit population in the field was increasing in resistance. When injected with a standard grade III strain of the virus, field samples of rabbits in 1950–51 had a case mortality of nearly 90%. But this had declined to less than 30% only eight years later (Marshall & Douglas, 1961).

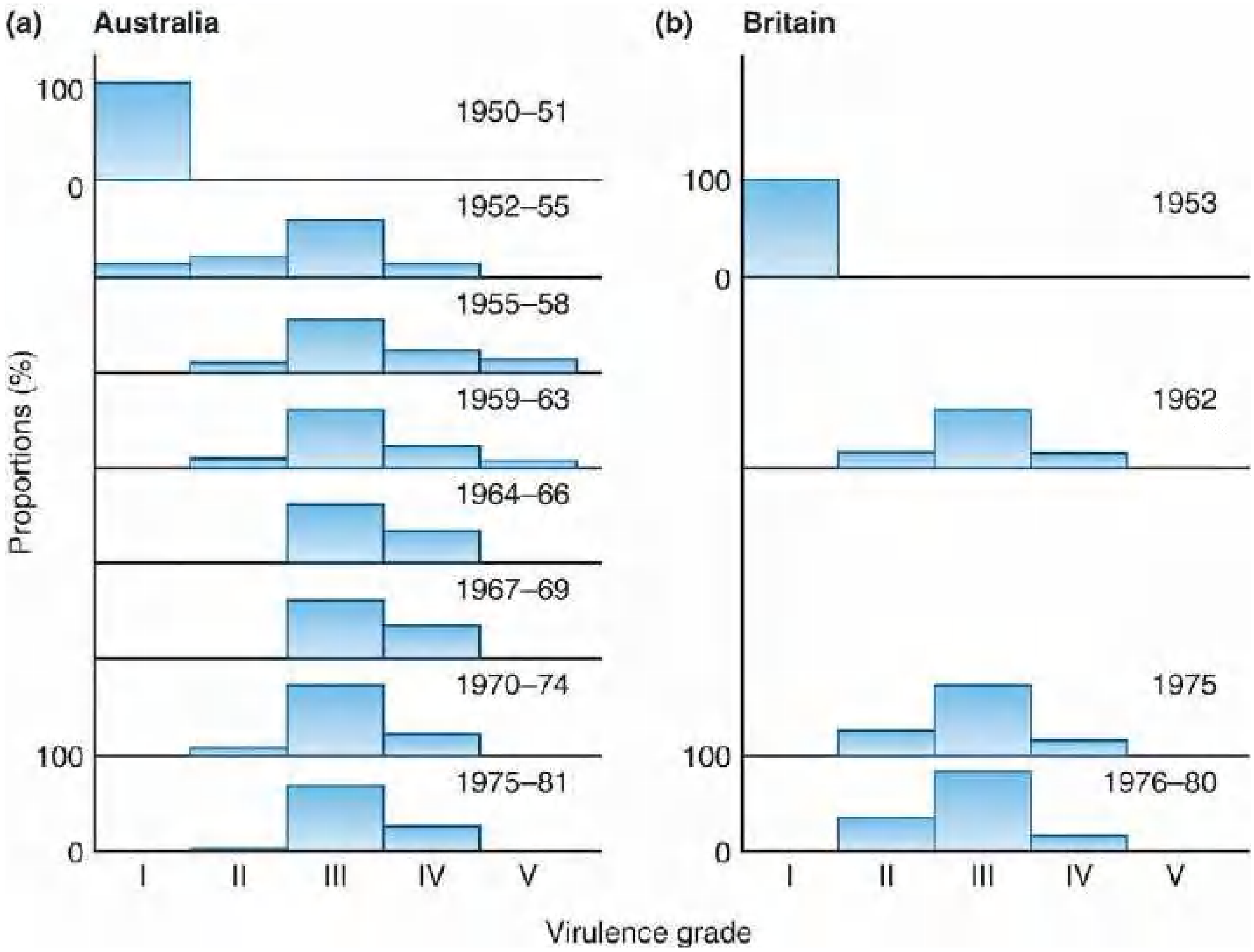


Figure 12.16 The myxoma virus in European rabbits evolved from high to intermediate virulence. (a) The percentages in which various grades of myxoma virus have been found in wild populations of rabbits in Australia at different times from 1950 to 1981. Grade I is the most virulent. (b) Similar data for wild populations of rabbits in Great Britain from 1953 to 1980.

Source: (a) After Fenner (1983). (b) After May & Anderson (1983), from Fenner (1983).

Relatively resistant European rabbits were obviously favoured by natural selection. The case of the virus, however, is subtler. The lack of virulence of the myxoma virus in the American host with which it had coevolved, combined with the attenuation of its virulence in Australia and Europe after its introduction, fit a commonly held view that parasites evolve toward becoming benign to their hosts in order to prevent the elimination of their own habitat. This is wrong. The parasites favoured by natural selection are those with the greatest fitness (broadly, the greatest value of R_0 – see Section 12.7). There are various ways this can occur.

In the rabbit–myxomatosis system, the myxoma virus is blood-borne and is transmitted from host to host by blood-feeding insect vectors. In Australia in the first 20 years after its introduction, the main vectors were mosquitoes (especially *Anopheles annulipes*), which feed only on live hosts. The problem for the virulent grade I and II viruses is that, while they produce many infective particles, they kill the host so quickly that there is only a very short time in which the mosquito can transmit them. Hence, there was selection against grades I and II and in favour of less virulent grades, giving rise to longer periods of host infectiousness. At the other end of the virulence scale, however, the mosquitoes are unlikely to transmit grade V virus because it produces very few infective particles in the host skin that could contaminate the vectors' mouthparts. An initial decline to intermediate grades of virulence was therefore favoured, but further declines were not.

the evolution of virulence

We see in Application 12.3, therefore, an example of a trade-off between the length of the infectious period (shorter for more virulent strains) and the rate at which infectious viral particles are shed (greater for more virulent strains). Increases in either would lead to an increase in parasite fitness (see Section 12.7), but since increases in one carry the cost of decreases in the other, the intermediate virulence that has been favoured is an optimal combination of the two, maximising virus fitness. This is usually described as a transmission–virulence trade-off: increases in virulence (increasing the rate of virus production but shortening the infectious period) lead to decreases in the transmission rate (see Cressler *et al.* (2016) for a review of the evolution of virulence). Many parasites of insects, on the other hand, rely on killing their host for effective transmission. In these, very high virulence is favoured. In other cases, natural selection acting on parasites has clearly favoured very low virulence: for example, the human herpes simplex virus may do very little tangible harm to its host but effectively gives it lifelong infectiousness. These variations no doubt reflect differences in the underlying host–parasite ecologies (Cressler *et al.*, 2016). For example, parasites reliant on stages that persist long term in the environment are predicted to evolve high virulence (since associated reductions in host lifespan are less important), whereas parasites attacking hosts with poor powers of dispersal are predicted to evolve low virulence (because of the dangers of quickly 'using up' local, accessible susceptible hosts). In all cases, though, we can expect virulence to have been driven by evolution towards increased parasite fitness, which in turn coevolves with changes in host resistance.

antagonistic coevolution in plants

In other cases, coevolution is more definitely antagonistic, driving towards increased resistance in the host and increased infectivity in the parasite. This has been the perspective in the study of plants, especially agricultural plants, and their pathogens, where a variety of models for the underlying genetic basis – gene-for-gene, matching allele, inverse matching allele – have been proposed and found support (see Thrall *et al.*, 2016). This antagonistic coevolution may give rise to polymorphism in the parasite and host, either as a result of different genes or alleles being favoured in different subpopulations, or because several alleles are simultaneously in a state of flux within their population, each being favoured when they (and their matching allele in the other partner) are rare. It may give rise, too, to a correlation between the resistance of hosts in different populations and the virulence of their pathogens (Figure 12.17).

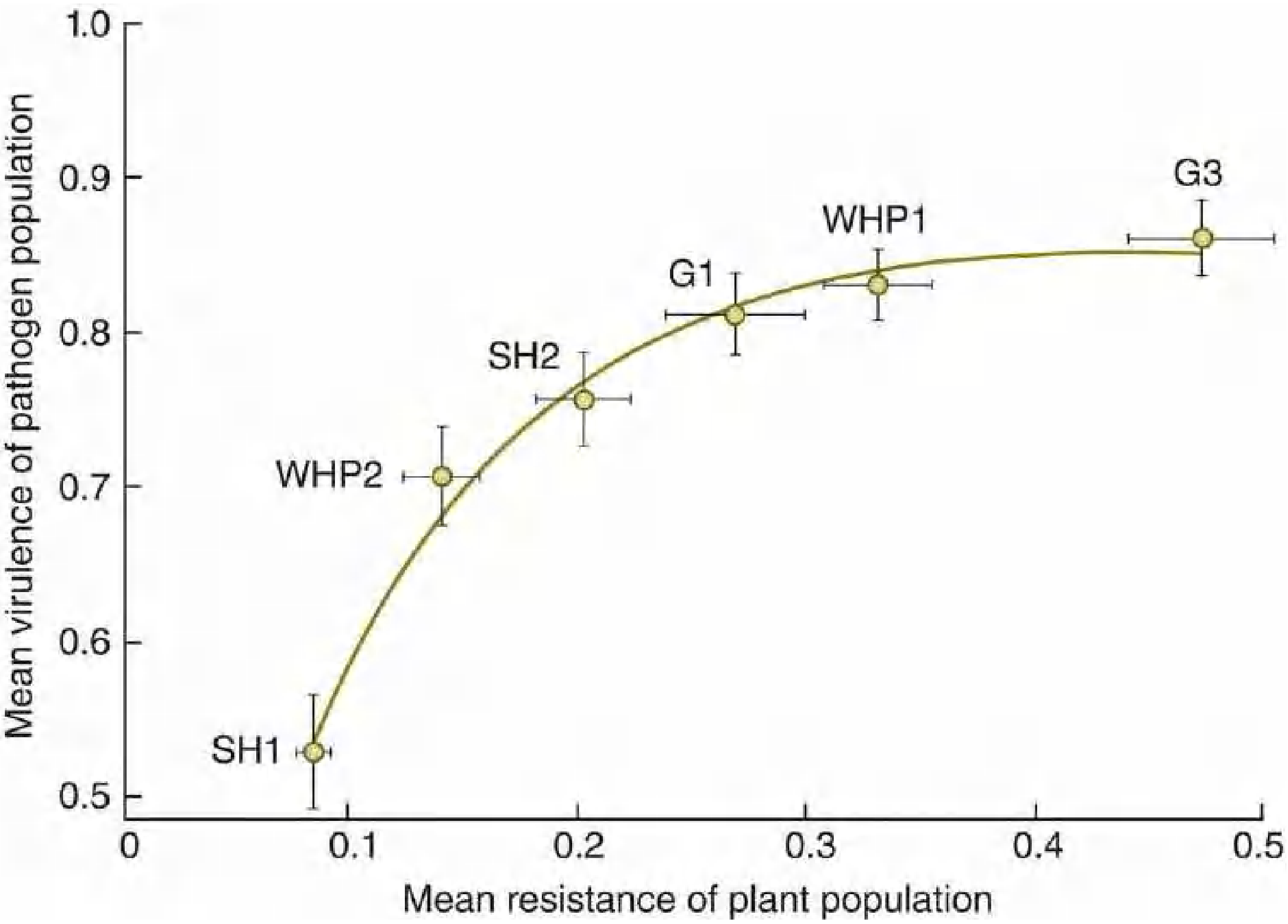


Figure 12.17 Coevolution leading to rust fungus pathogens being more virulent in more resistant populations of their host. The relationship between the mean virulence of the rust fungus pathogen, *Melampsora lini*, and the mean resistance of its host, the Australian flax plant, *Linum marginale*, are given in six populations in New South Wales, Australia, as indicated. Mean virulence was calculated as the fraction of hosts that could be successfully attacked; mean resistance was calculated as the fraction of pathogens against which resistance was observed. Bars are 95% CIs. The fitted line is the best-fit second-order power function.

Source: After Thrall & Burdon (2003).

bacteria and bacteriophages

The detailed coevolutionary processes leading to these patterns are difficult to observe, but an increasing number of studies have utilised microorganisms in studies of 'experimental evolution' (see also Section 8.10.3), where generation times are short enough for us to be able to see evolution in action. Evolved strains of both host and parasite can be tested not only against their contemporaries but against strains from their past. We can see not only where they have got to, but how far they have come. An example is shown in Figure 12.18, tracing the coevolution of the host bacterium *Pseudomonas fluorescens* and its viral parasite, the bacteriophage (or phage) SBW25Φ2, in laboratory culture bottles containing 6 ml of growth medium. The host and parasite were grown together in replicate populations and, every second day, 1% of the population was transferred to a fresh 6 ml of medium. In addition, and crucially, the residual populations at each transfer were retained and frozen, so that they could be tested against hosts and parasites from both earlier and later in the coevolutionary sequence. There were 16 transfers in all – around 120 bacterial generations. Then, every two transfers, the resistance of the bacteria (or, from the alternative point of view, the infectivity of the phage) was tested against their phage antagonist from the contemporary population and from populations two transfers previously and two transfers subsequently. Over (evolutionary) time the bacteria became increasingly resistant to the phage: the triplets of previous, contemporary and subsequent phage moved higher up the resistance scale in Figure 12.18. But the phage also became more virulent: within each triplet, the bacteria were more resistant to phage from two transfers previously (when phage virulence had not yet evolved to its contemporary level) but less resistant to phage two transfers into the future (by which time the phage virulence had evolved further). Antagonistic coevolution was confirmed.

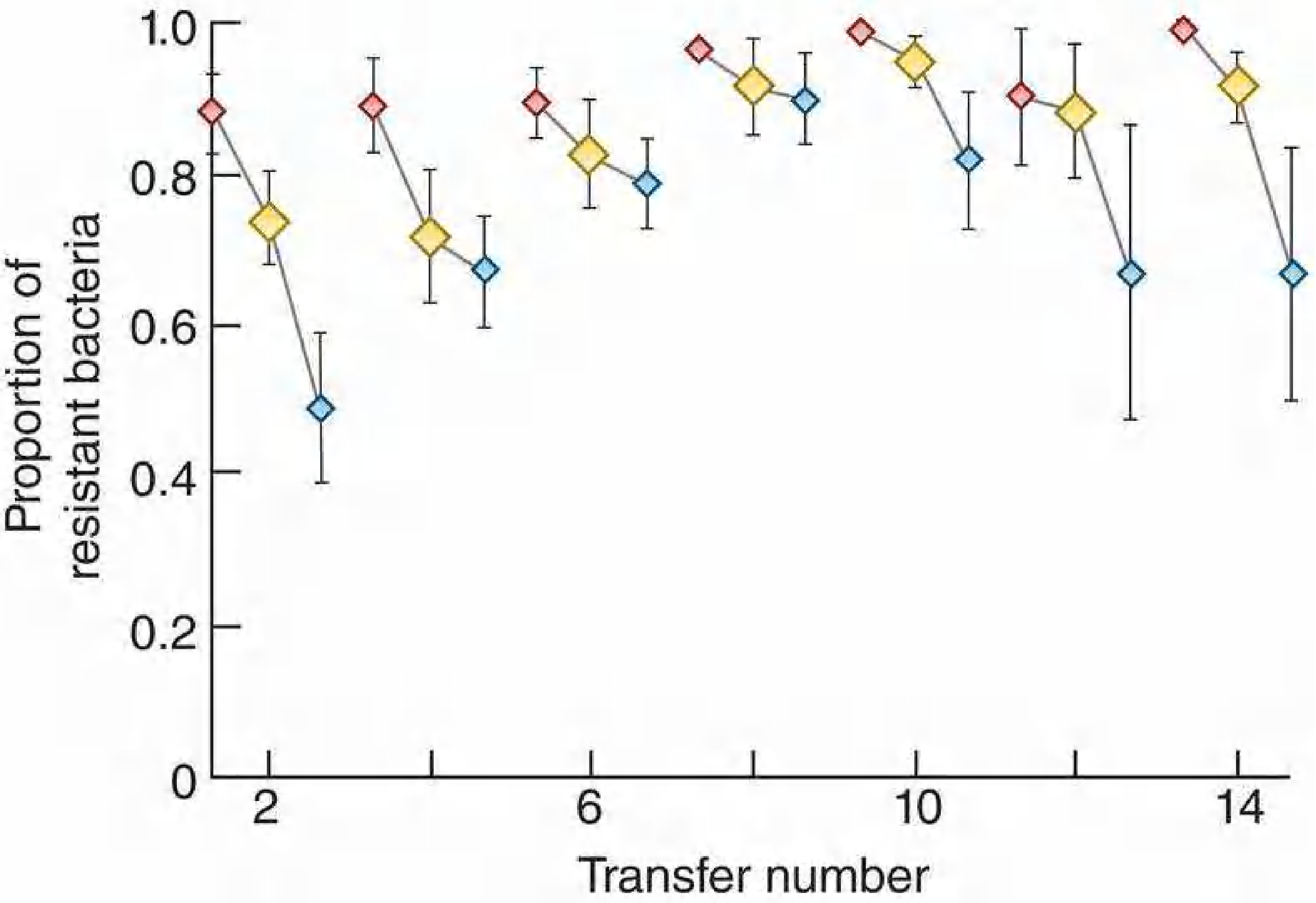


Figure 12.18 Antagonistic coevolution of the bacterial host, *Pseudomonas fluorescens*, and its parasite, the phage SBW25Φ2. Each triplet of points shows the results of bacteria tested against contemporary phage (yellow) and phage two transfers earlier (pink) and later (blue). The increase in resistance of the triplets over time indicates the evolving resistance of the host. The negative slope within each triplet additionally indicates the evolving virulence of the phage. Bars are 95% CIs.

Source: After Brockhurst *et al*. (2003).

12.5 The transmission of parasites amongst hosts

12.5.1 Transmission dynamics

direct and indirect transmission; short-and long-lived agents

Different species of parasite are, of course, transmitted in different ways between hosts. The most fundamental distinction, perhaps, is between parasites that are transmitted directly from host to host and those that require a vector or intermediate host for transmission. Amongst the former, we should also distinguish between those where infection is by physical contact between hosts or by a very short-lived infective agent (for example, coughs and sneezes), and those where hosts are infected by long-lived particles (e.g. dormant and persistent spores). Essentially the same patterns apply in plants. For example, many soil-borne fungal diseases are spread from one host plant to another by root contacts, or by the growth of the fungus through the soil from a base established on one plant, which gives it the resources from which to attack another. Many other plant diseases are spread by spores carried on the wind.

The dynamics of transmission are in a very real sense the driving force behind the overall population dynamics of pathogens, but they are often the aspect about which we have least data (compared, say, with the fecundity of parasites or the death rate of infected hosts). We can, nonetheless, build a picture of the principles behind transmission dynamics (Begon *et al*., 2002). The rate of production of new infections in a population depends on the *per capita* transmission rate (the rate of transmission per susceptible host 'target') and on the number of susceptible, target hosts (which we can call *S*). The per capita rate is usually proportional, first, to the contact rate between susceptible hosts and whatever it is that carries the infection. We refer to this as *k*. It is also proportional to the probability that a contact that *can* transmit infection actually does so, which clearly depends on the infectiousness of the parasite, the susceptibility of the host, and so on. We refer to this as *p*. Putting these three components together we can say:

$$\text{the rate of production of new infections} = k \cdot p \cdot S. \tag{12.1}$$

the contact rate

The nature of the contact rate is different for different types of transmission. For parasites transmitted directly from host to host, we deal with the rate of contact between infected hosts and susceptible (uninfected) hosts. For hosts infected by long-lived infective agents that have been shed into the environment, we deal with the rate of contact between these and susceptible hosts. And with vector-transmitted parasites, we deal with the contact rate between host and vector (the 'host-biting rate'), which goes to determine both the transmission rate from infected hosts to susceptible vectors and that from infected vectors to susceptible hosts.

But what is it that determines the per capita contact rate between susceptibles and infecteds? For long-lived infective agents, it is usually assumed that the contact rate is determined simply by their density – the more of them there are, the more likely it is that a susceptible host will contact one. For direct and vector-borne transmission, however, the contact rate needs to be broken down further into two components. The first is the contact rate between a susceptible host and all other hosts (or between a susceptible host and all other vectors). We refer to this as *c*. The second is then the proportion of those hosts or vectors that are infectious. We refer to this as I/N, where I is the number of infecteds and N the total number of hosts (or vectors). Our expanded equation is now:

$$\text{the rate of production of new infections} = c \cdot p \cdot S \cdot (I/N). \tag{12.2}$$

We need to look a little more deeply into *c* and I/N, in turn.

12.5.2 Contact rates: density- and frequency-dependent transmission

density-dependent transmission

For most infections, it is typically assumed that the contact rate, *c*, increases in proportion to the density of the population, N/A, where A is the area occupied by the population: the denser the population, the more hosts come into contact with one another (or vectors contact hosts). Assuming for simplicity that the area occupied, A, remains constant, the Ns in the equation then cancel, all the other constants can be combined into a single constant, β, the 'transmission coefficient', and the equation becomes:

$$\text{the rate of production of new infections} = \beta \cdot S \cdot I. \tag{12.3}$$

This, unsurprisingly, is known as *density-dependent transmission*.

frequency-dependent transmission

On the other hand, it has long been asserted that for sexually transmitted diseases, the contact rate is constant: the frequency of sexual contacts is independent of population density. This time the equation becomes:

$$\text{the rate of production of new infections} = \beta' \cdot S \cdot (I/N), \tag{12.4}$$

where the transmission coefficient again combines all the other constants, but this time acquires a 'prime', β', because the combination of constants is slightly different. This is known as *frequency-dependent transmission*.

Increasingly, however, it has become apparent that it is not simply that most transmission is density-dependent except for sexual transmission, which is frequency-dependent. For example, frequency dependence was found to be a better descriptor than density dependence for transmission of a directly transmitted infection in bank voles (*Myodes glareolus*; Begon *et al*., 1998); while density dependence was found to be a better descriptor than frequency dependence for transmission for a sexually transmitted infection in ladybirds (*Adalia bipunctata*; Ryder *et al*., 2005). One likely explanation is that sexual contact is not the only aspect of behaviour where the contact rate varies little with population density – many social contacts may come into the same category – while not all sexual contact may be intimate or long-lived, and they may therefore occur more often in denser populations.

ends of a spectrum

What is more, density and frequency dependence are themselves increasingly recognised not as simple truths but as benchmarks against which real examples of transmission might be measured, or perhaps as ends of a spectrum along which real transmission terms could be assembled. For example, the nature of transmission of the directly transmitted cowpox virus in natural populations of the field vole, *Microtus agrestis*, was explored by tracking the numbers of voles susceptible to, infected with and recovered from the virus (S, I and R, adding up to the total number of voles, N) and determining which transmission term best predicted the number of infected voles, I, in one month, based on S, I and N in the previous month. The transmission term was taken to be $\beta S(I/N^q)$, so that by estimating the most appropriate value of q it was possible to see if transmission was density-dependent (for which q would have been 0) or frequency-dependent ($q = 1$). In fact, the estimated value of q was 0.62 (with a confidence interval of 0.49–0.74), distinguishing it significantly from both (Figure 12.19). Rather than the rate of contacts between voles (appropriate for transmission) remaining the same as population size increases (frequency dependence), or increasing in direct proportion to population size (density dependence), it seems the rate increases rapidly with density in small populations but reaches a point in larger populations where increases in density have little effect on contact rate, perhaps because territorial behaviour limits contacts to near neighbours.

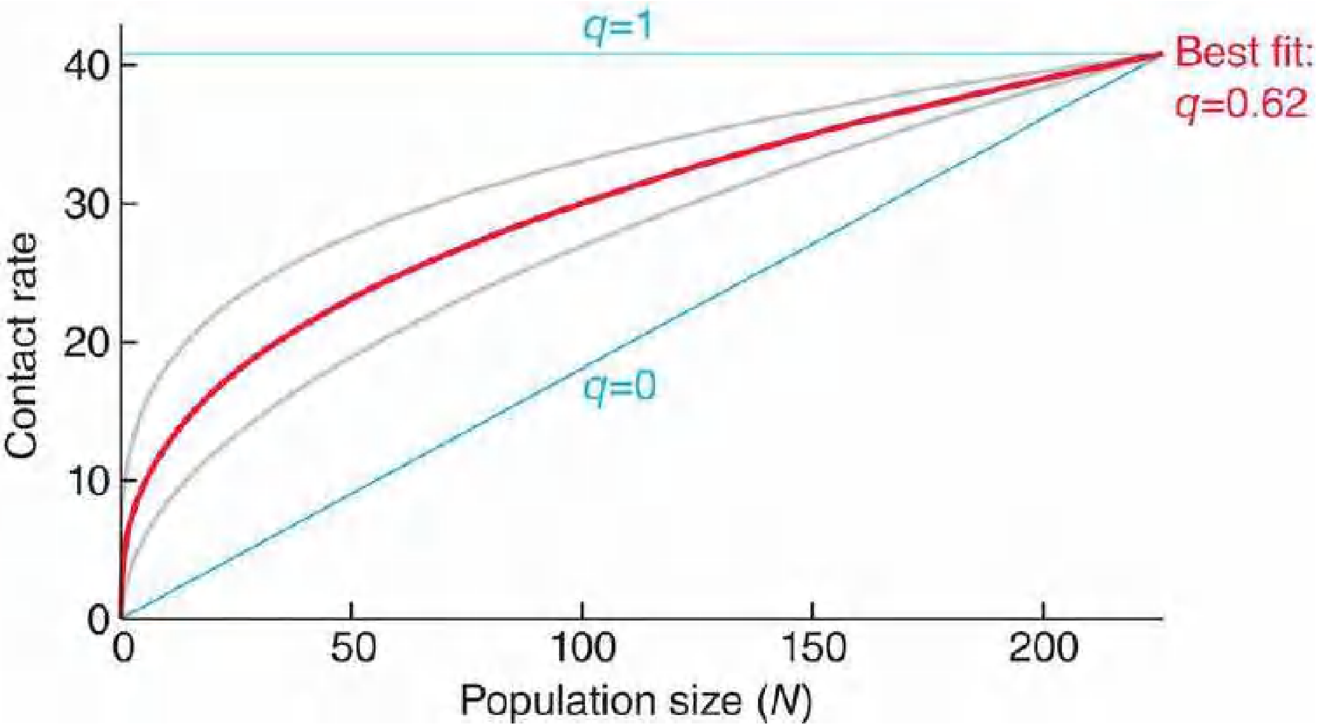
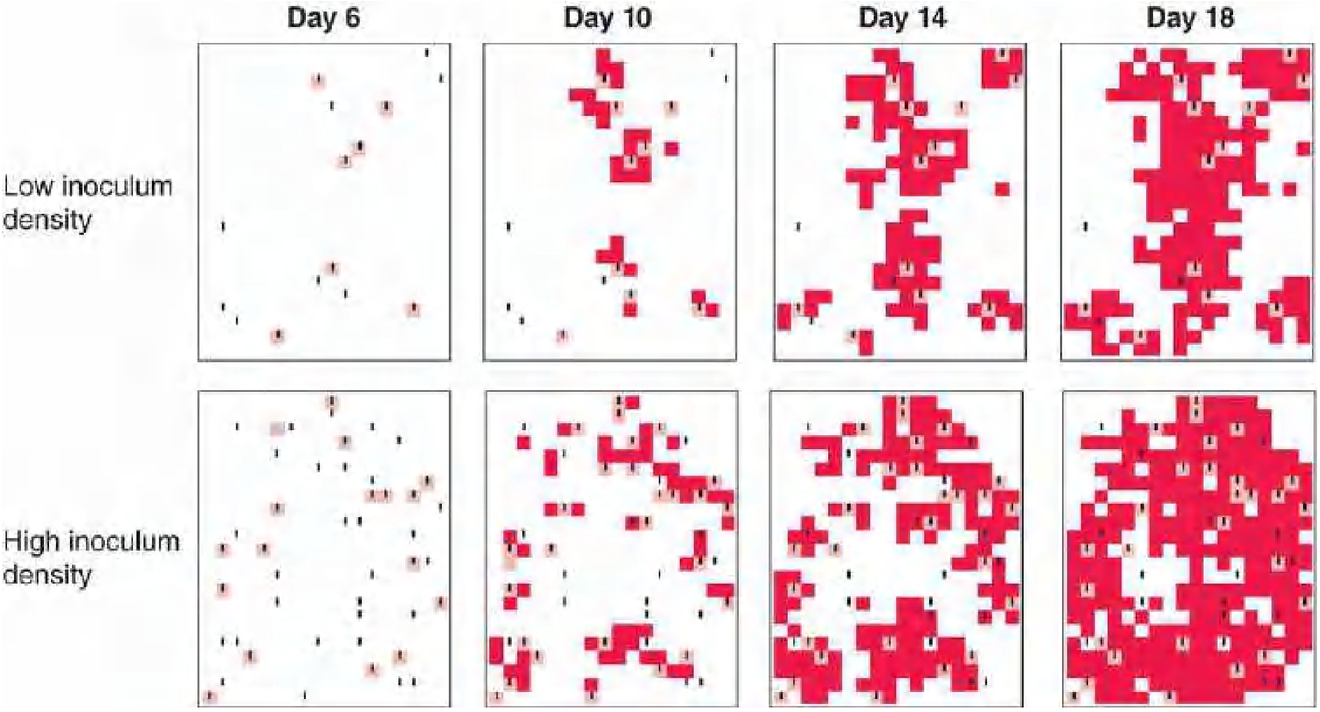


Figure 12.19 **Transmission of cowpox virus in field voles was neither density- nor frequency-dependent.** Data on the changing numbers of field voles, *Microtus agrestis*, susceptible to, infected with or recovered from infection with cowpox virus were used to estimate the best transmission term of the form $\beta S(I/N^q)$ to account for these (see text). The value of q in the term would indicate if transmission was density-dependent ($q = 0$) or frequency-dependent ($q = 1$), both shown as thin blue lines, but in fact the most appropriate value, 0.62 (bold red line; 95% CIs in grey) was significantly different from both of these.

Source: After Smith *et al*. (2009).

local hot spots

Turning to the I/N term, the usual simplifying assumption is that this can be based on numbers from the whole of a population. That is, the use of such a term assumes either that all individuals in a population are intermingling freely with one another, or, slightly more realistically, that individuals are distributed approximately evenly across the population, so that for all susceptibles the probability that a contact is with an infectious individual is I/N. In reality, however, transmission typically occurs locally, between nearby individuals, so there are likely to be hot spots of infection in a population, where I/N is high, and corresponding cool zones. Transmission, therefore, often gives rise to growing patches of infection in a population (e.g. Figure 12.20), rather than simply the overall rise in infection implied by a global transmission term like βSI. This, of course, is the price paid in diminished realism when a complex process is boiled down into a simple term (such as βSI). Nonetheless, as we shall see (and have seen previously in other contexts) without such simple terms to help us, progress in understanding complex processes would be impossible.



[Figure 12.20](#) **Local transmission of damping off disease in radishes leads to growing patches of infection.** The spatial spread over 18 days of damping off disease, caused by the fungus *Rhizoctonia solani*, in a population of radish plants, *Raphanus sativus*, in trays 550 × 350 mm overall, planted in a 23 × 18 rectangular grid. The fungus was inoculated at contrasting densities: 15 ('low') or 45 ('high') discs of fungus. Following initiation of the disease at randomly selected points (inoculation discs placed on plants – small vertical lines), plants either became infected from the initial inoculum (light squares), or as secondary infections when the epidemic spread to neighbouring plants (dark squares). Even at the experiment's end, the distribution of the disease clearly reflects local spread from the primary infections.

Source: After Otten *et al*. (2003).

APPLICATION 12.4 Superspreaders and their identification

Spatial hot spots of infection are an example of a more general truth – that, in terms of onward transmission, not all hosts are equal. A small minority of hosts may harbour more parasites than the rest (see Figure 12.2), or shed more infective particles, or contact more susceptible hosts. Any or all of these can make those hosts *superspreaders*, responsible for a very large proportion of new cases. The idea goes back at least as far as Mary Mallon, 'Typhoid Mary', who infected 51 other people with typhoid fever between 1902 and 1909, working in the food service in New York City while being an asymptomatic carrier, before being involuntarily quarantined (locked up!). We get some sense of the generality of this pattern from Figure 12.21, where the observed numbers of secondary cases (cases generated by the original host), reported in superspreading events for a range of human diseases, are shown to be far greater that what we would expect if all hosts were intrinsically equal and all we saw was random variation around that mean. Strategies to control the spread of human infection will clearly be most effective if these superspreaders can be identified and targeted (Lloyd-Smith *et al.*, 2005).

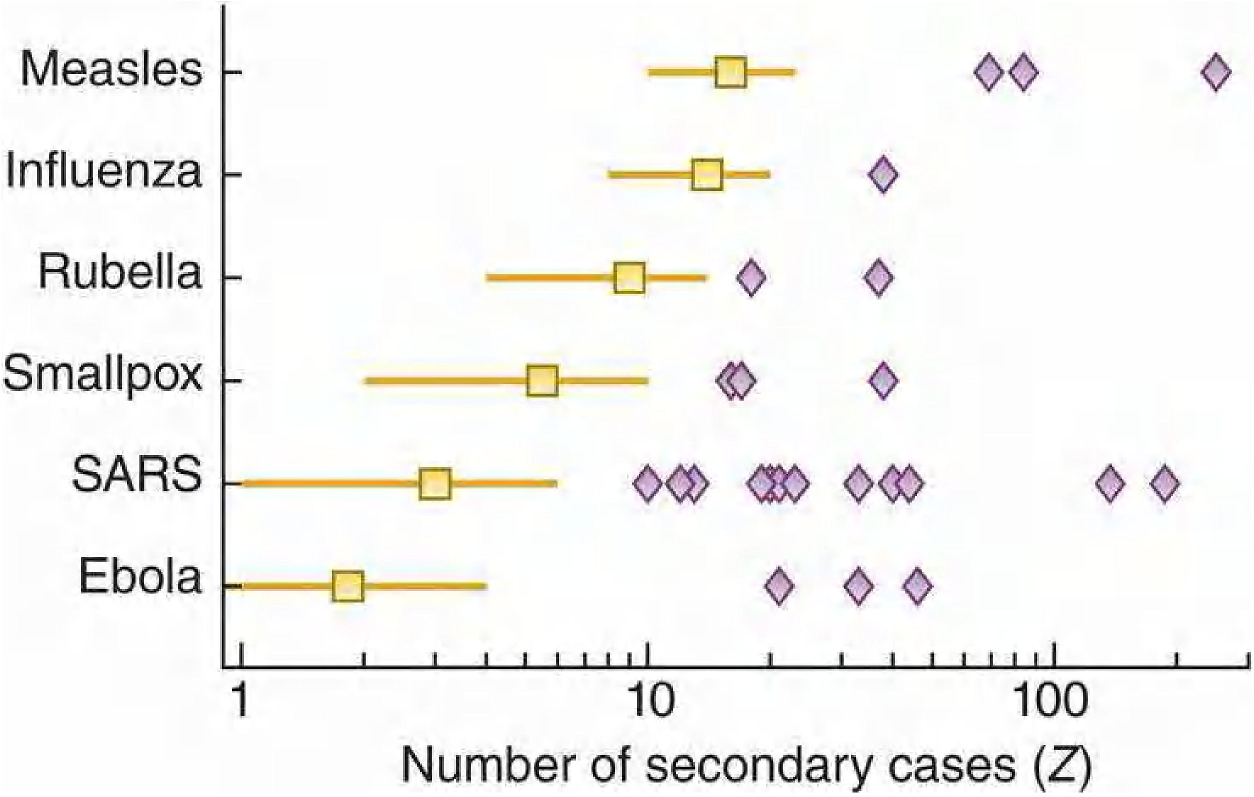


Figure 12.21 Superspreading events during outbreaks of six human diseases, as indicated. In each case, the yellow squares represent the number of secondary cases to be expected when all hosts are considered equal (the lines are their 99% CIs). The superspreaders spawned far more secondary cases (purple diamonds) than these expectations.

Source: After Lloyd-Smith *et al.* (2005).

12.5.3 Host diversity and the spread of disease

We noted earlier that host species naturally differ in their susceptibility to infection, and it follows from this that the spread or transmission of infection in a community of species depends on the mixture of susceptible and resistant types represented there. A simple example is shown in Figure 12.22 for essentially the same system as that in Figure 12.20, except that in this case there are two host species: relatively susceptible radish plants, *Raphanus sativus*, as before, but also more resistant mustard plants, *Sinapsis alba*. The spread amongst radish plants, generally, is slowed down by the mustard plants that separate them; and radishes may even be protected almost entirely by being surrounded by a cordon of mustard plants that the fungal infection finds it difficult to cross.

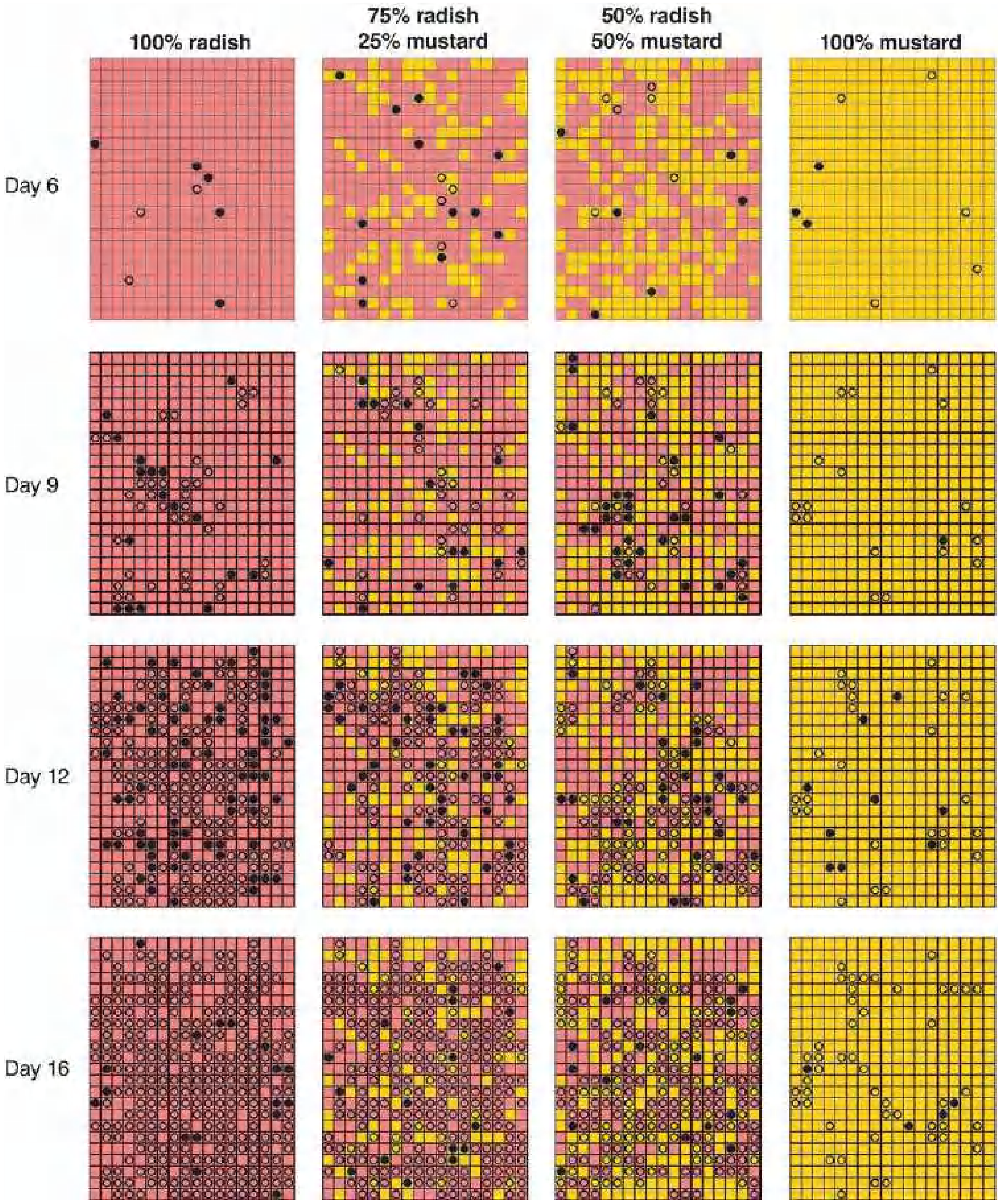

100% radish
75% radish
25% mustard
50% radish
50% mustard
100% mustard
Day 6
Day 9
Day 12
Day 16

Figure 12.22 The spread of damping off among radish plants is slowed down by mustard plants that separate them. The effect of various mixtures of susceptible and resistant hosts on the spread, up to day 16 of growth, of damping off disease, caused by the fungus *Rhizoctonia solani*. The susceptible hosts, radish plants, *Raphanus sativus*, are shown as pink squares; the resistant hosts, mustard plants, *Sinapsis alba*, are shown as yellow squares. Open circles are infected plants; closed circles are plants that became infected on the day in question. Overall infection rates are obviously lower when there are more resistant plants, but even amongst the susceptible radishes, 'protected' by the mustard plants, infection rates fell, especially in the middle of the period of spread on day 12, from 0.51 with 100% radishes, to 0.46 with 75% radishes and 0.39 with 50% radishes.

Source: After Otten *et al.* (2005).

APPLICATION 12.5 Lyme disease and the dilution effect

Even in more freely mixing populations, where for practical purposes the spatial locations of hosts can be ignored, the species mix in a community of hosts can have profound effects on the transmission and ultimately on the prevalence of infection. And if the infection itself is zoonotic, the mix of species may in turn determine the level of risk to humans.

Lyme disease is a good example. It is caused by a spirochaete bacterium (*Borrelia burgdorferi*) carried by ticks in the genus *Ixodes*, and affects tens of thousands of people around the world each year, though the strongest focus of infection is in the north-eastern USA. If untreated it can damage the heart and nervous system and lead to a type of arthritis. The spirochaete is transmitted naturally amongst a range of host species, carried by the ticks. But if an infected tick bites a human, Lyme disease may result. The ticks take two years to pass through four developmental stages. Eggs are laid in the spring and uninfected larvae take a single blood meal from a host (usually a small mammal or bird) before dropping off and moulting into a nymph, which overwinters. Infected hosts transmit the spirochaete to the larval ticks, which remain infective throughout their lives. Next year the nymph seeks a host in the spring/early summer for another single blood meal. This is the most risky stage for human infection because the nymphs are small and difficult to detect and attach to hosts at a time of peak human recreation in forests and parks. Between 1% and 40% of nymphs carry the spirochaete in Europe and the USA (Ostfeld & Keesing, 2000). Having fed, the nymph drops off and moults into an adult that takes a final blood meal and reproduces on a third host, often a larger mammal such as a deer. The adult females lay the eggs that initiate the next generation.

There is considerable variation among the potential mammal, bird and reptile hosts in the efficiency with which they are competent to transmit the spirochaete to the tick. In the eastern USA, by far the most competent transmitter of the spirochaete, and the most abundant small mammal host, is the white-footed mouse (*Peromyscus leucopus*). A number of studies have suggested that the prevalence of infection in the ticks is lower when the number of potential host species (host biodiversity) is higher (see for example Figure 12.23), and that this, combined with the abundance of the ticks, translates into a reduced risk to humans. Originally, this association between host species number and infection prevalence was attributed to a *dilution effect* (Van Buskirk & Ostfeld, 1995). The idea is that in communities where species numbers are low, white-footed mice are almost always present and often dominate, and so a high proportion of the tick bites are on these mice and result in effective transmission; whereas in communities where the numbers of host species are high, ticks often bite incompetent hosts bringing onward transmission to a dead end. Thus, the incompetent hosts 'dilute' the population of competent hosts – much like the mustard plants diluted the radish populations in Figure 12.22 – and overall rates of transmission, and the prevalence of infection, are reduced as a result. Subsequently, the use of the term dilution effect has been expanded to apply to the pattern – infection prevalence declining with host species number – irrespective of, or without actually knowing, the process leading to the pattern (Keesing *et al*., 2006). That process could be dilution in the strict sense, but it may, for example, be that white-footed mice are simply at higher densities when they are not sharing space and resources with other species and so transmission is more effective between them.

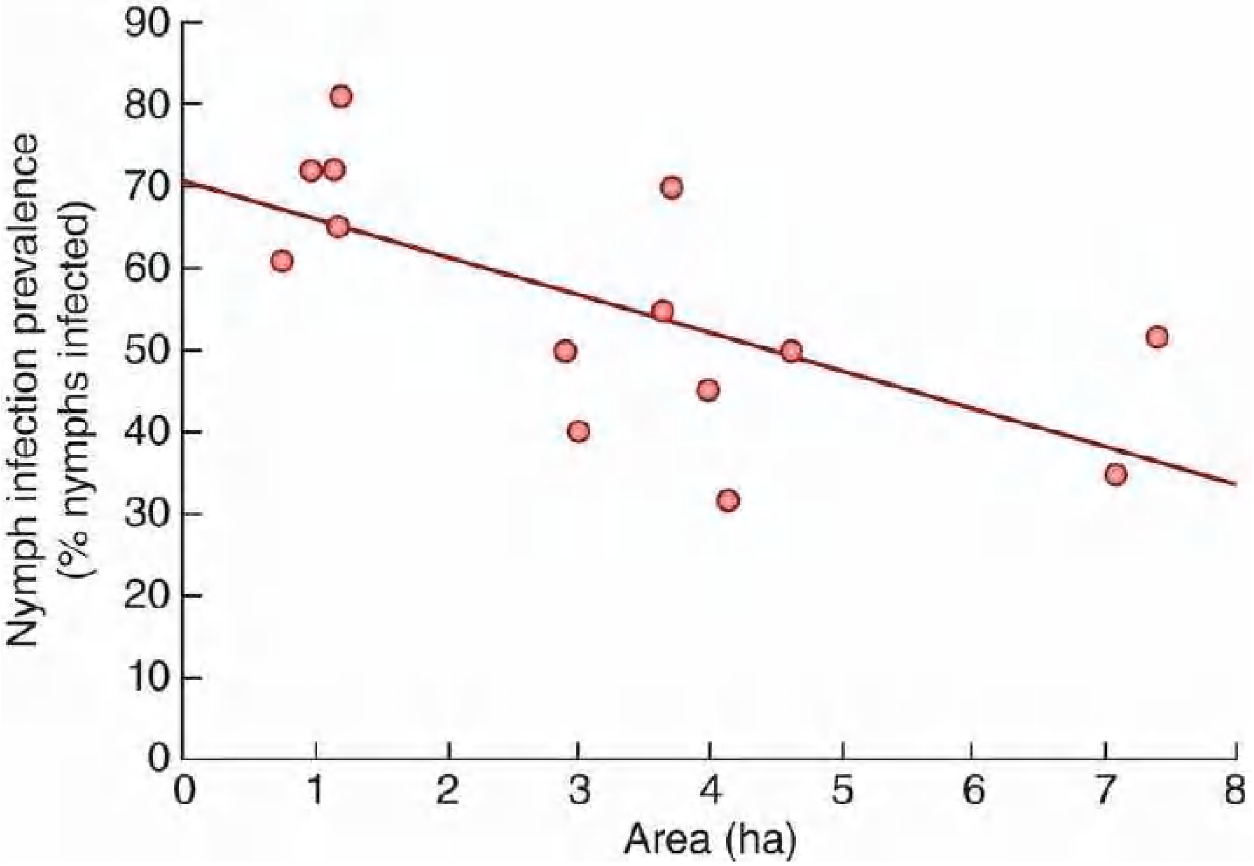


Figure 12.23 A dilution effect for the Lyme disease pathogen – or is it? The prevalence in nymphs of ticks, *Ixodes scapularis*, of infection with the agent causing Lyme disease, *Borrelia burgdorferi*, in woodlands of different areas in Dutchess County, New York. Higher prevalences of infection typically give rise to higher risks of human infection from tick bites. Smaller woodlands support fewer host species, though the most competent host, the white-footed mouse, *Peromyscus leucopus*, is present in all of them. The relationship can therefore be seen as reflecting an association between lower diversity of hosts (to the left) and higher prevalences of infection – a dilution effect. On the other hand, the densities of white-footed mice are also higher in the smaller woodlands, which will also favour the transmission of *B. burgdorferi* between them, and higher prevalences.

Source: After Allan *et al.* (2003).

There has been considerable, often heated, disagreement amongst workers in the field about whether the term dilution effect should be applied to the pattern irrespective of the process, or only to the pattern when it is generated by dilution, and especially about whether the pattern itself is common or perhaps even comes close to being a general rule. The question is an important one, for practical as well as scientific reasons, because it relates directly to the issue of whether the conservation of biodiversity, important for many other reasons (see Section 15.4), is valuable in terms of the protection it might afford to humans from zoonotic infectious diseases. On the one hand, Keesing *et al.* (2010) claim that 'current evidence indicates that preserving intact ecosystems and their endemic biodiversity should generally reduce the prevalence of infectious diseases', whereas Wood and Lafferty (2013) are more cautious, suggesting that biodiversity–disease relationships 'are likely to be complex and scale dependent, and that they might be negative, positive or neutral, depending on scale and ecological context'. In fact, as is often the case with these ecological (and other scientific?) disputes, the distance between the protagonists seems far shorter to the outsider that it does to the protagonists themselves. Both seem to agree, in fact, that it is not biodiversity in the limited sense of simply the *number* of species that is typically important, but rather the presence or absence of particular host species. This is an idea that we will meet again when we examine the relationship between biodiversity and food web stability (see Section 17.2).

12.6 The effects of parasites on the survivorship, growth and fecundity of hosts

According to strict definition, parasites cause harm to their host. But it is not always easy to demonstrate this harm, which may be detectable only at some peculiarly sensitive stage of the host's life history or under particular circumstances. There are of course, nonetheless, examples in which a detrimental effect of a parasite on host fitness has been demonstrated (see Tompkins & Begon (1999), for example, for a compilation).

effects are often subtle ...

The effects of parasitism on host fitness, though, are often subtle. For example, *Philornis downsi* is a fly parasite whose larvae inhabit birds' nests, chewing through the skin of nestlings and consuming blood and other fluids (see also Section 12.3.6). They were seen first on the Galápagos Islands in 1997, where they attacked a number of species including nine species of Darwin's finches. One of these, the medium ground finch, *Geospiza fortis*, has been the subject of field experiments to determine the effect of the parasite. Twenty-four nests constructed by the birds in tree cacti on Santa Cruz Island were fitted with a nylon liner to prevent the larvae reaching the nestlings from the nest material; a further 24 acted as unmanipulated (unlined) controls. The manipulation was only partially successful; there were around 22 *P. downsi* larvae per lined nest compared with around 37 in the unlined ones. Nonetheless, the experiment demonstrated that parasitism by the larvae was adversely affecting the finch nestlings. The nestlings grew faster in lined than in unlined nests (Figure 12.24a), and more lined nests fledged birds than unlined nests, while the total number of birds fledged from lined nests was also higher (Figure 12.24b).

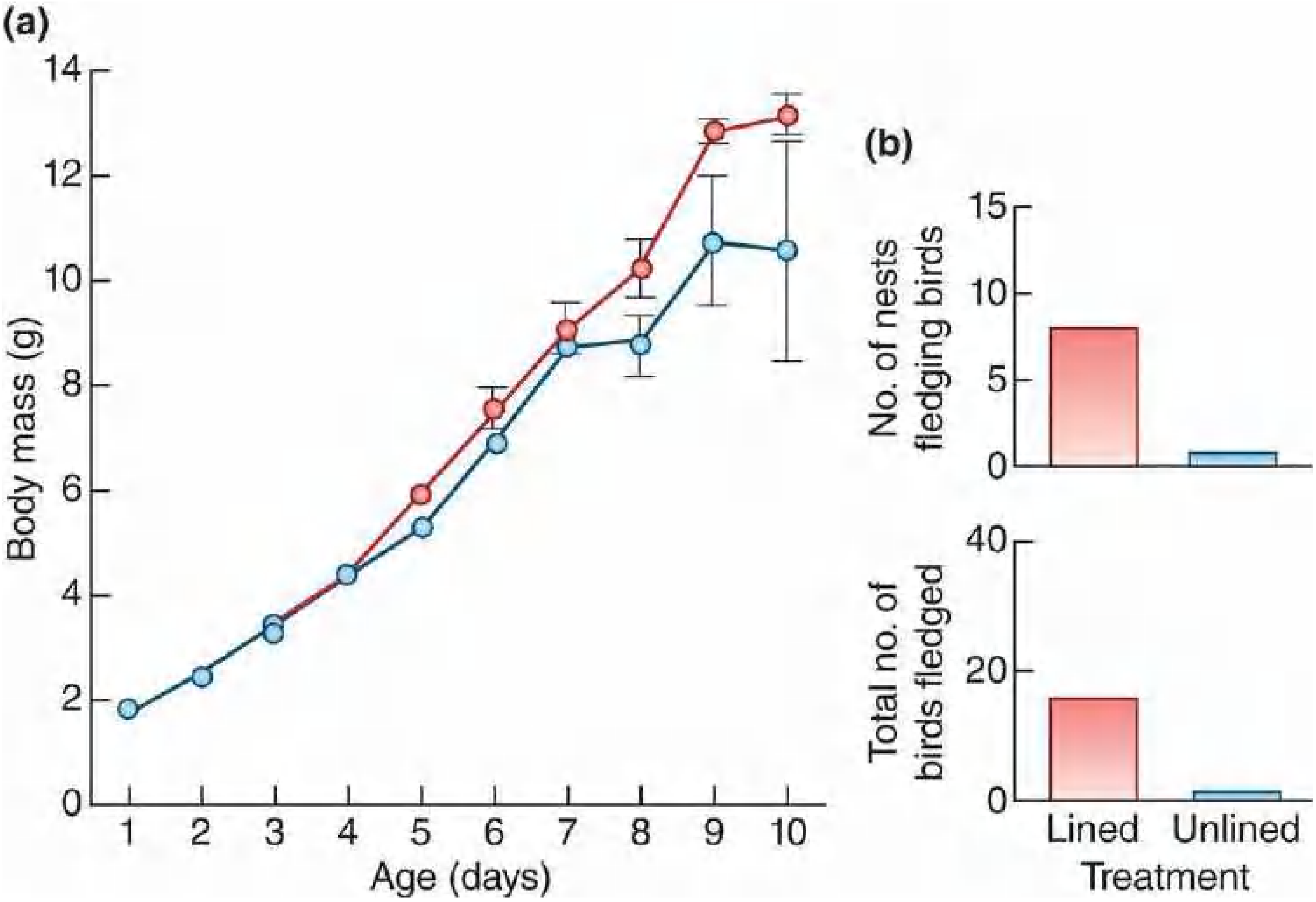


Figure 12.24 Harmful effects of fly larva parasitism on the medium ground finch, *Geospiza fortis*. (a) Nestlings in lined nests (red), providing protection from parasitism, grew faster than those in unlined nests (blue). Bars are SEs. (b) With equal numbers of nests lined and unlined, the number of lined nests fledging birds was higher (above; $P = 0.02$) as was the total number of bird fledged from lined nests (below; $P < 0.001$).

Source: After Koop *et al.* (2011).

... affecting an interaction

Effects, too, may not be direct but rather express themselves through an interaction with some other threat to host fitness, for example by making hosts more susceptible to predation. Thus, postmortem examinations of red grouse (*Lagopus lagopus scoticus*) showed that birds killed by predators carried significantly greater burdens of the parasitic nematode *Trichostrongylus tenuis* than the presumably far more random sample of birds that were shot (Hudson *et al.*, 1992a). Alternatively, the effect of parasitism may be to weaken an aggressive competitor and so allow weaker associated species to persist. For example, of two Anolis lizards that live on the Caribbean island of St Maarten, *Anolis gingivinus* is the stronger competitor and appears to exclude *A. wattsi* from most of the island. The malarial parasite *Plasmodium azurophilum* very commonly affects *A. gingivinus* but rarely affects *A. wattsi*. Wherever the parasite infects *A. gingivinus*, the species coexists with *A. wattsi*, but wherever the parasite is absent, only *A. gingivinus* occurs: *A wattsi* is excluded (Schall, 1992). Similarly, the holoparasitic plant, dodder (*Cuscuta salina*), which has a strong preference for *Salicornia* in a southern Californian salt marsh, is highly instrumental in determining the outcome of competition between *Salicornia* and other plant species within several zones of the marsh (Figure 12.25).

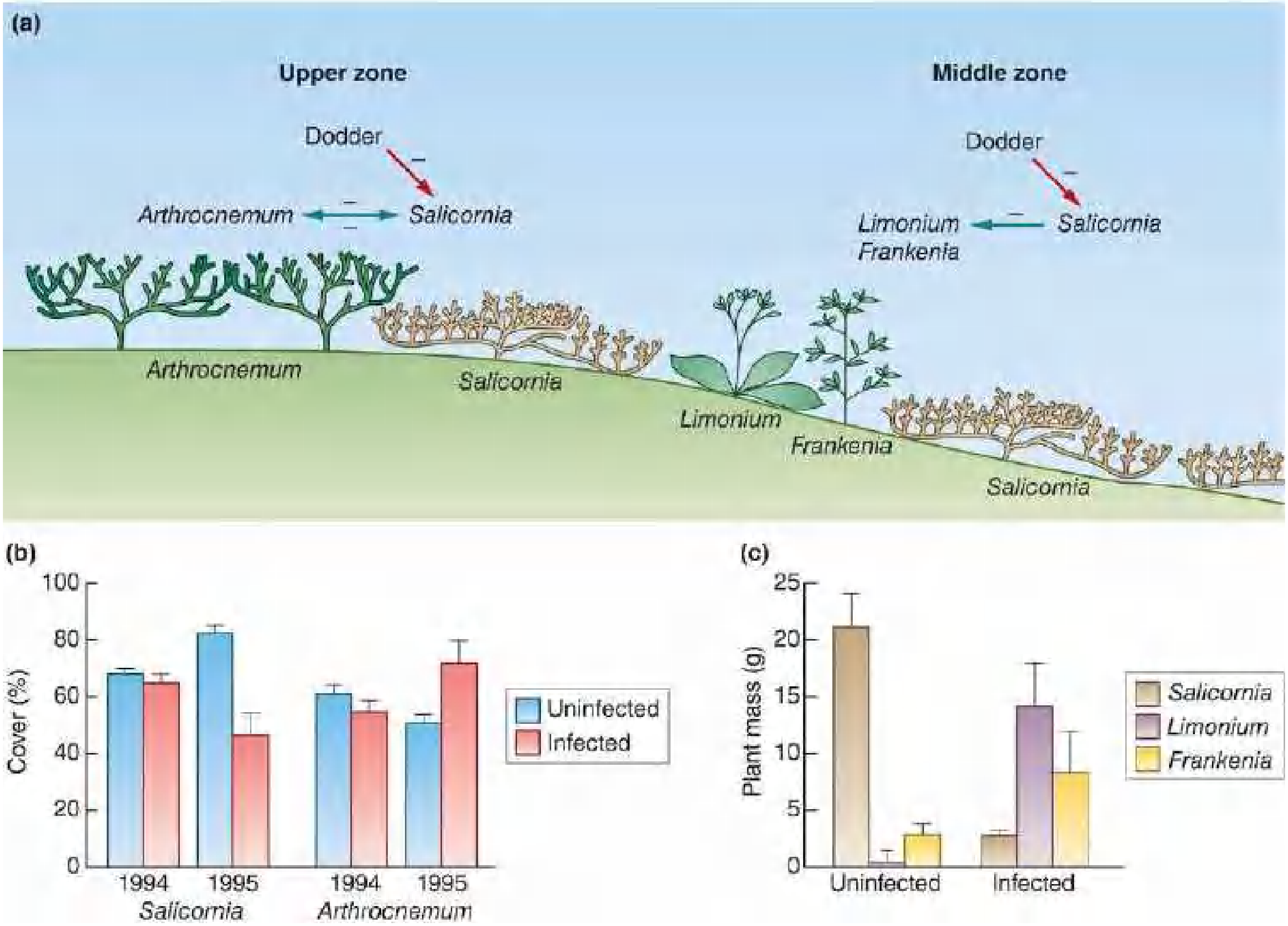


Figure 12.25 The effect of dodder, *Cuscuta salina*, on competition between *Salicornia* and other species in a southern Californian salt marsh, USA. (a) A schematic representation of the main plants in the community in the upper and middle zones of the marsh and the interactions between them (red arrows, parasitism; blue arrows, competition). *Salicornia* (the relatively low-growing plant in the figure) is most attacked by, and most affected by, dodder (which is not itself shown in the figure). When uninfected, *Salicornia* competes strongly and symmetrically with *Arthrocnemum* at the *Arthrocnemum–Salicornia* border, and is a dominant competitor over *Limonium* and *Frankenia* in the middle (high *Salicornia*) zone. However, dodder significantly shifts the competitive balances. (b) Over time, *Salicornia* decreased and *Arthrocnemum* increased in plots infected with dodder. (c) Large patches of dodder suppress *Salicornia* and favour *Limonium* and *Frankenia*.

Source: After Pennings & Callaway (2002).

Although parasites often affect their hosts not in isolation, but through an interaction with some other factor, this does not mean that the parasites play only a supporting role. Both partners in the interaction may be crucial in determining not only the overall strength of the effect but also which particular hosts are affected.

12.7 The population dynamics of infection

Establishing that parasites have a detrimental effect on host characteristics of demographic importance is a critical first step in establishing that parasites influence the population and community dynamics of their hosts. But it is only a first step. A parasite may increase mortality, directly or indirectly, or decrease fecundity, without this affecting levels or patterns of abundance. The effect may simply be too trivial to have a measurable effect at the population level, or other factors and processes may act in a compensatory fashion – for example, loss to parasites may lead to a weakening of density-dependent mortality at a later stage in the life cycle. The effects of rare, devastating epidemics, whether in humans, other animals or plants, are easy to see; but for more typical, endemic parasites and pathogens, moving from the host–individual to the host–population level offers an immense challenge.

effects on health or morbidity

The observations we made in Chapter 10 regarding the population dynamics of predator–prey and herbivore–plant interactions can be extended to parasites and hosts. Effects vary with the densities of both the parasites and hosts. Both infected and uninfected hosts can exhibit compensatory reactions that may greatly reduce the effects of infection on the host population as a whole. A range of outcomes is possible: varying degrees of reduction in host–population density, varying levels of parasite prevalence and various fluctuations in abundance. In contrast to predators, however, parasites often cause a reduction in the health or 'morbidity' of their host rather than its immediate death, and it is therefore usually difficult to disentangle the effects of the parasites from those of other factors with which they interact (see Section 12.6). Indeed, even when parasites cause a death, this may not be obvious without a detailed postmortem examination (especially in the case of microparasites). Also, plant pathologists and medical and veterinary parasitologists have generally studied parasites with known severe effects that live typically in dense and aggregated populations of hosts, paying little attention to the more typical effects of parasites in populations of 'wildlife' hosts. Elucidation of the role of parasites in host–population dynamics is therefore one of the major challenges facing ecologists.

Here, we begin by looking at the dynamics of infection within host populations without considering any possible effects on the total abundance of hosts. This 'epidemiological' approach (Anderson, 1991) has especially dominated the study of human disease, where we think of total abundance as being determined by a whole spectrum of factors and thus effectively independent of the prevalence of any one infection. We then take a more ecological approach by considering the effects of parasites on host abundance in a manner much more akin to conventional predator–prey dynamics.

12.7.1 The basic reproductive number and the transmission threshold

R_0, the basic reproductive number

In studies of the dynamics of parasite populations or the spread of infection, there are a number of key concepts. The first is the basic reproductive number, referred to as R_0. For microparasites, because infections (infected, infectious hosts) are the unit of study, this is defined as the average number of new infections that would arise from a single infectious host introduced into a population of susceptible hosts. For macroparasites, it is the average number of established, reproductively mature parasite offspring produced by a mature parasite throughout its life in a population of uninfected hosts.

the transmission threshold

This allows us to identify a transmission threshold, which must be crossed if an infection is to spread in a population. That threshold is defined by the condition $R_0 = 1$. An infection will eventually die out for $R_0 < 1$ (each present infection or parasite leads to less than one infection or parasite in the future), but an infection will spread for $R_0 > 1$. Insights into the dynamics of infection can be gained by considering the various determinants of R_0. We do this in some detail for directly transmitted microparasites, and then deal more briefly with related issues for indirectly transmitted microparasites and directly and indirectly transmitted macroparasites.

12.7.2 Directly transmitted microparasites: R_0 and the critical population size

For microparasites with direct, density-dependent transmission (see Section 12.5.2), R_0 increases with: (i) the average period of time over which an infected host remains infectious, L; (ii) the number of susceptible individuals in the host population, S, because greater numbers offer more opportunities for transmission of the parasite; and (iii) the transmission coefficient, β (see Section 12.5.2). Thus, overall:

$$R_0 = S\beta L. \tag{12.5}$$

Note immediately that the greater the number of susceptible hosts, the higher the basic reproductive number of the infection.

the critical population size ...

The transmission threshold can now be expressed in terms of a *critical population size*, S_T, where, because $R_0 = 1$ at that threshold:

$$S_T = 1/(\beta L). \tag{12.6}$$

This tells us that in populations with numbers of susceptibles less than this, the infection will die out ($R_0 < 1$), while with numbers greater than this the infection will spread ($R_0 > 1$). (S_T is often referred to as the critical community size because it has mostly been applied to human ‘communities’, but this is potentially confusing in a wider ecological context.) These simple considerations allow us to make sense of some very basic patterns in the dynamics of infection (Anderson, 1982 ; Anderson & May, 1991).

... for different types of parasite

Consider first the kinds of population in which we might expect to find different sorts of infection. If microparasites are highly infectious (β is large), or give rise to long periods of infectiousness (L is large), then they will have relatively high R_0 values even in small populations and will therefore be able to persist there (S_T is small). Conversely, if parasites are of low infectivity or have short periods of infectiousness, they will have relatively small R_0 values and will only be able to persist in large populations. Many protozoan infections of vertebrates, and also some viruses such as herpes, are persistent within individual hosts (large L), often because the immune response to them is either ineffective or short-lived. A number of plant diseases, too, like club root, have very long periods of infectiousness. In these cases, the critical population size is therefore small, explaining why they can and do survive endemically even in small host populations.

On the other hand, the immune responses to many other human viral and bacterial infections are powerful enough to ensure that they are only very transient in individual hosts (small L), and they often induce lasting immunity. Thus, for example, a disease like measles has a critical population size of around 300 000 individuals, and is unlikely to have been of great importance until quite recently in human biology. However, it generated major epidemics in the growing cities of the industrialised world in the 18th and 19th centuries, and in the growing concentrations of population in the developing world in the 20th century, such that in 1980 it was estimated to be causing 2.6 million deaths each year, though this was down to around 100 000 in 2013 (GBD

2013 Mortality and Causes of Death Collaborators, 2015), largely as a result of vaccination (see Section 12.7.5).

frequency-dependent transmission

Suppose, however, that transmission is frequency-dependent (see Section 12.5.2), as it may be, for example, with sexually transmitted diseases. Then there is no longer the same dependence on the number of susceptibles, and the basic reproductive number is simply given by:

$$R_0 = \beta' L. \tag{12.7}$$

Here, there is apparently no threshold population size and such infections can therefore persist even in extremely small populations where the rate of sexual contact may be the same as in large populations. Of course, long-term persistence in an isolated small population is unlikely, since the probability of stochastic extinction is not negligible. But if there are multiple small populations, a metapopulation perspective will be more appropriate (see Section 12.7.7).

12.7.3 Epidemic curves

The value of R_0 is also related to the nature of the *epidemic curve* of an infection. This is the time series of new cases following the introduction of the parasite into a population of hosts. Assuming there are sufficient susceptible hosts present for the parasite to invade (i.e. the critical population size, S_T, is exceeded), the initial growth of the epidemic will be rapid as the parasite sweeps through the population of susceptibles. But as these susceptibles either die or recover to immunity, their number, S, will decline, and so too therefore will R_0 (see Equation 12.5). Hence, the rate of appearance of new cases will slow down, reach a peak, and then decline. And if S falls below S_T and stays there, the infection will disappear – the epidemic will have ended. Not surprisingly, the higher the initial value of R_0, the more rapid will be the rise in the epidemic curve. But this will also lead to the more rapid removal of susceptibles from the population and hence to an earlier end to the epidemic: higher values of R_0 tend to give rise to shorter, sharper epidemic curves. Epidemic curves for Ebola in West Africa between 2013 and 2016 are shown in Figure 12.26a.

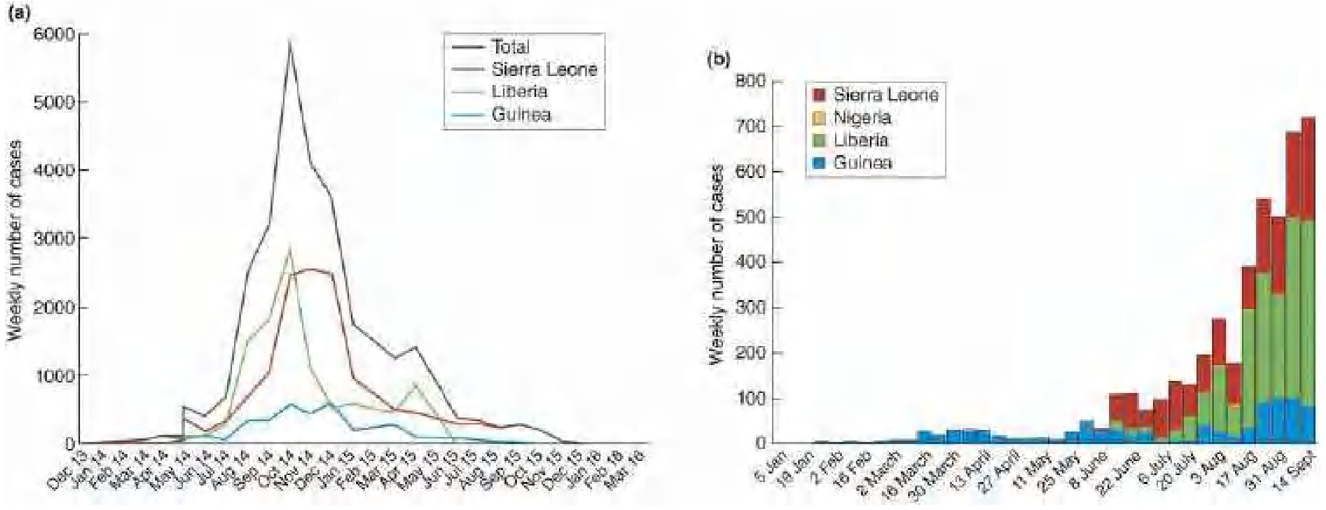


Figure 12.26 Epidemic curves for Ebola virus disease. (a) Epidemic curves in Guinea, Liberia and Sierra Leone (and the total of these) from December 2013 to March 2016. (b) More detail of the first nine months of the epidemic in 2014, also including Nigeria.

Source: (a) After Shultz *et al.* (2016). (b) After WHO Ebola Response Team (2014).

APPLICATION 12.6 Epidemic forecasting for Ebola

An understanding of how epidemic curves are generated in populations can enable us to anticipate the future phases of a curve so that, for example, future risk of a human, veterinary or agricultural disease might be predicted, or the likely consequences of alternative intervention strategies compared. Thus, Figure 12.26a shows the final curves for the West African Ebola epidemic, right through to its conclusion. But after the first nine months of the epidemic, the early phases of the curves (Figure 12.26b) were analysed by a World Health Organization (WHO) team to allow the likely seriousness of the threat to be precisely assessed. Ignoring details, the analysis involved using the available data to estimate R_0, and then projecting forward to where that R_0 value, if sustained, would lead. Those estimated R_0 values were 1.81, 1.51 and 1.38 for Guinea, Liberia and Sierra Leone, respectively. Clearly, the epidemic was expected to spread in all three countries as R_0 was greater than 1. Indeed, projecting forward from early September 2014, these values suggested a weekly number of cases of around 9000 across the three countries by mid-November, and the WHO team sent out a warning: 'The risk of continued epidemic expansion and the prospect of endemic EVD [Ebola virus disease] in West Africa call for the most forceful implementation of present control measures and for the rapid development and deployment of new drugs and vaccines'. Fortunately, that warning, based on the epidemic curve analysis, was heeded, at least to the extent that a major local and global response was mounted. The weekly number of cases in mid-November, in the face of this response, was less than 6000 and that was the peak, falling below 2000 by the turn of the year (Figure 12.26a). Of course, these numbers are deeply saddening, but without the analysis, the warning and the response, they could have been far worse.

12.7.4 Dynamic patterns of different types of parasite

Turning now to the longer term patterns in the dynamics of different types of endemic infection, we should note first that the immunity induced by many bacterial and viral infections reduces S, the number of susceptibles, which reduces R_0, which therefore tends to lead to a decline in the incidence of the infection itself. However, in due course, and before the infection disappears altogether from the population, there is likely to be an influx of new susceptibles into the population, a subsequent increase in S and R_0, and so on. Thus, such infections tend to generate a sequence from 'many susceptibles (R_0 high)', to 'high incidence', to 'few susceptibles (R_0 low)', to 'low incidence', to 'many susceptibles', etc. – just like any other predator–prey cycle. This undoubtedly underlies the observed cyclic incidence of many human diseases, with the differing lengths of cycle reflecting the differing characteristics of the diseases: measles with peaks every one or two years (Figure 12.27a), pertussis (whooping cough) every 3–4 years (Figure 12.27b), diphtheria every 4–6 years, and so on (Anderson & May, 1991).

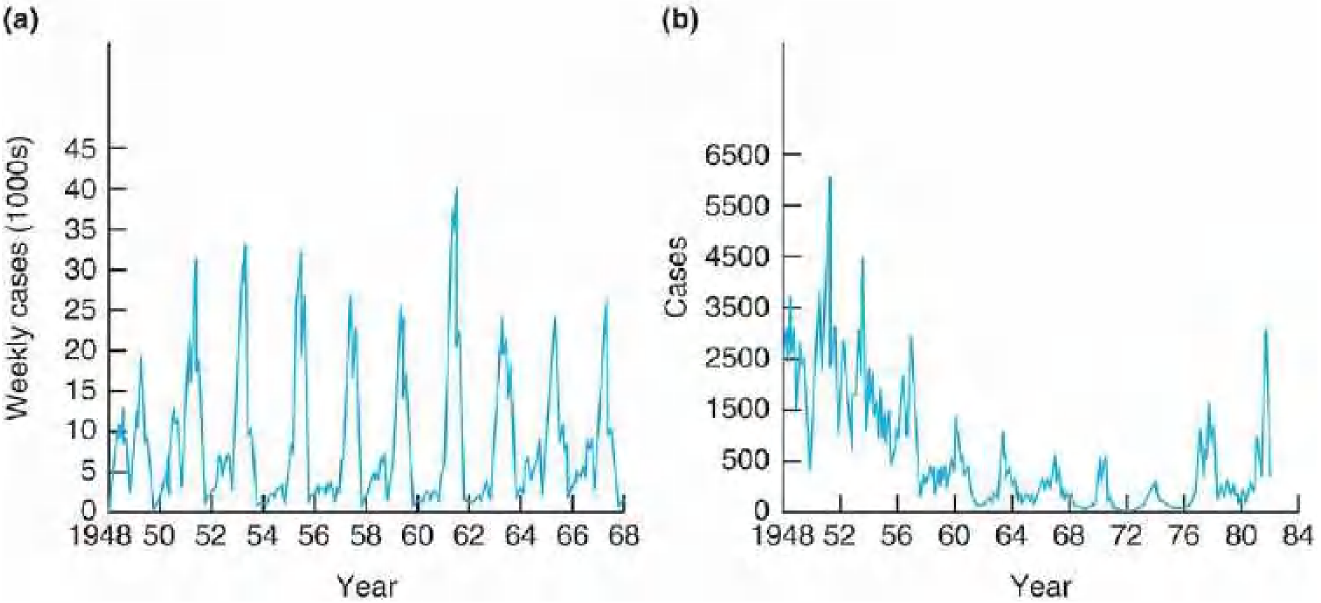


Figure 12.27 **Cycles in the incidence of human infections.** (a) Reported cases of measles in England and Wales from 1948 to 1968, prior to the introduction of mass vaccination. (b) Reported cases of pertussis (whooping cough) in England and Wales from 1948 to 1982. Mass vaccination was introduced in 1956.

Source: After Anderson & May (1991).

By contrast, infections that do not induce an effective immune response tend to be longer lasting within individual hosts, but not to give rise to the same sort of fluctuations in S and R_0. Thus, for example, protozoan infections are generally much less variable (less cyclic) in their prevalence.

12.7.5 Immunisation and herd immunity

Focusing on critical population sizes also sheds light on immunisation programmes, in which susceptible hosts are rendered non-susceptible without ever becoming diseased (showing clinical symptoms), usually through exposure to a killed or attenuated pathogen. The direct effects here are obvious: the immunised individual is protected. But, by reducing the number of susceptibles, such programmes also have the indirect effect of reducing R_0. Indeed, seen in these terms, the fundamental aim of an immunisation programme is clear – to hold the number of susceptibles below S_T so that R_0 remains less than 1. To do so is said to provide *herd immunity*.

APPLICATION 12.7 Critical vaccination coverage

A simple manipulation of Equation 12.5 gives rise to a formula for the critical proportion of the population, p_c, that need to be immunised in order to provide herd immunity (reducing R_0 to a maximum of 1, at most). If we define S_0 as the typical number of susceptibles prior to any immunisation, and note that S_T is the number still susceptible (not immunised) once the programme to achieve $R_0 = 1$ has become fully established, then the proportion immunised is:

$$p_c = 1 - (S_T/S_0). \tag{12.8}$$

The formula for S_T is given in Equation 12.6, whilst that for S_0, from Equation 12.5, is simply $R_0/\beta L$, where R_0 is the basic reproductive number of the infection prior to immunisation. Hence:

$$p_c = 1 - (1/R_0). \tag{12.9}$$

This reiterates the point that in order to eradicate a disease, it is not necessary to immunise the whole population – just a proportion sufficient to bring R_0 below 1. It also shows that this proportion will be higher the greater the ‘natural’ basic reproductive number of the disease (without immunisation). This general dependence of p_c on R_0 is illustrated in Figure 12.28, with the estimated values for a number of human diseases indicated on it. Note that smallpox, the only disease where in practice immunisation seems to have led to eradication, has unusually low values of R_0 and p_c. In practice, of course, finding a vaccination proportion that is effective in real populations will require more than a simple calculation of R_0. A study of rabies in domestic dogs in Tanzania provides a good illustration: R_0 calculations suggest a figure of around 20–45%, but wider considerations indicate that a figure of around 70% may be necessary (Hampson *et al.*, 2009).

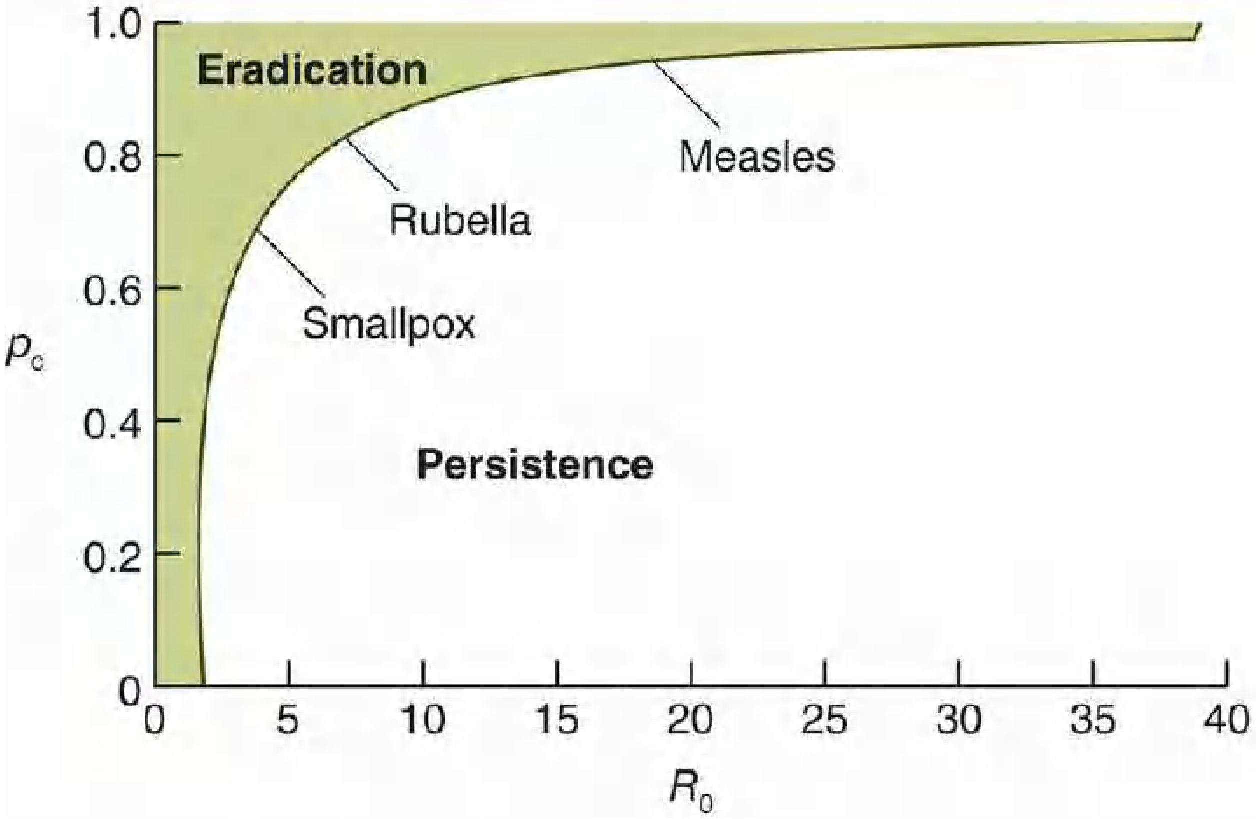


Figure 12.28 Diseases with a greater R_0require greater vaccination coverage. The dependence of the critical level of vaccination coverage required to halt transmission, p_c, on the basic reproductive number, R_0, with values for some common human diseases, as indicated.

Source: After Anderson & May (1991).

Figure 12.28 also shows that a particularly high rate of coverage is required to eliminate measles, making it particularly vulnerable to drops in coverage rate. It is therefore especially tragic that a widely publicised but now utterly discredited study in the late 1990s sought to link the MMR vaccine (against measles, mumps and rubella) with autism and bowel disease, leading to a sharp drop in vaccine uptake. Moreover, while in England, for example, rates of uptake among young children have now mostly recovered, a vulnerable, undercovered cohort has now grown up, and it is therefore notable that in an outbreak in 2018 (876 confirmed cases by 10 September compared with 267 for the whole of 2017), 82% had not been vaccinated and a large proportion of these, more than 50%, were over 15 years old (Bedford & Elliman, 2018).

12.7.6 Crop pathogens: macroparasites viewed as microparasites

Most of plant disease epidemiology has been concerned with the dynamics of diseases within crops, and hence with the spread of a disease within a generation. Moreover, although most commonly studied plant pathogens are macroparasites in the sense we have defined them (lesions of infection generating infective spores), they are typically treated like microparasites with disease being monitored on the basis of the proportion of the population infected (i.e. prevalence). The progress of infection can be modelled simply by distinguishing y_t, the proportion of the population affected by disease lesions at time t, from $(1 - y_t)$, the proportion without lesions and thus susceptible to infection, but also acknowledging that there is usually a latent period, length p, between a lesion being initiated and becoming infectious itself, and a further period, l, over which it remains infectious. Hence, the proportion of the population affected by *infectious* lesions at time t is $(y_{t-p} - y_{t-p-l})$. The rate of increase in the proportion of a plant population affected by lesions (see Gilligan, 1990) may thus be given by:

$$dy_t/dt = D(1 - y_t)\left(y_{t-p} - y_{t-p-l}\right), \tag{12.10}$$

which is essentially a βSI formulation, with D the plant pathologists' version of a transmission coefficient. This gives rise to S-shaped curves for the progress of diseases within a crop that broadly match the data derived from many crop–pathogen systems (Figure 12.29). Increasingly, though, more complex formulations are used to track the progress of crop disease that distinguish establishment and secondary phases of infection and acknowledge a potential loss of infectiousness over time (see, for example, Motisi *et al.*, 2013).

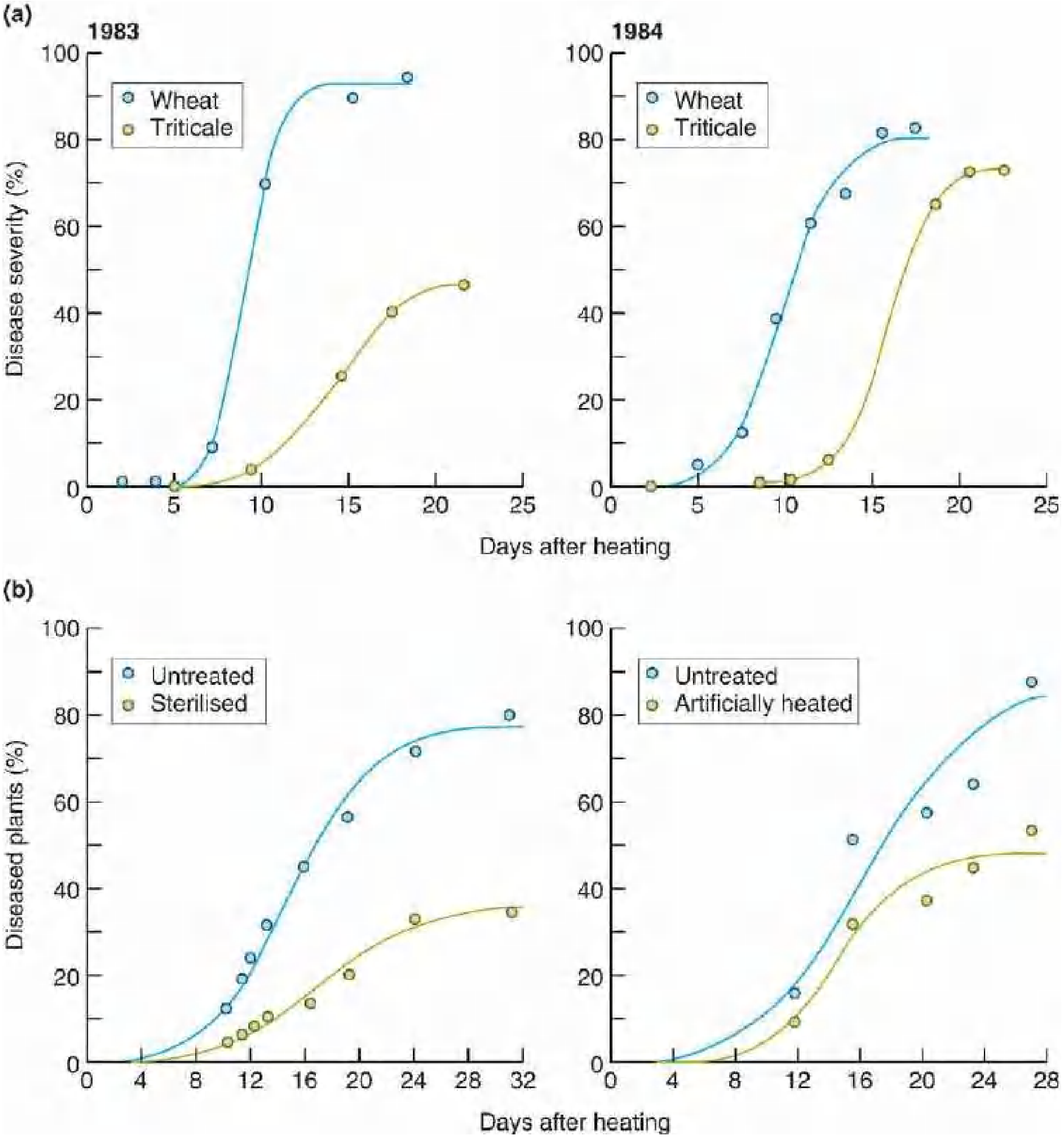


Figure 12.29 **S-shaped curves of the progress of diseases through crops from an initial inoculum to an asymptotic proportion of the total population infected.** (a) *Puccinia recondita* attacking wheat (cultivar Morocco) and triticale (a crop derived from the hybridisation of wheat and rye) in 1983 and 1984. (b) *Fusarium oxysporum* attacking tomatoes in experiments comparing untreated and sterilised soil and untreated and artificially heated soil. The original data sources and methods of curve-fitting may be found in Gilligan (1990).

Source: After Gilligan (1990).

APPLICATION 12.8 Other classes of parasite and their control

vector-borne infections

For microparasites that are spread from one host to another by a vector, the life cycle characteristics of both the host and vector enter into the calculation of R_0. In particular, the transmission threshold ($R_0 = 1$) is dependent on a ratio of vector : host numbers. For a disease to establish itself and spread, that *ratio* must exceed a critical level – hence, disease control measures are usually aimed directly at reducing the numbers of vectors, and are aimed only indirectly at the parasite. Many virus diseases of crops, and vector-transmitted diseases of humans and their livestock (malaria, onchocerciasis, etc.), are controlled by insecticides rather than chemicals directed at the parasite. For example, of the three primary interventions recommended by the WHO Global Malaria Program (WHO, 2007), the second and third are focused directly on the mosquito vector, namely: (i) diagnosis of malaria cases and treatment with effective medicine; (ii) distribution of insecticide-treated bed nets to achieve full coverage of populations at risk of malaria; and (iii) indoor residual spraying with insecticide.

directly transmitted macroparasites

The effective reproductive number of a directly transmitted macroparasite (no intermediate host) is directly related to the length of its reproductive period within the host (i.e. again, to *L*) and to its rate of reproduction (rate of production of infective stages). Both of these are subject to density-dependent constraints that can arise either because of competition between the parasites, or commonly because of the host's immune response (see Section 12.3.7). The intensity of these constraints varies with the distribution of the parasite population between its hosts and, as we have seen, aggregation of the parasites is the most common condition. This means that a very large proportion of the parasites exist at high densities where the constraints are most intense, and this tightly controlled density dependence undoubtedly goes a long way towards explaining the observed stability in prevalence of many helminth infections (such as hookworms and roundworms) even in the face of perturbations induced by climatic change or human intervention (Anderson, 1982).

Most directly transmitted helminths have an enormous reproductive capability. For instance, the female of the human hookworm *Necator* produces roughly 15 000 eggs per worm per day, whilst the roundworm *Ascaris* can produce in excess of 200 000 eggs per worm per day. The critical threshold densities for these parasites are therefore very low, and they occur and persist endemically in low-density human populations, such as hunter–gatherer communities.

indirectly transmitted macroparasites

Density dependence within hosts also plays a crucial role in the epidemiology of indirectly transmitted macroparasites, such as schistosomes that pass through both vertebrates and snails to complete their life cycle. In this case, however, the regulatory constraints can occur in either or both of the hosts. With human schistosomiasis, adult worm survival and egg production are influenced in a density-dependent manner in the human host; but production of infective *cercariae* by the snail is virtually independent of the number of infective *miracidia* that penetrate the snail. Thus, levels of schistosome prevalence tend to be stable and resistant to perturbations from outside influences.

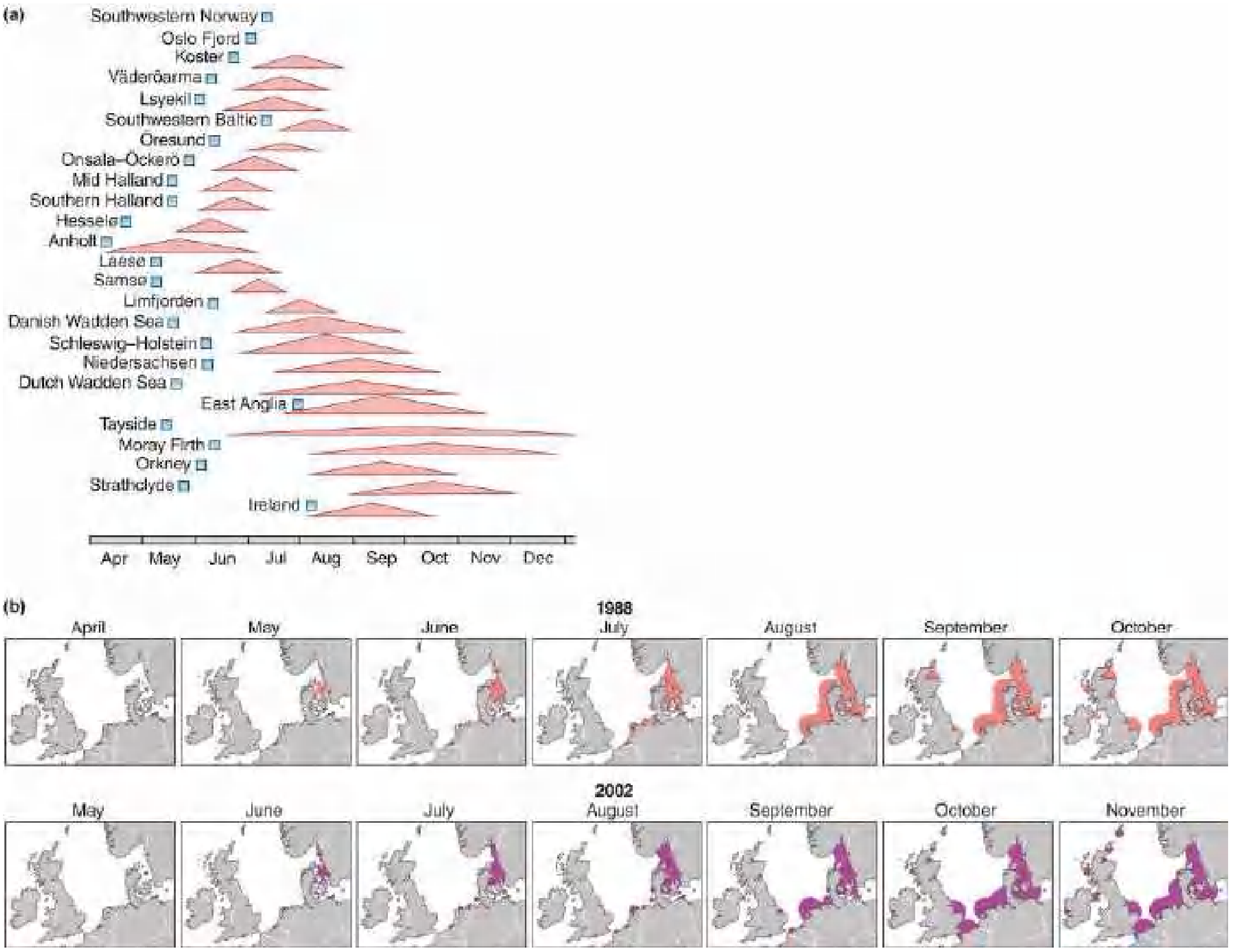
The threshold for the spread of infection depends directly on the abundance of both humans and snails (i.e. a product as opposed to the ratio that was appropriate for vector-transmitted microparasites). This is because transmission in both directions is by means of free-living infective stages. Thus, since it is inappropriate to reduce human abundance, schistosomiasis has often been controlled by reducing snail numbers with molluscicides in an attempt to depress R_0 below unity (the transmission threshold). The difficulty with this approach, however, is that the snails have an enormous reproductive capacity, and they rapidly recolonise aquatic habitats once molluscicide treatment ceases. Modern emphasis is therefore on water, sanitation and hygiene education (WASH), eliminating snail habitat where possible and limiting contact with infected water and live infective stages, but also on large-scale administration of the anti-schistosomal drug praziquantel to school-aged children and others at high risk (Grimes *et al*., 2015).

12.7.7 Parasites in metapopulations

With host–parasite dynamics, as with other areas of ecology, there is increasing recognition that populations cannot be seen as either homogeneous or isolated. Rather, hosts are usually distributed amongst a series of subpopulations, linked by dispersal between them, which together comprise a 'metapopulation' (see Section 6.7). Such a perspective immediately changes our view of what is required of a host population if it is to support a persistent population of parasites. Most fundamentally, a population may be too small, taken as a whole, to sustain an infection but may then become capable of doing so if broken into fragments. This arises as a kind of 'rescue effect' (see Section 6.7.3) – an infection may repeatedly die out in individual subpopulations, which are then reinfected by migrants. However, this is not an inevitability. Whether or not it occurs depends on there being the right balance between host birth and death rates (there needs to be a sufficient flow of new susceptibles into the subpopulations), the epidemiology of the infection (for example, hosts need to remain infectious for long enough) and the number and spatial arrangement of the subpopulations – too much exchange of hosts and the metapopulation is barely different from an undivided one: too little and there is no rescue effect (Swinton *et al*., 1998).

APPLICATION 12.9 Phocine distemper virus in harbour seals

Following an epidemic of phocine distemper virus spreading through and decimating the North Sea population of harbour seals, *Phoca vitulina*, in 1988 (Figure 12.30a), the metapopulation properties that would allow the infection to persist were analysed. The conclusion was that the high rates of transmission between the subpopulations, combined with the low rates of birth, made persistence effectively impossible (Swinton *et al.*, 1998). This, certainly, is consistent with the disease spreading rapidly though the metapopulation (there was an epidemic) but then disappearing (no endemic state was achieved). Something very similar, from the same starting point, happened in 2002 (Figure 12.30b). It seems likely that this is an example of an infection spilling over from other species where it causes little or no mortality (see Section 12.3.2) – in this case, other seal species in the Arctic. Seal conservation managers clearly need to understand the details of this process (Härkönen *et al.*, 2006).



[**Figure 12.30**](#) **The spread of phocine distemper virus epidemics**. (a) The spread of the epidemic afflicting harbour seal subpopulations in the North Sea in 1988, starting in Anholt. Sites above Anholt in the list represent a spread from Denmark to Norway and then Sweden. Those below Anholt represent a spread to the Netherlands, Great Britain and Ireland. Blue squares denote first recorded cases. Red triangles represent the period and number (log scale) of the median 90% of cases. (b) The cumulative spread of the disease in 1988 and then again in 2002, showing how both started from Anholt off the coast of Denmark.

Source: (a) After Swinton *et al*. (1998). (b) After Härkönen *et al*. (2006).

12.8 Parasites and the population dynamics of hosts

The previous section confirmed how important the host population is for the dynamics of infection. But what role, if any, do parasites and pathogens play in the dynamics of their hosts? This is a key but largely unanswered question in population ecology (see Tompkins *et al*. (2011) for a review). Data in Section 12.6 showed that parasites may affect host characteristics of demographic importance (birth and death rates), though even these data are relatively uncommon; and there are mathematical models showing that parasites have the potential to have a major impact on the dynamics of their hosts. But neither of these establishes that dynamics are actually affected. There are epidemics that may have devastating effects on the host, as we saw for phocine distemper virus in seals, though it is significant in that case that the infection appeared to have spilled over from reservoir hosts in which the virus had no demonstrable effect on host dynamics. We saw something similar in Section 8.8 when considering the apparent competition between red and grey squirrels in the UK, mediated by squirrelpox virus, which has no demonstrable effect on grey squirrel abundance, but spills over into reds, where its effect has been devastating. Epidemics occur in plant populations, too. For example, Dutch elm disease, caused by fungi of the genus *Ophiostoma* and carried by bark beetles, has spread across huge swathes of Europe and the USA (Figure 12.31), often eradicating populations of elm trees (*Ulmus* spp.) in its wake. In this case, it seems to have originated in Asian elm species where it does little harm, spreading through Europe from around 1910 and arriving in the USA with a shipment of timber in 1928.

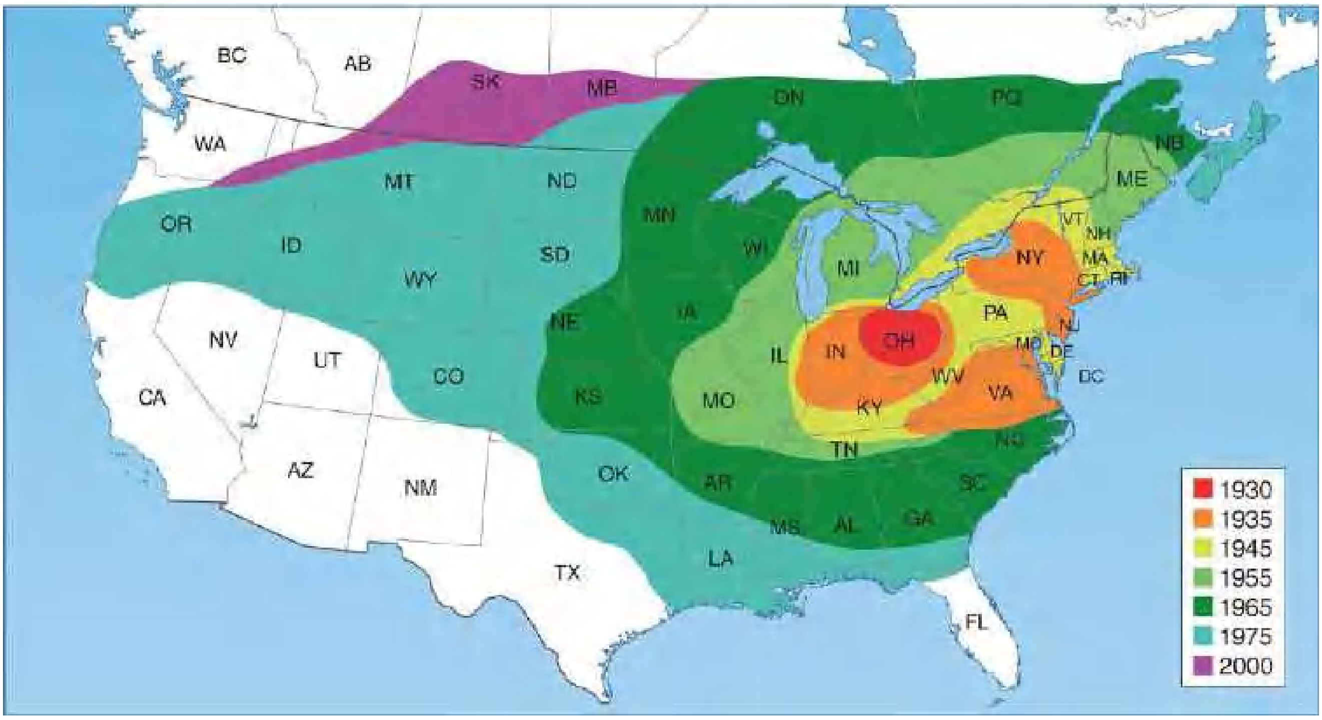
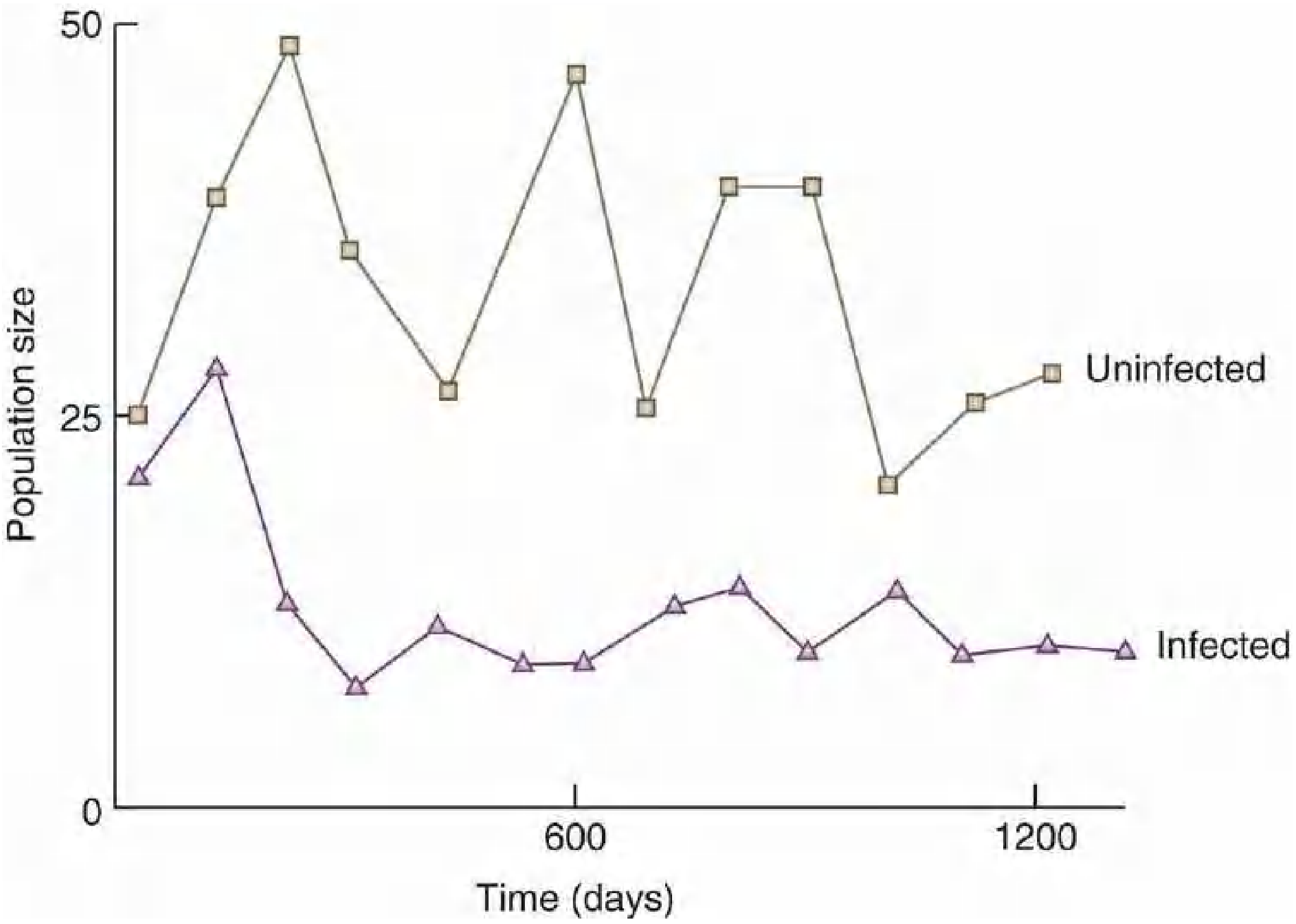


Figure 12.31 The spread of Dutch elm disease in North America from its introduction in infected timber in 1928 through to the end of the 20th century.

Source: Courtesy of Sunday Oghiake, University of Manitoba Entomology Department.

The widespread and intensive use of sprays, injections and medicines in agricultural and veterinary practice, too, all bear witness to the disease-induced loss of yield that would result in their absence. Datasets from controlled, laboratory environments showing reductions in host abundance due to parasites have also been available for many years (Figure 12.32). However, good evidence for endemic parasites having significant effects on the dynamics of natural populations is extremely rare. Part of the problem is the difficulty of proving that effects are significant even if they do exist. Even when a parasite is present in one population but absent in another, the parasite-free population is certain to live in an environment that is different from that of the infected population; and it is likely also to be infected with some other parasite that is absent from or of low prevalence in the first population.



[Figure 12.32](#) A protozoan parasite reduces the abundance of a beetle. Depression of the population size of the flour beetle, *Tribolium castaneum*, infected with the protozoan parasite *Adelina triboli*.

Source: After Park (1948).

One relatively straightforward, though nonetheless still rare, way in which an effect can be indicated is illustrated in Figure 12.33. Anther smut is a sexually transmitted disease carried by pollinating insects, caused by the fungus *Microbotryum violaceum*, that affects plants of the genus *Silene*, including *S. latifolia*, white campion. The figure shows that populations of white campion in Virginia, USA, are usually able to increase in abundance when the prevalence of infection is low, but decrease when prevalence is high.

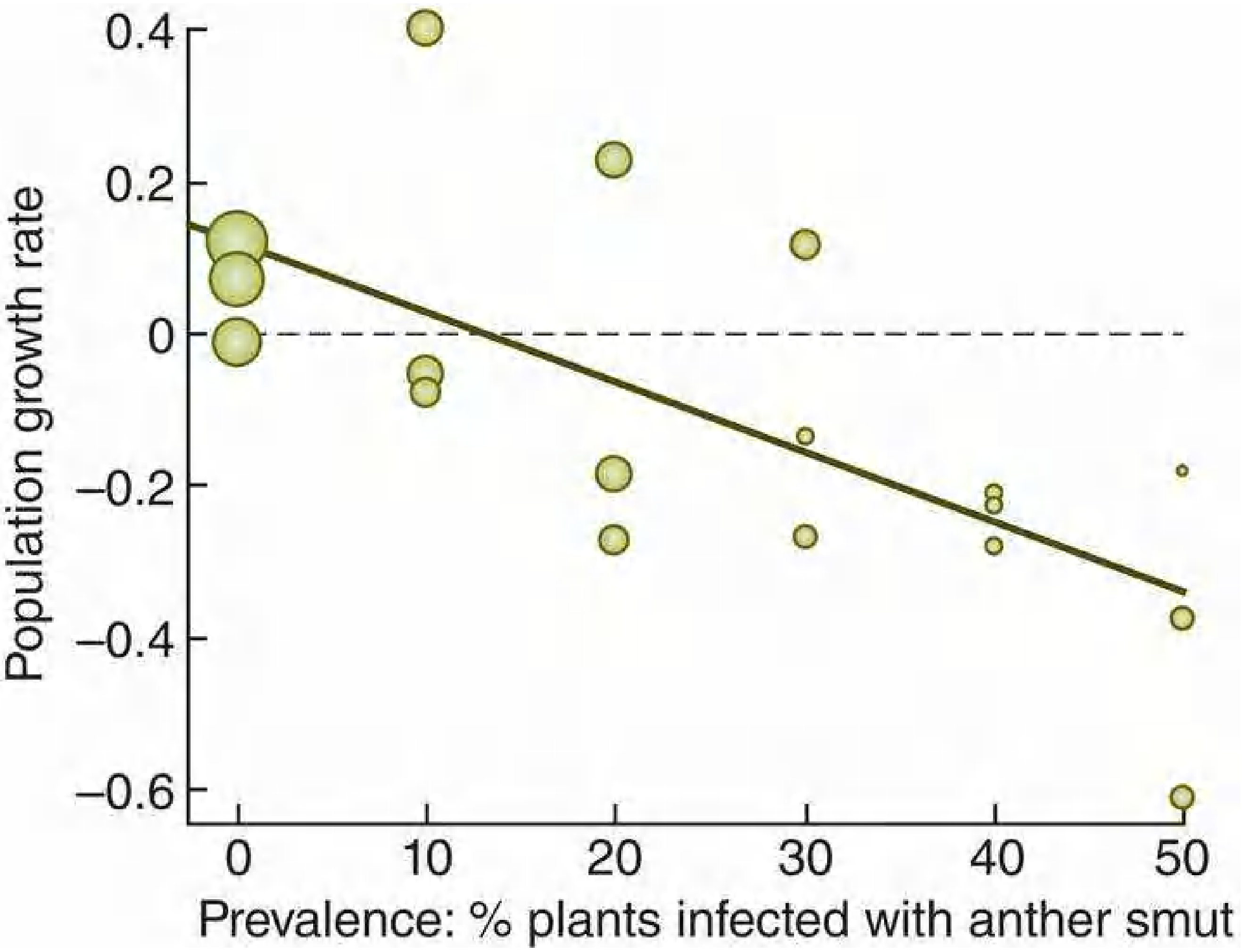
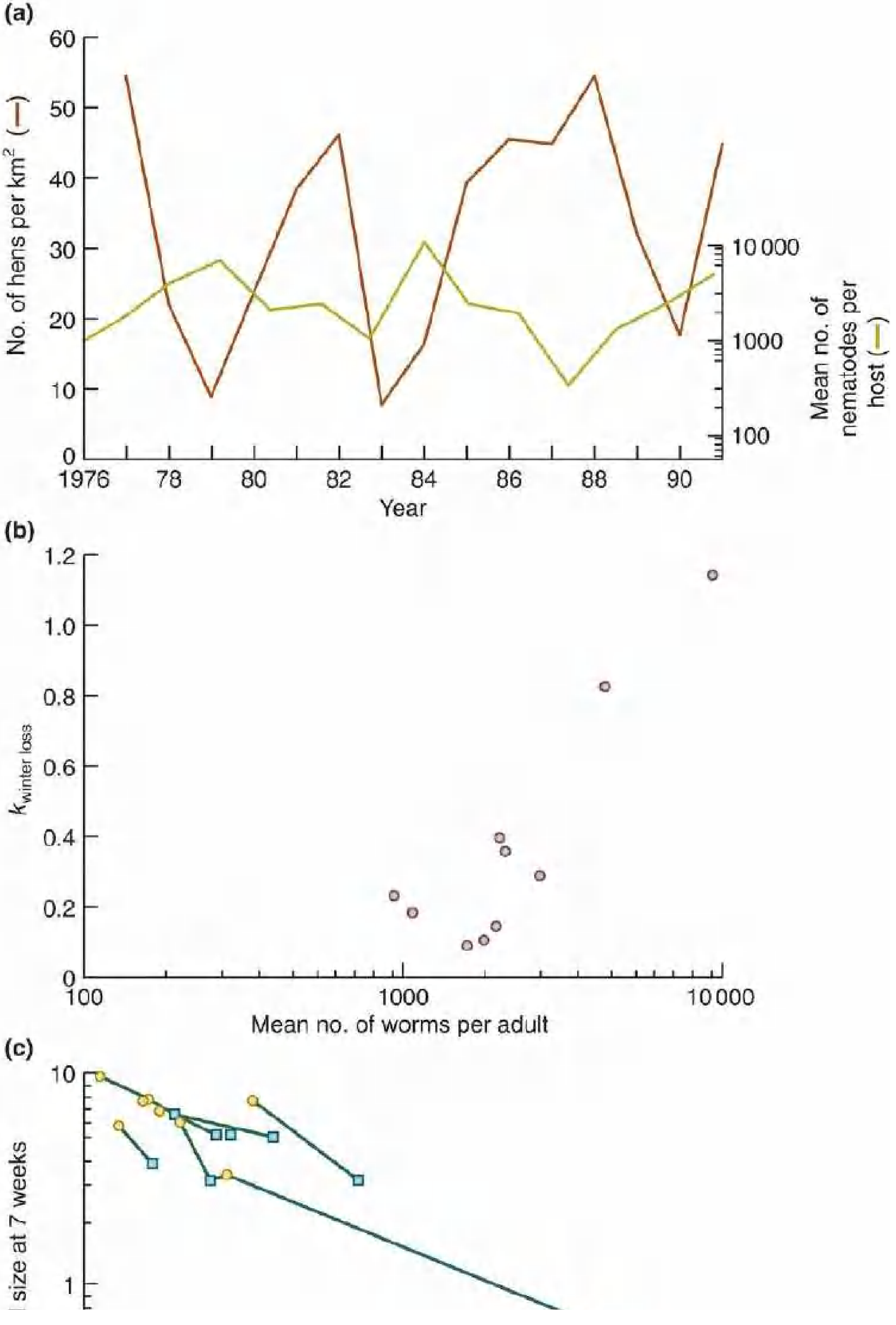


Figure 12.33 Anther smut has a negative effect on the population growth rate of white campion. The year-on-year rate of growth of populations of *Silene latifolia*, in Virginia, USA between 1989 and 1993 was lower when the prevalence of anther smut infection in the population was higher. Growth rate was measured as the ratio of the natural logs of population sizes in successive years, so zero represents an unchanging population. At each prevalence level, there are three dots representing populations grouped into three size classes (8–15, 16–31, 32–63) and dot sizes reflect the sample sizes in each ($n = 3$–239).

Source: After Bernasconi *et al.* (2009).

12.8.1 Red grouse and nematodes

Attempts to demonstrate an effect of infection on the *pattern* of dynamics are especially rare. One example, though, comes from studies of red grouse, *Lagopus lagopus scoticus* – of interest both because it is a ‘game’ bird, and hence the focus of an industry in which British landowners charge for the right to shoot it, and also because it is a species that often, though not always, exhibits regular cycles of abundance (Figure 12.34a). The underlying cause of these cycles has been disputed (Hudson *et al.*, 1998 ; Lambin *et al.*, 1999 ; Mougeot *et al.*, 2003a ; see also Section 14.6.1), but one mechanism receiving strong support has been the influence of the parasitic nematode, *Trichostrongylus tenuis*, occupying the birds’ gut caeca and reducing survival and breeding production (Figure 12.34b, c).


(a)
60
50
40
30
20
10
0
No. of hens per km^2 (—)
10 000
1000
100
Mean no. of nematodes per host (—)
1976
78
80
82
84
86
88
90
Year
(b)
1.2
1.0
0.8
0.6
0.4
0.2
0
$k_{\text{winter loss}}$
100
1000
10 000
Mean no. of worms per adult
(c)
10
1
size at 7 weeks

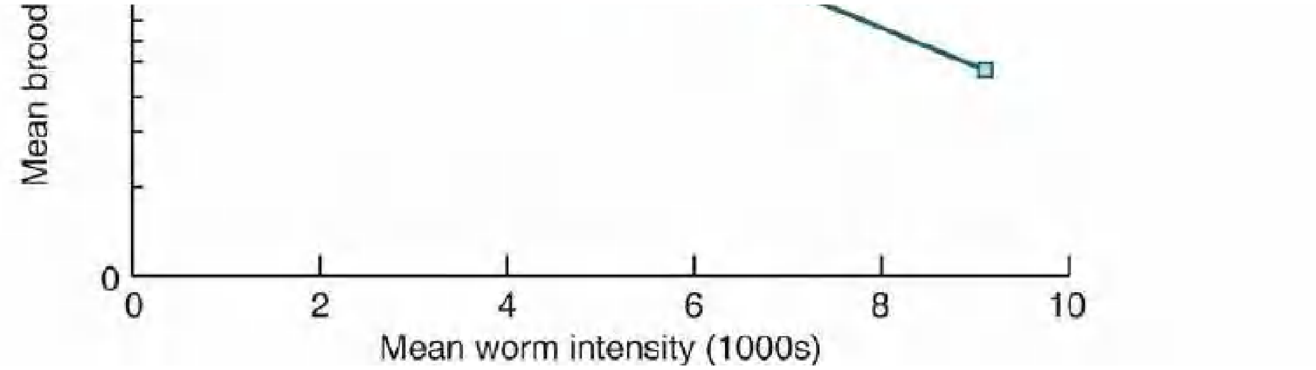
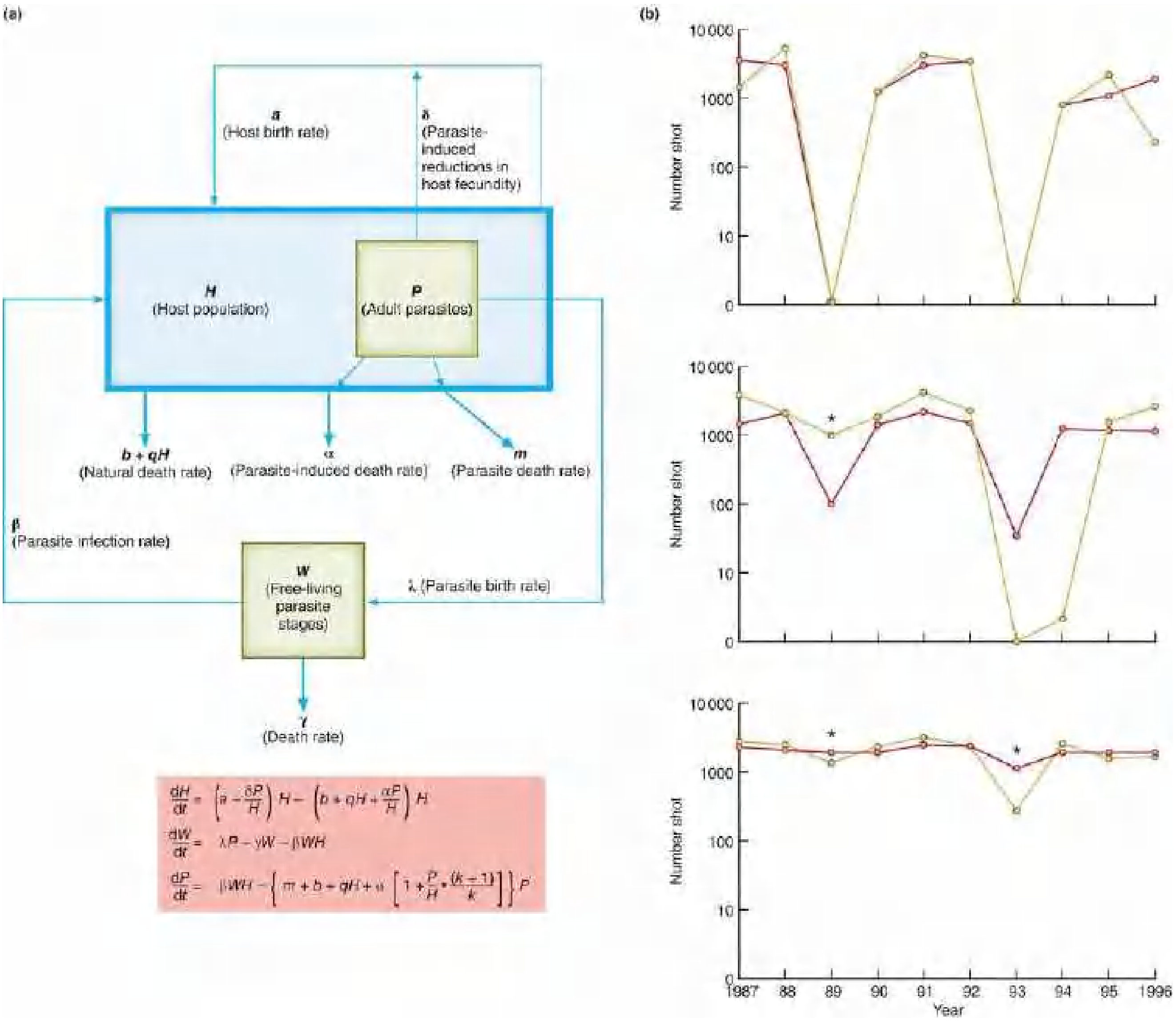


Figure 12.34 A nematode has detrimental effects on the survival and fecundity of red grouse. (a) Regular cycles in the abundance (breeding hens per km^2) of red grouse and the mean number of nematodes, *Trichostrongylus tenuis*, per host at Gunnerside, UK. (b) *Trichostrongylus tenuis* reduces survival in the red grouse: over 10 years (1980–89) winter loss (measured as a *k*-value) increased significantly ($P < 0.05$) with the mean number of worms per adult. (c) *T. tenuis* reduces fecundity in the red grouse: in each of eight years, females treated with a drug to kill nematodes (yellow circles; representing mean values) had fewer worms and larger brood sizes (at seven weeks) than untreated females (blue squares). Lines join pairs of points from each year.

Source: (a) After Dobson & Hudson (1992) and Hudson *et al.* (1992b). (b, c) After Hudson *et al.* (1992b).

A model for this type of host–macroparasite interaction is described in Figure 12.35a. Its analysis suggests that regular cycles both of host abundance and of mean number of parasites per host will be generated if:

$$\delta > \alpha k. \tag{12.11}$$



[**Figure 12.35**](#) **Treatment against its nematode parasite reduces the amplitude of fluctuations in the abundance of red grouse.** (a) Flow diagram (above) depicting the dynamics of a macroparasitic infection such as the nematode *Trichostrongylus tenuis* in red grouse, where the parasite has free-living infective stages; and (below) the model equations describing those dynamics. Taking the equations in order, they describe: (i) hosts (H) increasing as a result of (density-independent) births (which, however, are reduced at a rate dependent on the average number of parasites per host, P/H), but decreasing as a result of deaths – both natural (density-dependent) and induced by the parasite (again dependent on P/H); (ii) free-living parasite stages (W) increasing as a result of being produced by parasites in infected hosts, but decreasing both as a result of death and by being consumed by hosts; and (iii) parasites within hosts (P) increasing as a result of being consumed by hosts, but decreasing as a result of their own death within hosts, of the natural death of the hosts themselves and of disease-induced death of the hosts. This final term is dependent on the distribution of parasites amongst hosts – here assumed to follow a negative binomial distribution, parameter k, accounting for the term in square brackets. (b) Population changes of red grouse, as represented through bag records in two control sites (above), two populations with a single treatment each against nematodes (middle), and two populations with two treatments each (below). Asterisks represent the years of treatment, when worm burdens in adult grouse were reduced by an anthelmintic.

Source: (a) After Anderson & May ([1978](#)) and Dobson & Hudson ([1992](#)). (b) After Hudson *et al.* ([1998](#)).

Here, δ is the parasite-induced reduction in host fecundity (relatively delayed density dependence: destabilising), α is the parasite-induced host death rate (relatively direct density dependence: stabilising), and k is the 'aggregation parameter' for the (assumed) negative binomial distribution of parasites amongst hosts. Cycles arise when the destabilising effects of reduced fecundity outweigh the stabilising effects of both increased mortality and the aggregation of parasites. Data from a cyclic study population in the north of England indicated that this condition was indeed satisfied there. Grouse populations that fail to show regular cycles, or show them only sporadically, are often those in which the nematode cannot properly establish (Dobson & Hudson, 1992 ; Hudson *et al*., 1992b).

Such results from models are supportive of a role for the parasites in grouse cycles. But they fall short of the type of 'proof' that can come from a controlled experiment. A simple modification of the model in Figure 12.35a, however, predicted that if a sufficient proportion (20%) of the population were treated for their nematodes with an anthelmintic, then the cycles would die out. This set the scene for a field-scale experimental manipulation designed to test the parasite's role (Hudson *et al*., 1998). In two populations, the grouse were treated with anthelmintics in the years when two successive population crashes were expected. In two other populations, the grouse were treated only in the expected year of one crash. Two further populations were monitored as unmanipulated controls. Grouse abundance was measured as 'bag records': the number of grouse shot. It is clear that the anthelmintic changed the dynamics (Figure 12.35b), and it is therefore equally clear that the parasites normally have an effect: that is, the parasites affected host dynamics.

The precise nature of that effect, however, remains a matter of some controversy. Those who carried out the study believed that the experiment demonstrated that the parasites were 'necessary and sufficient' to generate the observed host cycles. Others felt that rather less had been fully demonstrated, suggesting for example that the cycles may have been reduced in amplitude rather than eliminated, especially as the very low numbers normally 'observed' in a trough are exaggerated (1 on their logarithmic scale equates to 'zero abundance') because there is no shooting when abundance is low (Lambin *et al*., 1999 ; Tompkins & Begon, 1999). Indeed, a further field experiment failed to support a role for the parasite acting alone in generating the cycles, indicating that cycles could occur in red grouse populations without any measurable influence of parasites (Redpath *et al*., 2006). On the other hand, such controversy should not be seen as detracting from the importance of field-scale experiments in investigating the roles of parasites in the dynamics of host populations – nor, indeed, the roles of other factors. For example, a subsequent field manipulation supported the alternative hypothesis that red grouse cycles are the result of density-dependent changes in aggressiveness and the spacing behaviour of males (Mougeot *et al*., 2003b). This system is examined again in a wider discussion of population cycles in Section 14.6.1.

12.8.2 An integral role for parasites?

Taken together, the red grouse studies show us that the nematode parasite can play a key role in the dynamics of its host, but that even when it does so, it acts in concert with other factors (Redpath *et al*., 2006). We see something similar, and again as a result of field experimental manipulations, in a study of white-footed mice, *Peromyscus leucopus*, and deer mice, *P. maniculatus*, in western Virginia, USA (Pedersen & Grieves, 2008). The abundances of these species fluctuate, but typically rise with the appearance in the autumn of their main food source, acorns of white and red oak (*Quercus alba* and *Q. rubra*), especially dramatically following mast years (see Section 10.2.4). They then crash over the following winter. The question that was asked was: 'What are the respective roles played by food shortage and intestinal nematodes in these crashes?' To address this, field populations were either given supplementary food (to simulate a mast), or given the drug ivermectin, which kills the nematodes (and also reduces abundances of ectoparasitic ticks, fleas and parasitic flies), or given both, or left as unmanipulated controls. Both food supplementation and parasite treatment tended to lessen the depth of the crash, but it is when both were given that a really profound effect was seen (Figure 12.36). This suggests in turn that neither food shortage nor the parasites alone are responsible for the crashes that are observed, but rather the interaction between the two.

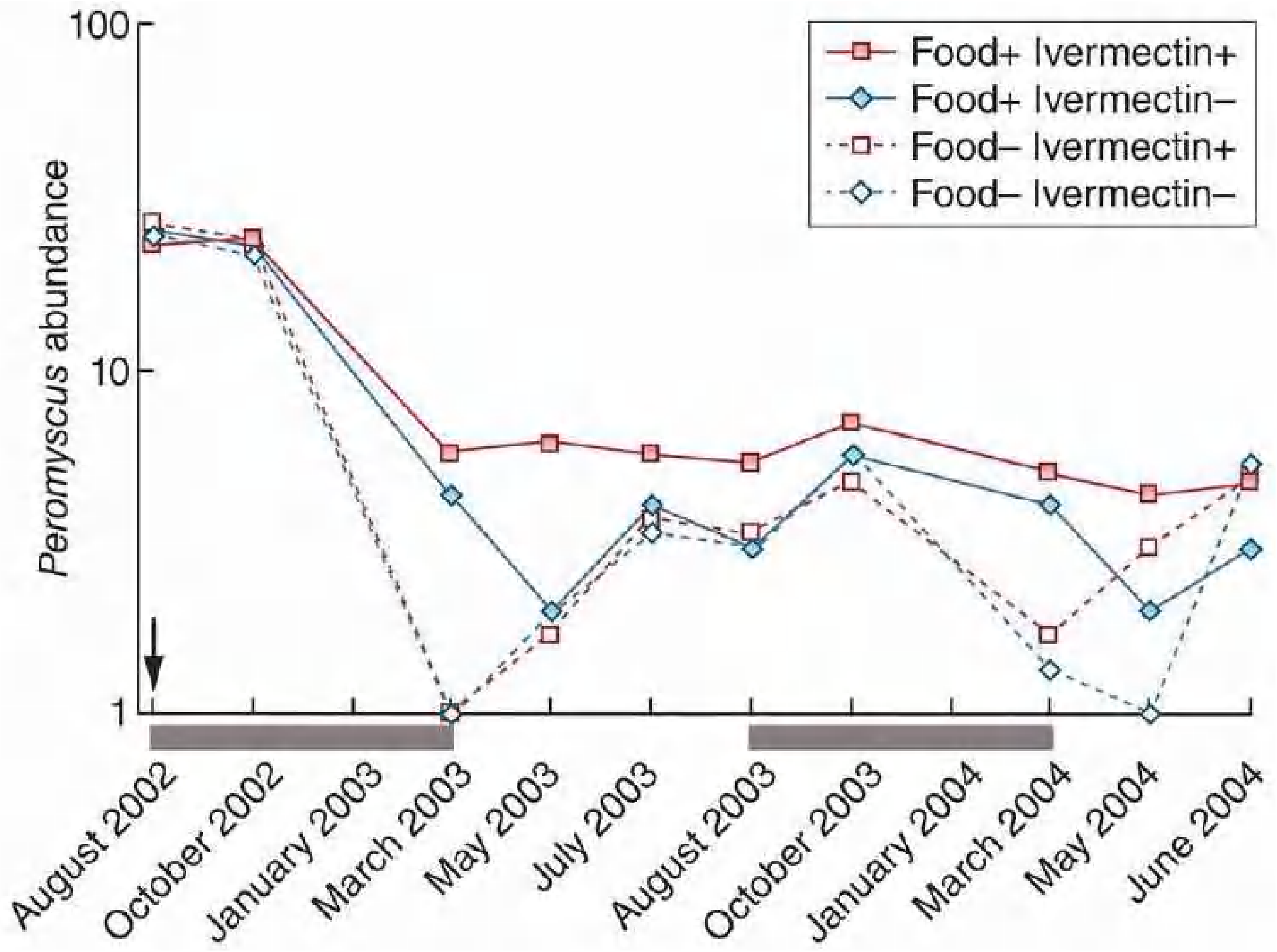


Figure 12.36 Food and parasites combine to determine the abundance of white-footed mice. Combined abundance (mean number per grid) of two species of *Peromyscus*, *P. leucopus* and *P. maniculatus*, on experimental grids subjected to four types of treatment: (i) given supplementary food, (ii) treated with ivermectin to kill gut parasites and ectoparasites, (iii) given both, and (iv) left as unmanipulated controls. Ivermectin treatment was initiated in August 2002, as indicated by the vertical black arrow, and supplementary food supplied during the period indicated by the grey bars. There were significant effects of food, ivermectin and an interaction between the two ($P < 0.05$ in all cases).

Source: After Pedersen & Grieves (2008).

Indeed, the ivermectin treatment, by its nature, did not point to an effect of one key parasite, but of intestinal nematodes collectively, and of other gut and ectoparasites. We saw, too, in Section 12.6, that parasites often act on their hosts by making them more vulnerable to predation or competition. And we saw in Section 12.3.8 how infection with one parasite may increase the chances of infection with other parasites. A picture emerges, therefore, of individual parasite species affecting the fitness of individual hosts, and ultimately the sizes of host populations, not alone, but through their interaction with other parasites and other factors. We can even think of there often being a vicious circle, through which an initial infection may make an individual only marginally less fit than its uninfected counterparts, but more vulnerable to competition or other parasites, further increasing its vulnerability to infection and its risk of immediate death at the hands of a predator, to infection and successively acting on small initial differences and turning them ultimately into ecologically important ones (Beldomenico & Begon, 2010). We might conclude from this that parasites play a less important role in determining the fitness and population dynamics of their hosts than, say, predators or competitors because, alone, any such role is often difficult to detect. But we might equally conclude that the effects of predators and competitors are often not as simple as they seem, and that parasites often play an integral role in determining not only who suffers the effects of predation or competition but how many do. We return in Chapter 14 to an examination of how a multiplicity of factors may combine to determine abundance.

Chapter 13

Facilitation: Mutualism and Commensalism

13.1 Introduction: facilitation, mutualists and commensals

So far, we have considered interactions between pairs of species where at least one of their populations suffers negative consequences – prey or hosts harmed by their predators or parasites ('+ –'), or competitors adversely affecting each other ('– –'). The term *facilitation*, on the other hand, encompasses interactions where at least one of the species gains a net benefit and neither is harmed (Bruno *et al.*, 2003). If both species benefit ('+ +') the interaction is a *mutualism*. If one benefits and the other is unaffected ('+ 0') the interaction is a *commensalism*.

Corals that form reefs provide good examples of commensalism, their structure making habitat available for a rich diversity of other inhabitants that have no influence, positive or negative, on the coral. Mutualisms, by contrast, usually involve the direct exchange between partners of goods or services (e.g. food, defence or transport) and typically result in the acquisition of novel capabilities and niche expansion for at least one of the species (Section 2.2; Moran, 2007). An extreme example is provided by siboglinid polychaetes of the genus *Osedax*. These worms lack a functional mouth and gut but harbour heterotrophic bacterial symbionts in root-like structures that allow them to obtain nutrients from the bones of whale carcasses (Goffredi *et al.*, 2005). Arguably, the ultimate example, and in a similar vein, is the acquisition by a 'proto-eukaryote' of aerobic heterotrophic bacteria that later evolved into mitochondria and provided the basis for all eukaryote diversification (Williams *et al.*, 2013).

mutualists ... and commensals ... need not be symbionts

No species lives in isolation, but often the association with other species is especially close: for many organisms, the habitat they occupy is an individual of another species. Mutualistic nitrogen-fixing bacteria live in nodules on the roots of leguminous plants, and, as we saw in Chapter 12, many parasites live within the body cavities or even the cells of their hosts. *Symbiosis* ('living together') is the term for such close physical associations between species, in which a 'symbiont' occupies a habitat provided by the body of a 'host'. Mutualists are sometimes symbionts, but mutualists need not be symbionts. For example, plants often gain dispersal of their seeds by offering a reward to birds or mammals in the form of edible fleshy fruits, and many plants assure effective pollination by offering a resource of nectar to animals that visit their flowers. These are mutualistic interactions but they are not symbioses. Likewise, commensalisms may or may not be symbiotic: sucker fish in the family Echeneidae are adapted to hitch a ride on sharks and are symbiotic commensals, in contrast to the many commensal species inhabiting reefs that have no intimate physical connection to the coral species that built them.

mutualism is typically reciprocal exploitation *not* a cosy partnership

Commensalistic and mutualistic interactions should not be viewed simply as conflict-free relationships from which nothing but good things flow for one or both partners. Rather, current evolutionary thinking views mutualisms as cases of reciprocal exploitation where, nonetheless, each partner is a *net* beneficiary. Moreover, among symbiotic relationships, mutualism, commensalism and parasitism can be regarded as occurring along a continuum, with commensalism occupying the middle ground. In practice, symbiotic commensals such as sucker fish on sharks may be indistinguishable from parasites that cause their host comparatively low levels of harm, or from mutualists that exert such a heavy price for their service that the host is barely compensated by the benefits of the association (Leung & Poulin, 2008).

It would be mistaken to think of mutualism and commensalism as less common or ecologically less important than competition, predation or parasitism (Bruno *et al.*, 2003). Some animals and many plants provide habitat for their commensal partners. And almost all the plants that dominate grasslands, heaths and forests have roots that have an intimate mutualistic association with fungi, most corals depend on the unicellular algae within their cells, many flowering plants need their insect pollinators, and many animals carry communities of microorganisms within their guts that they require for effective digestion. In the next section we consider commensalisms in more detail. Then we deal with examples of mutualism before finally considering mathematical models of commensalism and mutualism.

13.2 Commensalisms

'habitat' commensalisms: ecosystem engineers provide habitat for other species

There are many ‘interactions’ between two species in which the first provides a habitat for the second, but there is no real suspicion that the first either benefits or suffers in any measurable way as a consequence. Trees, for example, provide habitats for the many species of birds, bats and climbing and scrambling animals that are absent from treeless environments. Lichens and mosses develop on tree trunks, and climbing plants such as ivy, vines and figs, though they root in the ground, use tree trunks as support to extend their foliage up into a forest canopy. Trees are therefore good examples of what have been called ecological or ecosystem ‘engineers’ (Jones *et al.*, 1994). By their very presence, they create, modify or maintain habitats for others (see also Section 16.4). In marine communities, the solid surfaces of larger organisms are just as important contributors to biodiversity. Kelps and other seaweeds normally grow only where they can be anchored on rocks, but their fronds are colonised in turn by filamentous algae, tube-forming worms (*Spirorbis*) and modular animals such as hydroids and bryozoans that depend on seaweeds for anchorage and access to resources in the moving waters of the sea. Turning to freshwater habitats, many crayfish species build burrows as refuges that are particularly important in seasonal environments that may dry up. *Gramastacus insolitus*, a threatened species endemic to shallow swamps and stream margins in southern Australia, is a very small crayfish (<4 cm total length) that does not build burrows. It only occurs as a commensal associated with larger species (*G. falcata* and *Cherax destructor*) that build deep burrows connected to the ground water, the smaller species using cracks and depressions in the burrows as aestivation sites in dry seasons. *G. insolitus* is part of an assemblage of animals that take advantage of this ‘engineered' habitat (Johnston & Robson, 2009).

One of the best studied ecosystem engineers are the ‘cushion’ plants – species from diverse families that grow as spreading mats (no more than a few inches high), with relatively large and deep tap roots, and life histories adapted to slow growth in harsh alpine and arctic environments around the world. The low structure and compact architecture of cushion plants have been shown to attenuate severe environmental conditions, buffering both low and high temperatures at the soil surface, increasing humidity, and enhancing soil moisture, organic matter and nutrient concentrations, with benefits for both the cushion plants themselves and a myriad of commensalistic plant species that associate with them (Cavieres *et al.*, 2014). Because cushion plants can enhance plant species richness in harsh environments, they have earned the epithet of ‘nurse’ species.

‘feeding' commensalisms

Some large herbivores can dramatically change their landscape. Elephants transform savannah into open shrubland, while hippos convert tall grass into grazing lawns, with concomitant shifts towards other grazers that favour the new feeding regime. Feeding commensalisms among grazing mammals may also arise in two more subtle ways: one species making grass more accessible to another, usually smaller species, by reducing grass height and removing stems, or stimulating grass regrowth and enhancing nutritional quality. The evidence is stronger for the second case, with, for example, Thomson’s gazelles on the Serengeti plains clearly attracted to areas where prior grazing by wildebeest improved the quality of the grass sward (McNaughton, 1976).

Arsenault and Owen-Smith (2002) note, though, that while feeding facilitation is undoubtedly a structuring force in grazing ungulate communities, there is no strong evidence of this translating into population abundance. They argue that there may be a seasonal trade-off between commensalism in the productive wet season, enhancing reproductive performance of the smaller species, but competition via reduced forage in the dry season, increasing mortality. It is clear that the consequences of any commensalistic (or indeed mutualistic) relationship need to be viewed in the broader context of a web of population interactions, a theme that recurs throughout this chapter.

trade-offs between commensalism and competition

Trade-offs between commensalism and competition can be expected to be particularly marked in interactions between species occupying the same trophic level, and this applies both for herbivores in the Serengeti (above) and more particularly for interactions among plants. Thus, the well-established process whereby some early successional pioneer plants, including some cushion plants (discussed previously), facilitate later successional species (for example by enhancing soil nutrient concentrations; Section 18.4) can be viewed as temporarily commensalistic in its earliest stages, with the pioneer plants not suffering detectably through competition with the later commensals, until the latter become large enough and vigorous enough to outcompete the former. On the other hand, in some plant communities, for example in high mountain or arctic sites, the severity of the abiotic environment may prevent succession progressing to a stage where competition kicks in, and the facilitative benefits of pioneer cushion plants remain relevant in communities that are stable in the long term. The balance between commensalism and competition can also vary spatially, with the competitive effects of some cushion plants outweighing their commensalistic effects in less stressful environments (Cavieres *et al.*, 2014).

The physiological status of a cushion plant may also influence its facilitative effects. For example, cushions of *Arenaria tetraquetra* at high elevations in Spain’s Sierra Nevada mountains showed good physiological status (compact and large), produced higher soil water and organic matter content compared with open areas, and strongly facilitated many beneficiary species (in terms of numbers of individuals and of species within the cushions). On the other hand, physiological data at lower elevations indicated stressful abiotic conditions for *A. tetraquetra* (lower water availability). There they formed loose and small cushions and had reduced facilitative effects (Schöb *et al.*, 2013).

APPLICATION 13.1 Commensalism, restoration and intercropping agriculture

‘nurse’ plants and other forms of plant–plant facilitation

Against a background of rapid degradation of many ecosystems worldwide, a fuller understanding of commensalistic relationships could provide the basis for novel restoration techniques (Brooker *et al*., 2008). There are already established practices in terrestrial environments that have concentrated on the positive effects of 'nurse' plants, such as the use of naturally occurring shrubs to provide shade and improve water availability during reforestation of degraded, water-limited Mediterranean environments – an approach that runs counter to the previous reforestation practice of removing shrubs before replanting trees (see Application 18.2 and Figure 18.15). But other commensalistic processes could also be harnessed in restoration efforts, including augmentation of pollinator visits by locating attractive plants in the vicinity of less attractive commensal neighbours; bioremediation by metallicious nurse plants that accumulate lead or cadmium from the environment and increase the success of commensal neighbours in environments polluted by heavy metals (see Section 2.8); dispersion of fixed nitrogen to non-legume commensals from legumes, with their nitrogen-fixing, symbiotic bacteria (while the legumes are growing or decomposing); and the presence of unpalatable plants, reducing predation risk in neighbouring palatable commensals (Graff *et al*., 2007).

‘intercropping’

Intercropping is a farming practice, used in many subsistence or low-input agricultural systems, that involves two or more species (or genotypes) growing together and coexisting for a time. Brooker *et al*. (2016) argue that by allowing yield gains without increased inputs of fertiliser and pesticides, intercropping could be a route to delivering ‘sustainable agricultural intensification’. Thus, for example, crop production on acidic soils is often limited by phosphorus availability or aluminium toxicity. Roots of plants such as peanut, cowpea and sweet potato secrete organic acids and phosphatases into the rhizosphere, increasing soil phosphorus availability and protecting roots of beneficiary crops from aluminium toxicity. Likewise, crops in alkaline soils are often limited by the availability of phosphorus, iron, zinc, magnesium or copper. Crops tolerant of mildly alkaline soils, including brassicas, maize, beet and squash, acidify their rhizosphere and improve the mineral nutrition of beneficiary crops.

shift from facilitation to competition along environmental gradients

Early observations on the relative importance in plant communities of positive interactions (i.e. facilitation, mainly of the commensal variety) and negative interactions (competition) led to formulation of the stress-gradient hypothesis (Bertness & Callaway, 1994), which predicts that facilitation in communities increases and competition decreases with increasing environmental stress. Examples of facilitation in plants include those that ameliorate abiotic stresses, such as when one plant provides shade for another where light intensities would otherwise be too high, and those that ameliorate biotic effects, such as when the antiherbivore defences of one plant also protects other plants nearby. The idea of the stress-gradient hypothesis, simply, is that species are most likely to owe their continued existence to facilitation in environments in which they would fail to survive otherwise. In a meta-analysis of 727 published tests of the hypothesis from terrestrial, marine and freshwater ecosystems, He *et al.* (2013) defined stress as any abiotic factor (e.g. salinity, cold, wind), resource factor (e.g. water, nutrients, light) or biotic factor (e.g. grazing pressure) that can reduce survival, growth and/or reproduction, and thus fitness. For each study in the meta-analysis, stress was categorised as low or high, and the relative importance of facilitation and competition for target species was defined in terms of performance in treatments with and without neighbours of a different species. Effect sizes were calculated using the log of the ratio of the odds of survival with and without neighbours, and, for growth and reproduction, using Hedges g^* statistic, which measures the standardised mean difference in performance with and without neighbours (see He *et al.* (2013) for more details). Positive effect sizes indicate facilitation by neighbours; negative effect sizes indicate competition by neighbours.

Their results show that plant interactions do in general change with increased environmental stress, and always in the direction of a shift towards facilitation (typical for survival responses) or a decline in competition (typical for growth and reproduction responses) (Figure 13.1a–c). Their findings were generally consistent across stress types (abiotic, resource, biotic), plant growth forms (herb, grass, shrub, tree), life histories (annuals, perennials) and origins (native or invasive). They were also consistent in relation to climate and ecosystem type. Across ecosystems, the most pronounced shifts in survival with increasing stress occurred in coastal/marine, wetland and grassland/savannah systems, while significant changes in growth were recorded for all ecosystems. Thus, it turns out that a pattern of shift towards facilitation with increasing stress is not restricted to climates that are generally considered more stressful (cold, arid) but is also seen in moister and warmer regions (temperate and tropical) (Figure 13.1d–g).

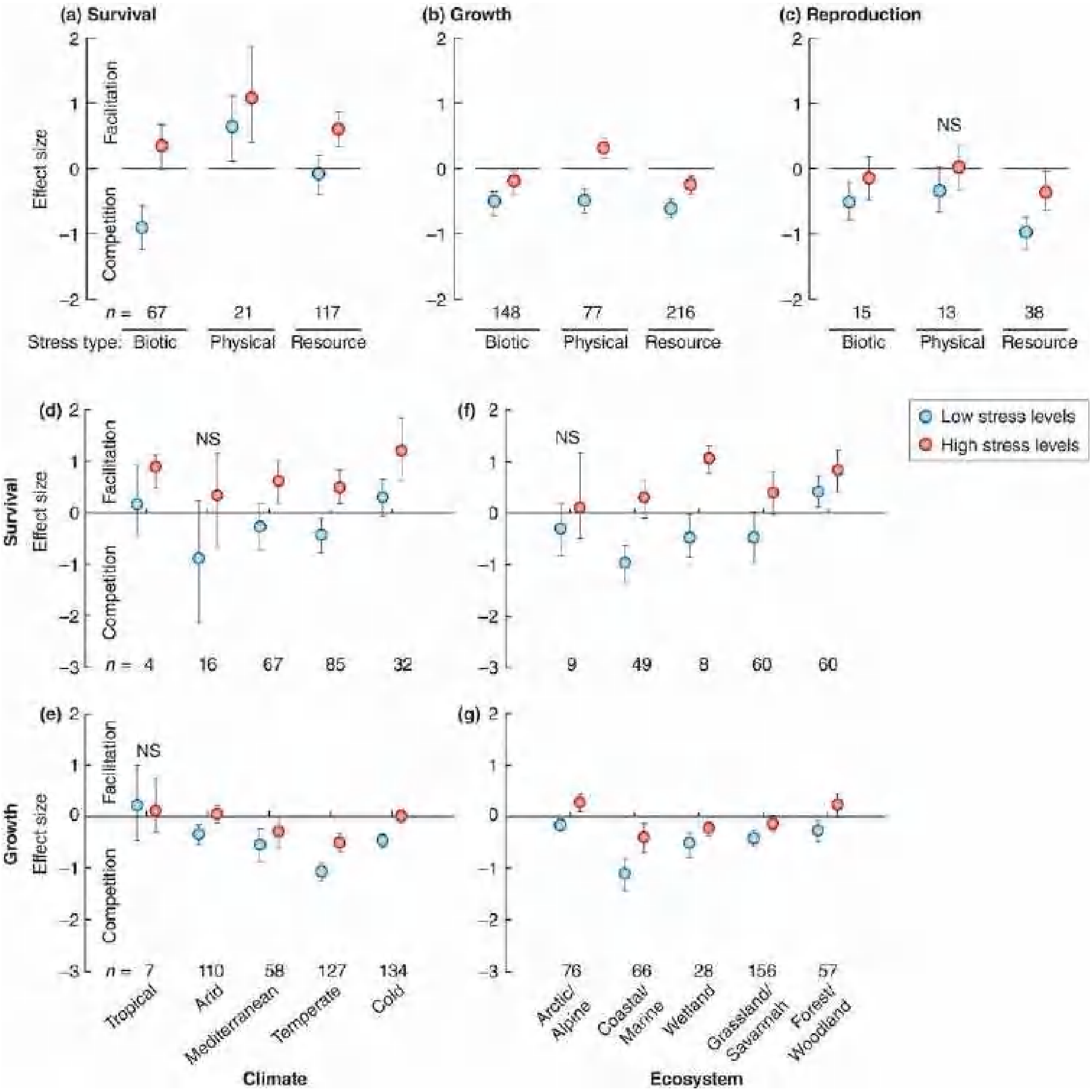


Figure 13.1 Plant interactions shift from competitive to facilitative in more stressed environments. Plant interaction effect sizes (medians with 95% CI) at low (blue circles) and high (red circles) stress levels sorted by stress type (biotic, physical, resource): (a) survival, (b) growth, and (c) reproduction. Positive values for effect sizes (comparing performance with and without neighbours) indicate facilitation, negative values indicate competition. Effect sizes for separate climate zones: (d) survival and (e) growth. Effect sizes for separate ecosystem types: (f) survival and (g) growth. Number of studies in each category are indicated. All differences are statistically significant unless indicated (NS, not significant).

Source: From He *et al.* (2013).

He *et al.* (2013) also emphasise the need for community-level studies and assessment of interactions among multiple species to understand better the consequences of facilitative interactions for the structure of whole communities. We pick up the stress-gradient hypothesis again when we discuss this and other population interactions in Chapter 16.

evolutionary aspects of commensalism

Much evolutionary research has focused on adaptation by natural selection to changing abiotic environments. However, species also create their own environment and modify that of their commensals, and Soliveres *et al*. (2015) stress the importance of studies testing patterns of evolution in response to the presence of neighbouring species. On an 'ecological' time scale, plant–plant commensalisms may, for example, affect the evolutionary consequences of grazing history. Thus, the 'naïve' (without grazing history) phenotype of the palatable annual herb *Persicaria longiseta* strongly benefits from commensalistic protection against browsing, while the grazing-adapted phenotype of the same species does not (Suzuki & Suzuki, 2012). Taking a much longer-term view, Valiente-Banuet *et al*. (2006) have asserted that recent plant lineages (quaternary; the last 2.58 million years) may have nursed older, drought-maladapted ones (tertiary; 65–2.58 million years ago) through the drier climatic conditions found during the quaternary period.

13.3 Mutualistic protectors – a behavioural association

When compared with commensalistic relationships, the evolutionary histories of mutualists are inevitably much more intimately tied together. We now turn to mutualisms, organised as a progression, starting here with those where no intimate symbiosis is involved and the association is largely behavioural: that is, each partner behaves in a manner that confers a net benefit on the other. By Section 13.6, when we discuss mutualisms between animals and the microbiota living in their guts, we will have moved on to closer, symbiotic associations (one partner living within the other), and in later sections we examine still more intimate symbioses in which one partner enters between or within another's cells.

13.3.1 Cleaners and clients

cleaner fish and their client fish

'Cleaner' fish, of which many species in several families have been recognised, remove ectoparasites, bacteria and necrotic tissue from the body surface of 'client' fish. Indeed, the cleaners often hold territories with 'cleaning stations' that their clients visit – and visit more often when they carry many parasites. The cleaners gain a food source and the clients are protected from infection. In fact, it has not always proved easy to establish that the clients benefit, but in experiments off Lizard Island on Australia's Great Barrier Reef, Grutter (1999) was able to do this for the cleaner fish *Labroides dimidiatus*, which eats parasitic gnathiid isopods from its client fish, *Hemigymnus melapterus*. Clients had significantly (3.8 times) more parasites 12 days after cleaners were excluded from caged enclosures (Figure 13.2a); but even in the short term (up to one day), although removing cleaners, which only feed during daylight, had no effect when a check was made at dawn (Figure 13.2b), this led to there being significantly (4.5 times) more parasites following a further day's feeding (Figure 13.2c).

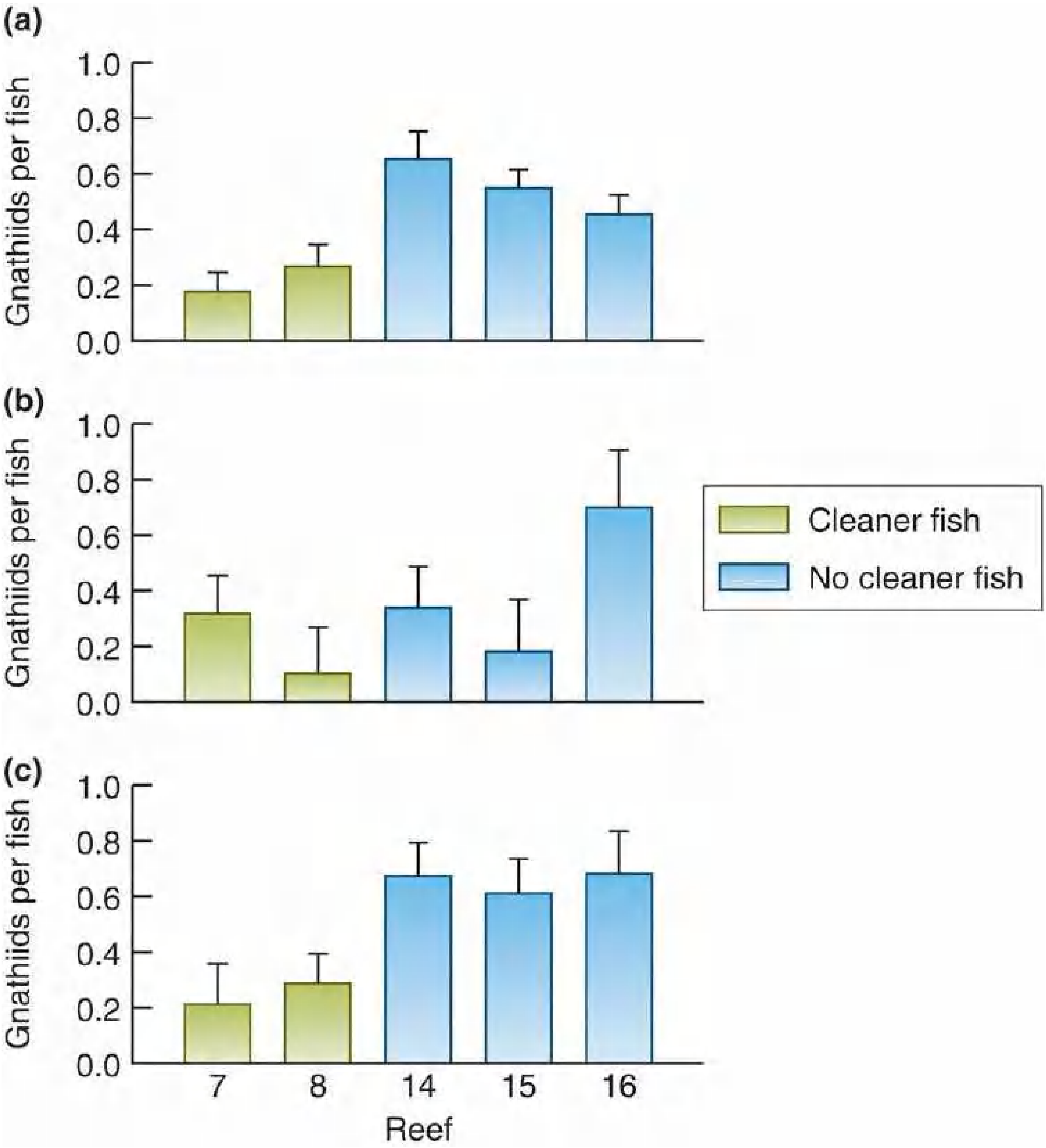


Figure 13.2 Cleaner fish really do clean their clients. The mean number of gnathiid parasites per client (*Hemigymnus melapterus*) at five reefs, from three of which (14, 15 and 16) the cleaners (*Labroides dimidiatus*) were experimentally removed. (a) In a 'long-term' experiment, clients without cleaners had more parasites after 12 days ($F = 7.6$, $P = 0.02$). (b) In a 'short-term' experiment, at dawn after 12 hours without cleaners, clients did not have significantly more parasites ($F = 1.8$, $P = 0.21$), presumably because cleaners do not feed at night, but (c) after a further 12 hours of daylight the difference was significant ($F = 11.6$, $P = 0.04$). Bars represent SEs.

Source: After Grutter (1999).

cleaner fish and the motivation to 'cheat'

However, while cleaner wrasses such as *Labroides* spp. readily eat ectoparasites such as gnathiid isopods and monogeneans from the skin of their clients, Grutter and Bshary (2003) found that their preferred food is actually client mucus and tissue, which are costly for the client to produce. Thus, if the cleaner fish actually fed according to their preference, they might become ‘cheaters' that impose a cost without providing much of a service, shifting the relationship along the spectrum from mutualism toward parasitism (Leung & Poulin, 2008). Indeed, Cheney and Côté (2005) found that in locations where clients had more ectoparasites, the cleaners removed less client mucus and tissue, and suggest that the availability of ectoparasites for cleaners to consume is the key to whether the association leans toward mutualism or parasitism.

crayfish and ‘cleaner worms' – another continuum from parasitism to mutualism

Cleaning symbioses also occur between freshwater crayfish and branchiobdellidan worms throughout Europe, North America and Asia. By feeding on accumulations of sediment and biofilm, particularly on respiratory surfaces, the worms can increase growth and survivorship of their crayfish hosts. But here too, the worms also come at a cost, sometimes consuming host tissue. When symbiosis decreases host fitness, host defences should evolve to limit colonisation of symbionts, whereas when symbiosis increases host fitness, those barriers to colonisation should disappear. The interaction between host and symbiont can be expected to shift along the continuum from parasitism to mutualism as the balance of benefits and costs changes in response to changing context, and Skelton *et al.* (2016) argued that young crayfish would not be likely to benefit much from a worm's cleaning services because of rapid growth and frequent moulting (which rids the crayfish of any accumulation of sediment and biofilm), whereas older crayfish should benefit from cleaning because they moult infrequently. Their prediction was confirmed in experiments where individual worms were introduced to individual crayfish (*Cambarus sciotensis*) and grooming responses were recorded. Young crayfish always groomed and were able to remove all but the smallest species of worm (*Cambarincola fallax* and *Pterodrilus alcicornis*). Older crayfish (carapace length >27–30 mm) never groomed, and consequently their bodies harboured a greater diversity of both large (*Cambarincola ingens* and *Ankyrodrilus koronaeus*) and small species of worms, but with the former predominating (Figure 13.3).

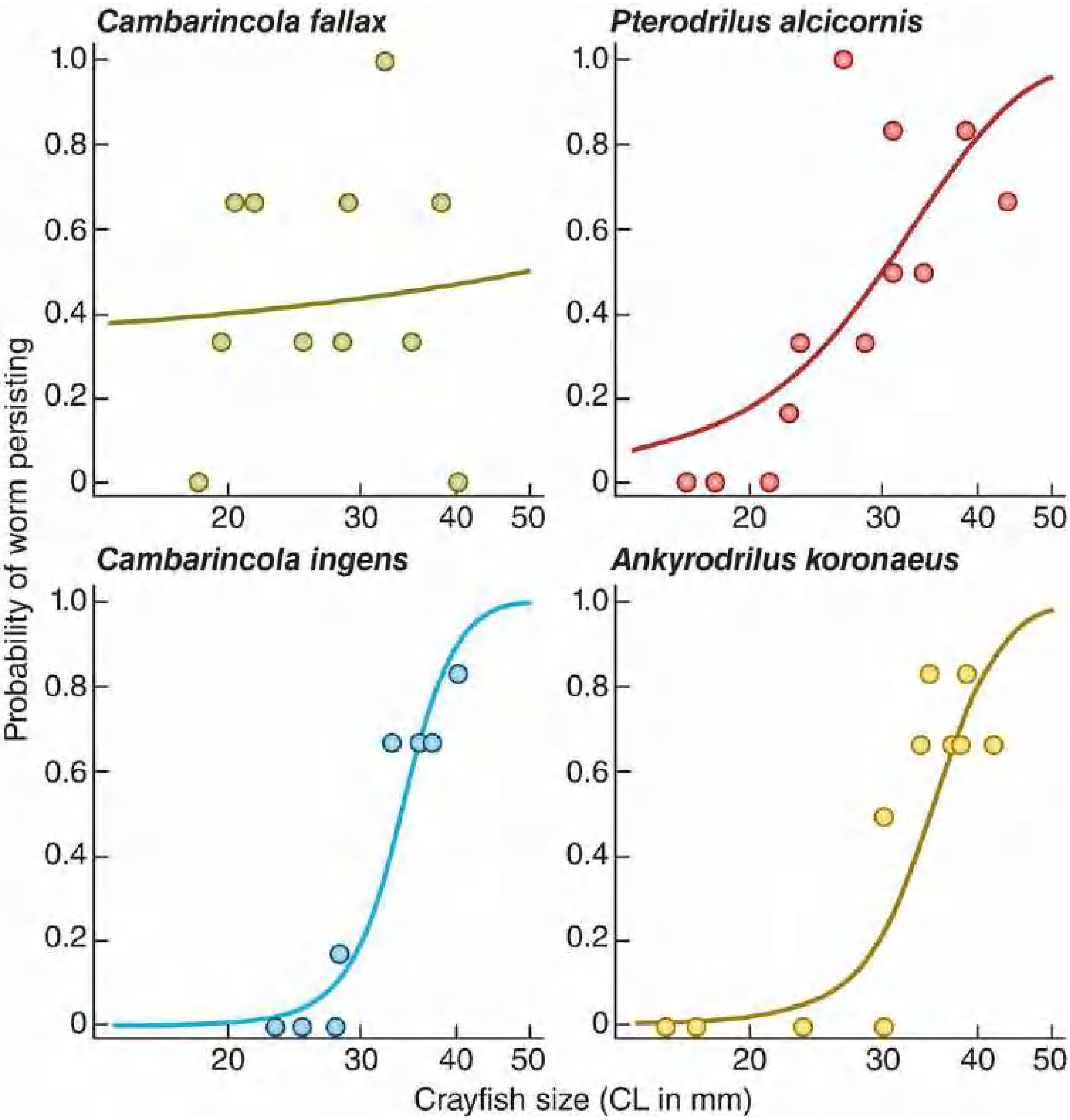


Figure 13.3 'Cleaner' worms benefit large but not small crayfish, which groom to remove them. The effect of crayfish size on the persistence on their bodies of four species of branchiobdellidan worms, in order of increasing worm species size. By their grooming responses, small crayfish were able to remove all but the smallest species (*Cambarincola fallax*). Large crayfish did not groom. CL, crayfish carapace length.

Source: From Skelton *et al*. (2016).

13.3.2 Ant–plant mutualisms

The idea that there are mutualistic relationships between plants and ants was put forward by Belt (1874) after observing the behaviour of aggressive ants on species of *Acacia* with swollen thorns in Central America. This relationship was later described more fully by Janzen (1967) for the Bull's horn acacia (*Acacia cornigera*) and its associated ant, *Pseudomyrmex ferruginea*. The plant bears hollow thorns that are used by the ants as nesting sites; its leaves have protein-rich 'Beltian bodies' at their tips that the ants collect and use for food; and it has sugar-secreting nectaries on its vegetative parts that also attract the ants. The ants, for their part, protect these small trees from competitors by actively snipping off shoots of other species and also protect the plant from herbivores – even large (vertebrate) herbivores may be deterred.

do the plants benefit?

In fact, ant–plant mutualisms appear to have evolved many times (even repeatedly in the same family of plants), and nectaries are present on the vegetative parts of plants of at least 39 families and in many communities throughout the world. The benefits to the plants are not always easy to establish, but in the case of the Amazonian rainforest tree *Duroia hirsuta*, they are at least two-fold. *D. hirsuta* exists naturally as large single-species stands that can last for as long as 800 years. They are known locally as Devil's gardens because of the traditional belief that they are tended by evil spirits that exclude other species. It seems, though, from a series of experiments, that those evils spirits are ants, especially *Myrmelachista schumanni*. On the one hand, these ants removed herbivores from the trees (Figure 13.4a), though they were not as effective in this as some other ant species (*Azteca* spp.). But in addition, *M. schumanni* attacked competing plants within a Devil's garden, injecting poison (formic acid) into their leaves, such that the leaves showed signs of dying within a day and had mostly been shed within five days (Figure 13.4b). The homes and nectaries the plants provide for their devils seem to be a price well worth paying.

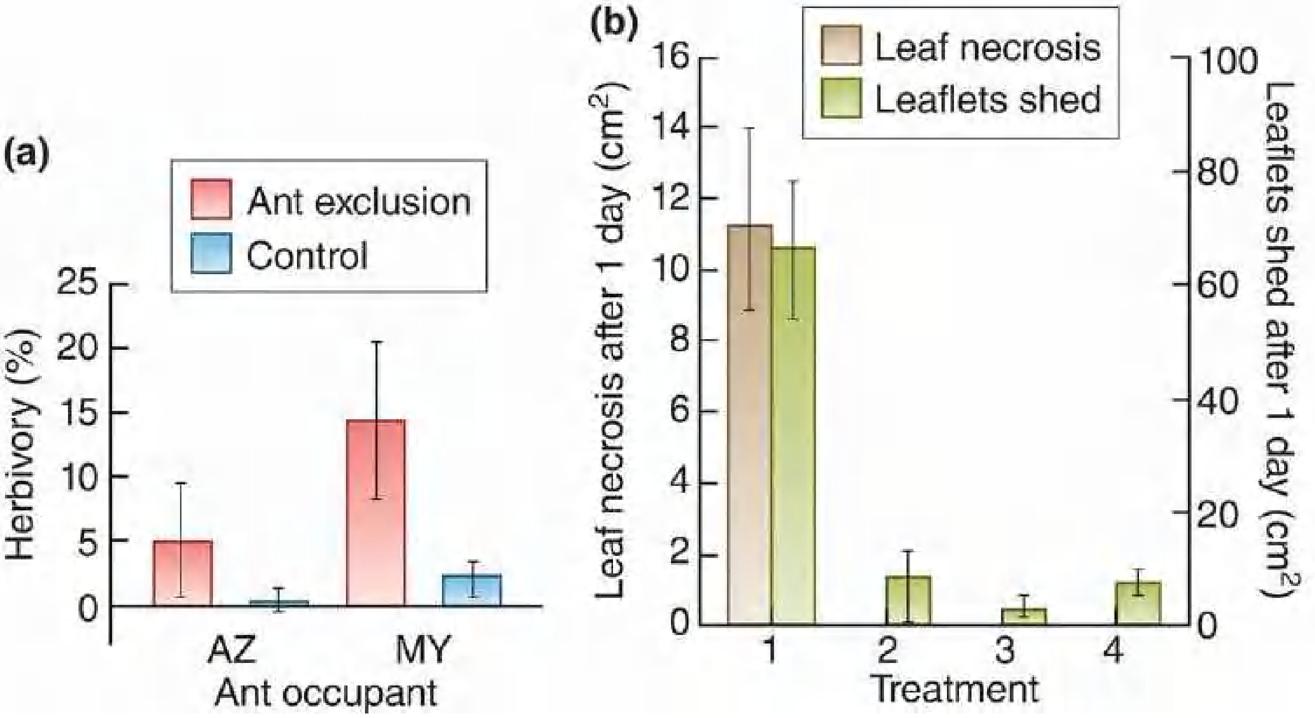


Figure 13.4 Ants provide their host plants with significant protection against herbivores and competing plants. (a) Levels of herbivory (percentage of leaf area loss) on *Duroia hirsuta* leaves, naturally occupied by ants or with ants artificially excluded. Ants: AZ, *Azteca* spp., MY, *Myrmelachista schumanni*. Bars are 95% CIs. (b) Saplings of the competitor plant *Clidemia heterophylla* subjected to different treatments: 1, planted among *D. hirsuta*, ants (*M. schumanni*) present; 2, planted among *D. hirsuta*, ants excluded; 3, planted away from *D. hirsuta*, ants present; 4, planted away from *D. hirsuta*, ants excluded. Bars are SEs. Treatments 2, 3 and 4 were all significantly different from treatment 1 ($P < 0.001$).

Source: (a) After Fredrickson (2005). (b) After Fredrickson *et al*. (2005).

competition amongst mutualistic ants

Mutualistic relationships, in this case between individual ant and plant species, should not, however, be viewed in isolation – a constant theme in this chapter. Palmer *et al.* (2000), for example, studied competition amongst four species of ant that have mutualistic relationships with *Acacia drepanolobium* trees in Kenya, nesting within the swollen thorns and feeding from the nectaries at the leaf bases. Experimentally staged conflicts and natural take-overs of plants both indicated a dominance hierarchy among the ant species. *Crematogaster sjostedti* was the most dominant, followed by *C. mimosae*, *C. nigriceps* and *Tetraponera penzigi*. Irrespective of which ant species had colonised a particular acacia tree, occupied trees tended to grow faster than unoccupied trees (Figure 13.5a). This confirmed the mutualistic nature of the interactions overall. But more subtly, changes in ant occupancy in the direction of the dominance hierarchy (take-over by a more dominant species) occurred on plants that grew faster than average, whereas changes in the opposite direction to the hierarchy occurred on plants that grew more slowly than average (Figure 13.5b).

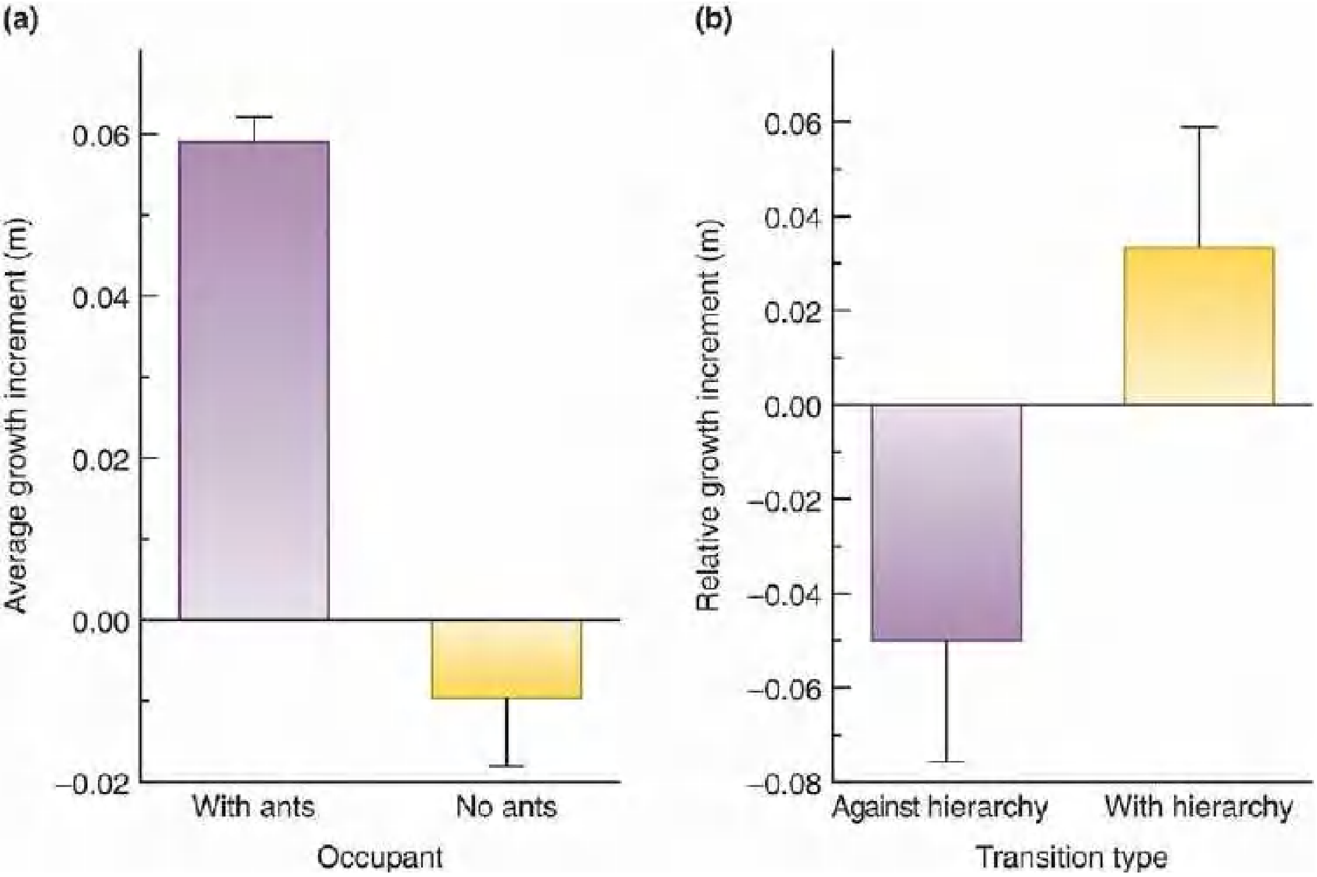


Figure 13.5 Ant species may compete in a competitive hierarchy for access to their plant mutualist, and the outcome is different on fast- and slow-growing trees. (a) As expected, average growth increment was significantly greater ($P < 0.0001$) for *Acacia drepanolobium* trees continually occupied by ants ($n = 651$) than for uninhabited trees ($n = 126$). ‘Continually occupied’ trees (‘With ants’) were occupied by ant colonies at both an initial survey and one six months later. Uninhabited trees (‘No ants’) were vacant at the time of both surveys. (b) Relative growth increments were significantly greater ($P < 0.05$) for trees undergoing transitions in ant occupancy in the direction of the ants’ competitive hierarchy ($n = 85$) than for those against the hierarchy ($n = 48$). Growth increment was determined relative to trees occupied by the same ant species when these ants were not displaced. Bars are SEs.

Source: After Palmer *et al.* (2000).

These data therefore suggest that take-overs are rather different on fast- and slow-growing trees, though the details remain speculative. It may be, for example, that trees that grow fastest also produce ant ‘rewards’ at the greatest rate and are actively chosen by the dominant ant species;

whereas slow-growing trees are more readily abandoned by dominant species, with their much greater demands for resources. Alternatively, competitively superior ant species may be able to detect and preferentially colonize faster growing trees. What is clear is that these mutualistic interactions are not cosy relationships between pairs of species that we can separate from a more tangled web of interactions. The costs and benefits accruing to the different partners vary in space and time, driving complex dynamics amongst the competing ant species that in turn determine the ultimate balance sheet for the acacias.

13.4 Farming mutualisms

13.4.1 Human agriculture

At least in terms of geographic extent, some of the most dramatic mutualisms are those of human agriculture. The numbers of individual plants of wheat, barley, oats, corn and rice, and the areas these crops occupy, vastly exceed what would have been present if they had not been brought into cultivation. The increase in human population since the time of hunter–gatherers is some measure of the reciprocal advantage to *Homo sapiens*. Even without doing the experiment, we can easily imagine the effect the extinction of humans would have on the world population of rice plants or the effect of the extinction of rice plants on the population of humans. The same comments apply to the domestication of cattle, sheep and other mammals.

Similar ‘farming’ mutualisms have developed in just three other animal groups (Schultz & Brady, 2008): bark beetles, termites and especially ants, where the farmers may protect individuals they exploit from competitors and predators and may even move or tend them.

13.4.2 Farming of insects by ants

farmed aphids: do they pay a price?

Ants farm many species of aphids (Hemiptera) in return for sugar-rich secretions of honeydew. The ‘flocks’ of aphids benefit through lower mortality rates caused by predators, showing increased feeding and excretion rates, and forming larger colonies. But it would be wrong, as ever, to imagine that this is a cosy relationship: the aphids are being manipulated – is there a price that they pay to be entered on the other side of the balance sheet (Stadler & Dixon, 1998)? This question has been addressed for colonies of the aphid *Tuberculatus quercicola* attended by the red wood ant *Formica yessensis* on the island of Hokkaido, northern Japan (Yao *et al*., 2000). As expected, in the presence of predators, aphid colonies survived significantly longer when attended by ants than when ants were excluded by smearing ant repellent at the base of the oak trees on which the aphids lived (Figure 13.6a). However, there *were* also costs for the aphids: in an environment from which predators were excluded, and the effects of ant attendance on aphids could thus be viewed in isolation, ant-attended aphids grew less well and were less fecund than those where ants as well as predators were excluded (Figure 13.6b).

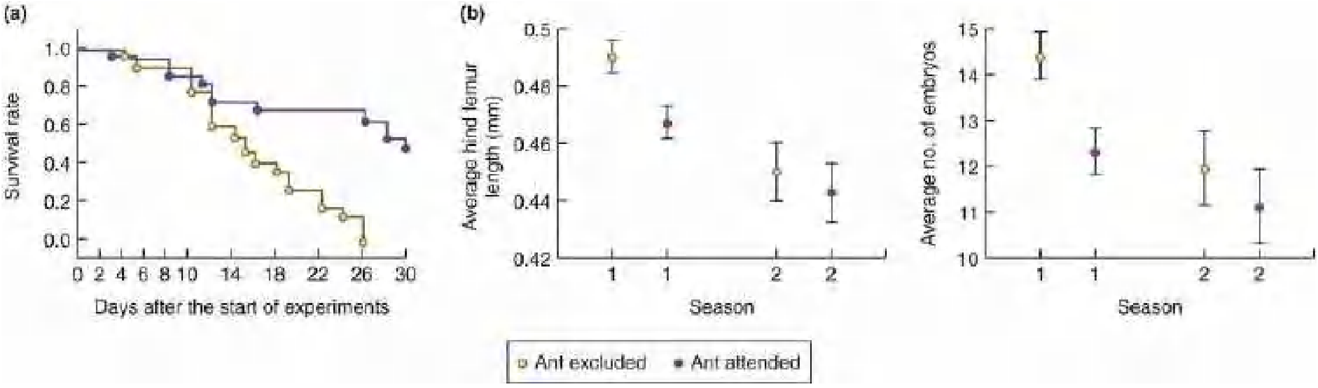


Figure 13.6 Aphid colonies survive longer when attended by ants but only if predators of aphids are present. (a) Ant-excluded colonies of the aphid *Tuberculatus quercicola* were more likely to become extinct than those attended by ants ($\chi^2 = 15.9$, $P < 0.0001$). (b) But in the absence of predators, ant-excluded colonies (green circles) perform better than those attended by ants (purple circles). Shown are the averages for aphid body size (hind femur length; $F = 6.75$, $P = 0.013$) and numbers of embryos ($F = 7.25$, $P = 0.010$), ± SE, for two seasons (23 July–11 August 1998 and 21–31 August 1998) in a predator-free environment.

Source: After Yao *et al*. (2000).

APPLICATION 13.2 A mutualistic ant-scale insect interaction may indirectly benefit coffee plants

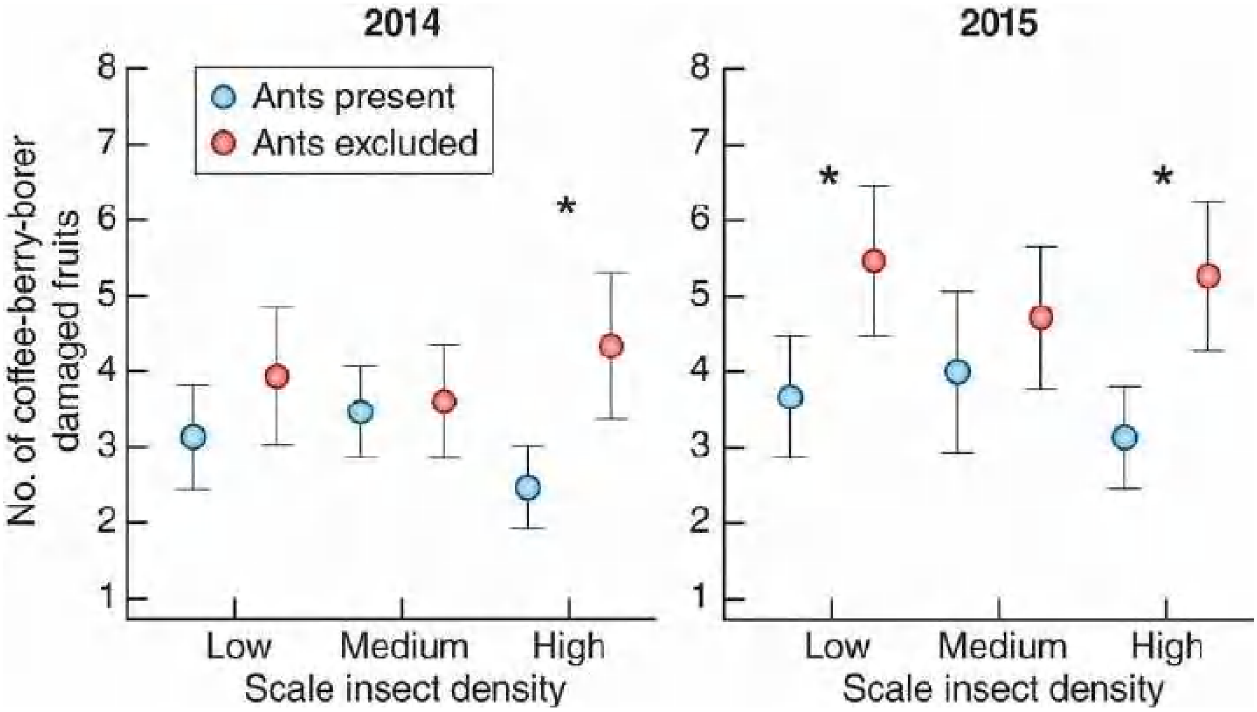
context dependence of mutualisms

Context dependence of mutualisms is a general theme of widespread importance, as we saw for interactions between cleaner wrasses and their client fish (benefit to the clients depends on level of infestation with ectoparasites) and between freshwater crayfish and branchiobdellidan worms (benefit to the crayfish depends on crayfish size and moulting rate). A context-dependent three-species interaction may also indirectly benefit coffee production.

a mutualism-dependent mutualism?

The interaction is common in natural and agroecosystems, and occurs between ants, honeydew-secreting scale insects (Hemiptera) and the plants on which they live. In coffee agroecosystems, the green coffee scale (*Coccus viridis*) can reach high densities but only when it is in a mutualistic relationship with ants (*Azteca sericeasur*). Like the aphids above, the scale insects provide sugar resources to the ants, while the ants protect the scale insects from their natural enemies (such as coccinellid beetles). Whether the ants should be classed as a pest of coffee, as has often been done in the past, depends on the trade-off between, on the one hand, harm to the coffee from the scale insect and, on the other, coincidental protection provided to the coffee plants by the ants against other pests, particularly the coffee berry borer (*Hypothenemus hampei*), which is far more damaging to coffee production than the scale insects. Note that coffee plants do not have a direct mutualistic relationship with ants, but there may be an indirect, so-called, ‘mutualism-dependent mutualism'. Rivera-Salinas *et al.* (2018) tested this idea in a coffee agroecosystem in southern Mexico in 2014 (a wetter year) and 2015 (a drier year). Each year they selected coffee plants: two branches per plant with low, intermediate or high densities of scale insects (medians of 15, 27–30 and 60–70 per branch, respectively). Where necessary they manipulated the numbers of scale insects on a branch. On one branch in each pair, ants were excluded by applying a sticky substance (‘tanglefoot') at its base. All coffee berry borers were removed from the branches before 20 adult borers were introduced per branch. The number of coffee fruits with coffee berry borer damage was recorded 24 hours later.

When scale insect density was high, the presence of ants did indeed benefit the coffee plants by reducing coffee berry borer attack (Figure 13.7). There also seems to be a weather dependent effect on this indirect mutualism, with benefits to the coffee plants even at low scale insect density in 2015, perhaps because scale insects produce higher quantities of sugar in their honeydew during drier times, enhancing ant activity. These results provide the basis for effective management action that takes into account species interactions and climatic factors, and call into question any blanket recommendation that ants be eliminated from coffee agroecosystems (Rivera-Salinas *et al.*, 2018).



[Figure 13.7](#) **When scale insect density is high, ants benefit coffee plants by reducing damage by coffee berry borers.** Number of coffee fruits damaged by coffee berry borer in relation to scale insect density on coffee branches and whether or not ants were excluded, in 2014 and 2015. Error bars are 95% CIs and asterisks show significant differences between treatments with and without ants.

Source: From Rivera-Salinas *et al*. (2018).

ants and blue butterflies

Another classic farming mutualism is that between ants and many species of lycaenid butterfly. In a number of cases, young lycaenid caterpillars feed on their preferred food plants usually until their third or fourth instar, when they expose themselves to foraging ant workers that pick them up and carry them back to their nests – the ants ‘adopt’ them. There, the ants ‘milk’ a sugary secretion from a specialised gland of the caterpillars, and in return protect them from predators and parasitoids throughout the remainder of their larval and pupal lives. On the other hand, in other lycaenid–ant interactions the evolutionary balance is rather different. The caterpillars produce chemical signals mimicking chemicals produced by the ants, inducing the ants to carry them back to their nests and allowing them to remain there. Within the nests, the caterpillars may either act as social parasites (‘cuckoos’, see Section 12.2.2), being fed by the ants (e.g. *Maculinea rebeli*, which feeds on the crossleaved gentian, *Gentiana cruciata*, and whose caterpillars mimic the larvae of the ant *Myrmica schenkii*), or they may simply prey upon the ants (e.g. *Maculinea arion* – see Application 13.3) (Elmes *et al*., 2002).

APPLICATION 13.3 The large blue – a butterfly in danger

The large blue, *Maculinea arion*, is the largest and rarest of Britain's blue butterflies. Extinct in Britain by 1979, it has been successfully reintroduced and by 2004 occurred in nine sites in southern England. Its caterpillars feed initially on the flower heads of wild thyme plants, *Thymus polytrichus*, but when about 4 mm long each drops to the ground and waits to be picked up by a red ant, *Myrmica sabuleti*, which takes it back to the ant nest to complete development, feeding on ant larvae (Figure 13.8). The large blue breeds in warm and well-drained grassland, predominantly acidic coastal grassland or limestone grassland. The conservation aim has been to produce short turf, less than 2 cm in height, with an abundance of wild thyme plus the host ant. When mean sward height exceeds 2 cm, the temperature in the ant nests is reduced by several degrees, allowing less thermophilous congeneric ant species, such as *M. scabrinodis*, to outcompete *M. sabuleti*. Survivorship of the butterfly is reduced five-fold when brooded by its non-preferred ant species, contributing significantly to the historic patterns of extinction (Thomas *et al.*, 2009). Grazing is essential to achieve the short turf conditions during spring and early summer, with rough coastal grasslands best grazed by hardy ponies or cattle, while limestone grassland can be grazed by either sheep or cattle. Then grazing must be removed in mid-summer to allow good wild thyme flowering. This successful restoration was enabled by a thorough understanding of the butterfly–ant mutualism.

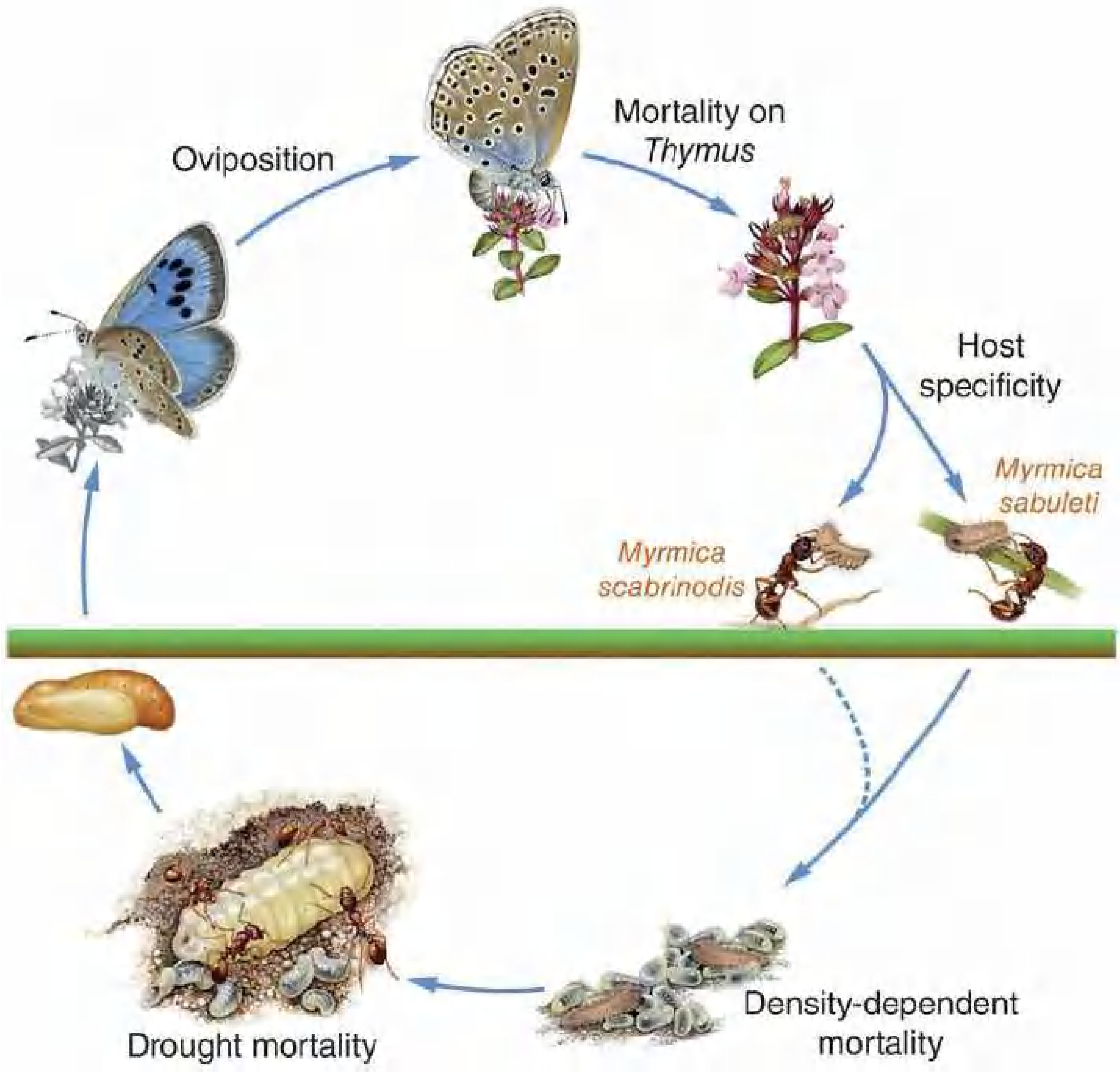


Figure 13.8 Life cycle of *Maculinea arion*. Adult butterflies oviposit on *Thymus polytrichus* from June to July. Causes of egg and larval mortality on the thyme flower heads are parasitism, predation and cannibalism. The final larval instar is adopted into the underground nest by the first ant worker to encounter it, but survival in the nest is substantially higher if the ant is *Myrmica sabuleti*. Mortality in the nest is higher if more than one larva is adopted into the nest and is amplified in drought years. After 10 months the larva pupates and an adult emerges two to three weeks later.

Source: From Thomas *et al*. (2009).

13.4.3 Farming of fungi by beetles and ants

ambrosia beetles

Much plant tissue, including wood, is unavailable as a direct source of food to most animals because they lack the enzymes that can digest cellulose and lignins (see Sections 3.7.1 and 11.3.1). However, many fungi possess these enzymes, and an animal that can eat such fungi gains indirect access to an energy-rich food. Some very specialised mutualisms have developed between animal and fungal decomposers. ‘Ambrosia' beetles in the subfamilies Scolytinae and Platypodinae tunnel deep into the wood of dead and dying trees, and ambrosia fungi that are specific for particular species of beetle grow in these burrows and are continually grazed by the beetle larvae. These ambrosia beetles may carry inocula of the fungus in their digestive tract, and some species bear specialised brushes of hairs on their heads that carry the spores. The fungi serve as food for the beetle and in turn depend on it for dispersal to new tunnels.

fungus-farming ants

Fungus-farming ants (subfamily Myrmicinae, tribe Attini) are found only in the New World, and the species, more than 230, appear to have evolved from a common ancestor: that is, the trait has appeared just once in evolution (Figure 13.9). They provide a case study of the evolution of sophistication in a mutualistic interaction. The first, most ‘primitive’ farmers (Paleoattini: extant ant genera *Mycocepurus, Myrmicocrypta, Apterostigma*), which evolved about 50 million years ago, typically use dead vegetative debris as well as insect faeces and corpses to manure their gardens of fungi from the tribe Leucocoprineae (Schultz & Brady, 2008). Because these fungi are also capable of independent existence, this is referred to as ‘lower agriculture'.

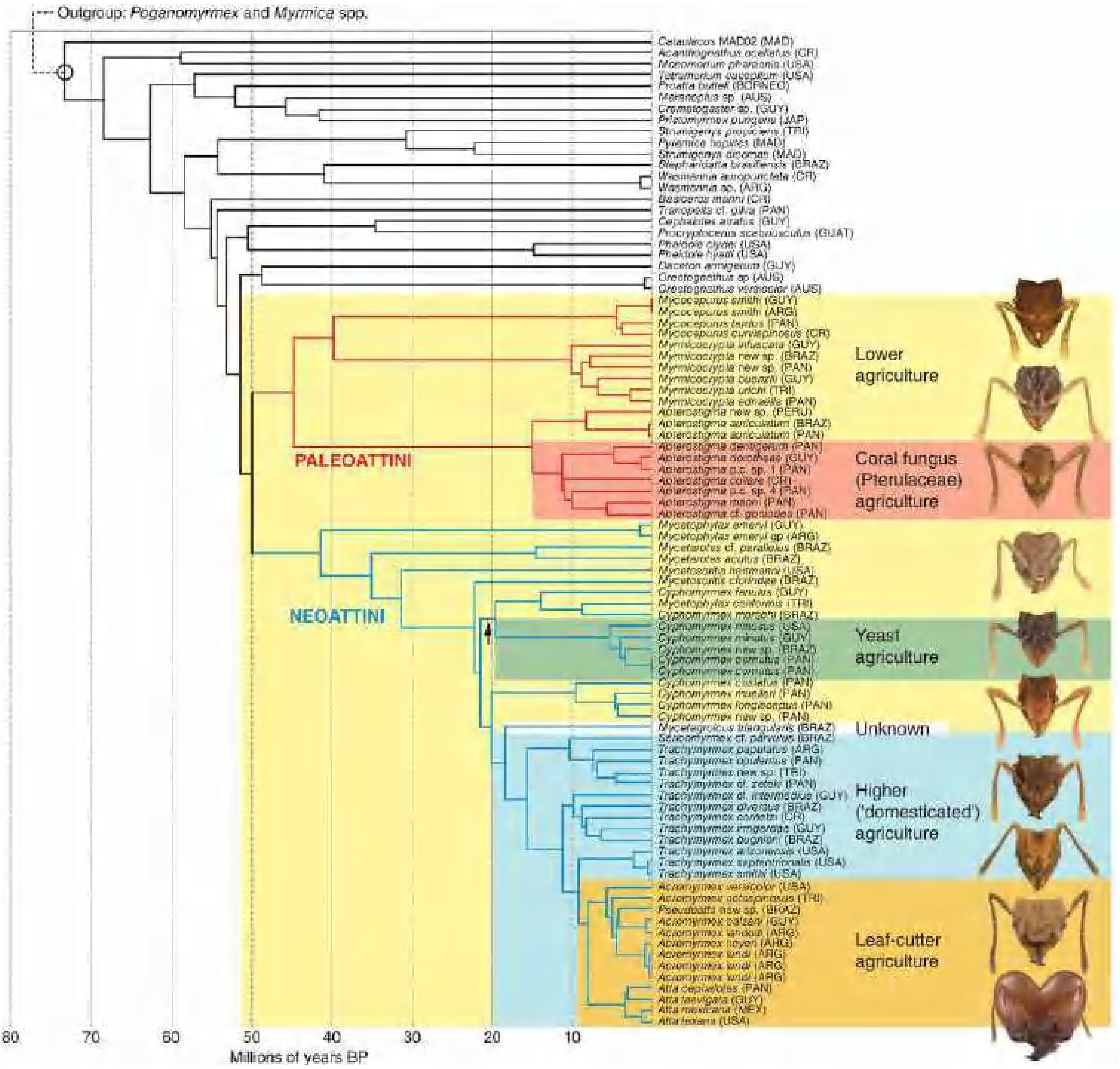


Figure 13.9 **Evolutionary origins of five categories of fungus farming by ants.** A time-calibrated phylogeny of fungus-farming ants (Attini) with age estimates for the origins of five ant agricultural systems, indicated by coloured rectangles and defined according to phylogenetically distinct groups of associated fungal cultivars. The geographical location of each species is indicated in brackets. The tree is based on analysis of nucleotides in four nuclear genes of 65 taxa of attine ants (fungus farmers, divided into Paleoattini and Neoattini) and 26 non-attine outgroups.

Source: From Schultz & Brady (2008).

lower agriculture – the farming of coral fungus and yeast gardens

Only one transition seems ever to have occurred from the use of Leucocoprineae fungi. This involved a switch, 10–20 million years ago, to farming 'coral' fungus (Pterulaceae) (Figure 13.9). The practice can still be considered 'lower agriculture' and extant ant species adopting it are a group within the genus *Apterostigma*. Another shift in fungal cultivar type (but still in the fungus family Leucocoprineae) occurred 5–25 million years ago, in which ants (extant species in the genus *Cyphomyrmex*) began farming fungus in the single-celled yeast phase rather than the mycelial phase as practiced by all other ant species. This may be viewed as an early stage of ‘domestication' of the ant's crop.

higher agriculture – domestication

The fungal cultivars of the ants that practice ‘higher agriculture' (Neoattini) have two derived features that suggest a significant degree of domestication for life with ants: firstly, they do not appear capable of a free-living existence separate from their ant hosts, and secondly, they are the only cultivars that produce gongylidia, nutritious swollen hyphal tips that the ants harvest. Most extant ant species farming these are in the genus *Trachymyrmex* (Figure 13.9).

higher agriculture – leaf-cutter ants

Leaf-cutting ants (in the genera *Atta* and *Acromyrmex*), which evolved 8–12 million years ago, are the most remarkable of the fungus-farming ants. They excavate 2–3-litre cavities in the soil, and in these their fungus is cultured on living leaves that are cut from neighbouring vegetation. The ant colony may depend absolutely on the fungus for the nutrition of their larvae. Workers lick the fungus colonies and remove specialised swollen hyphae, which are aggregated into bite-sized ‘staphylae’. These are fed to the larvae and this ‘pruning’ of the fungus may stimulate further fungal growth. The fungus gains from the association: it is both fed and dispersed by leaf-cutting ants and has never been found outside their nests. The reproductive female ant carries her last meal as a culture when she leaves one colony to found another, as indeed do daughter queens of all lower and higher agriculture species.

Most phytophagous insects have very narrow diets – indeed, many insect herbivores are strictly monophagous (see Section 3.7). The leaf-cutting ants are remarkable amongst insect herbivores in their polyphagy. Ants from a nest of *Atta cephalotes* harvest from 50% to 77% of the plant species in their neighbourhood; and leaf-cutting ants generally may harvest 17% of total leaf production in tropical rainforest and be the ecologically dominant herbivores in the community. It is their polyphagy that gives them this remarkable status. In contrast to the *A. cephalotes* adults though, the larvae appear to be extreme dietary specialists, being restricted to the nutritive gongylidia produced by the fungus *Attamyces bromatificus*, which the adults cultivate and which decompose the leaf fragments (Cherrett *et al*., 1989).

ants, farmed fungi, pest fungi and a bacterium: a three-way mutualism

Just as human farmers may be plagued by weeds, so fungus-farming ants have to contend with other species of fungus that may devastate their crop. Fungal pathogens in the genus *Escovopsis* are specialised (never found other than in fungus gardens) and virulent: in one experiment, nine of 16 colonies of the leaf-cutter *Atta colombica* that were treated with heavy doses of *Escovopsis* spores lost their garden within three weeks of treatment (Currie, 2001). But the ants have another mutualistic association to help them. A filamentous actinomycete bacterium in the genus *Pseudonocardia*, associated with the surface of the ants, is dispersed to new gardens by virgin queens on their nuptial flight, and the ants may even produce chemicals that promote the actinomycete’s growth. For its part, the actinomycete produces an antibiotic with specialised and potent inhibitory effects against *Escovopsis*. It even appears to protect the ants themselves from pathogens and to promote the growth of the farmed fungi (Currie, 2001). *Escovopsis* therefore has ranged against it not just two two-species mutualisms but a three-species mutualism amongst ants, farmed fungi and actinomycete bacteria.

13.5 Dispersal of seeds and pollen

13.5.1 Seed dispersal mutualisms

Very many plant species use animals to disperse their seeds and pollen. About 10% of all flowering plants possess seeds or fruits that bear hooks, barbs or glues that become attached to the hairs, bristles or feathers of any animal that comes into contact with them. They are frequently an irritation to the animal, which often cleans itself and removes them if it can, but usually after carrying them some distance. In these cases, the benefit is to the plant (which has invested resources in attachment mechanisms) and there is no reward to the animal.

fruits

Quite different are the true mutualisms between higher plants and the birds, mammals and other animals that feed on the fleshy fruits and disperse the seeds. Of course, for the relationship to be mutualistic it is essential that the animal digests only the fleshy fruit and not the seeds, which must remain viable when regurgitated, defaecated or discarded. Thick, strong defences that protect plant embryos are usually part of the price paid by the plant for dispersal by fruit-eaters. The plant kingdom has exploited a splendid array of morphological variations in the evolution of fleshy fruits. In their review of seed dispersal mutualisms in North American terrestrial plants (12 424 species), Vander Wall *et al.* (2017) found that 10% were dispersed by frugivorous mammals or birds, in comparison with 3.7% by scatter hoarding rodents and birds, and 3.9% by myrmecochory (seed dispersal by ants), where the seeds have an edible attachment (elaiosome, caruncle or aril) that ants eat, before burying or discarding the seeds in a waste pile.

Mutualisms involving animals that eat fleshy fruits and disperse seeds are seldom very specific to the species of animal concerned. Partly, this is because these mutualisms usually involve long-lived birds or mammals, and even in the tropics there are few plant species that fruit throughout the year and form a reliable food supply for any one specialist. But also, as will be apparent when pollination mutualisms are considered next, a more exclusive mutualistic link would require the plant's reward to be protected and denied to other animal species: this is much easier for nectar than for fruit (although myrmecochory is an obvious exception). In any case, specialisation by the animal is important in pollination, because interspecies transfers of pollen are disadvantageous, whereas with fruit and seed it is necessary only that they are dispersed away from the parent plant.

13.5.2 Pollination mutualisms

Most animal-pollinated flowers offer nectar, pollen or both as a reward to their visitors. Floral nectar seems to have no value to the plant other than as an attractant to animals and it has a cost to the plant, because the nectar carbohydrates might have been used in growth or some other activity.

Presumably, the evolution of specialised flowers and the involvement of animal pollinators have been favoured because an animal may be able to recognise and discriminate between different flowers and so move pollen between different flowers of the same species but not to flowers of other species. Passive transfer of pollen, for example by wind or water, does not discriminate in this way and is therefore much more wasteful. Indeed, where the vectors and flowers are highly specialised, as is the case in many orchids, virtually no pollen is wasted even on the flowers of other orchid species.

There are, though, costs that arise from adopting animals as mutualists in flower pollination. For example, animals carrying pollen may be responsible for the transmission of sexual diseases as well (Shykoff & Bucheli, 1995). The fungal pathogen *Microbotryum violaceum*, for example, is transmitted by pollinating visitors to the flowers of white campion (*Silene alba*) and in infected plants the anthers are filled with fungal spores.

insect pollinators: from generalists to ultraspecialists

Many different kinds of animals have entered into pollination liaisons with flowering plants, including hummingbirds, bats and even small rodents and marsupials. However, the pollinators *par excellence* are, without doubt, the insects. Pollen is a nutritionally rich food resource, and in the simplest insect-pollinated flowers, pollen is offered in abundance and freely exposed to all and sundry. The plants rely for pollination on the insects being less than wholly efficient in their pollen consumption, carrying their spilt food with them from plant to plant. In more complex flowers, nectar (a solution of sugars) is produced as an additional or alternative reward. In the simplest of these, the nectaries are unprotected, but with increasing specialisation the nectaries are enclosed in structures that restrict access to the nectar to just a few species of visitor. This range can be seen within the family Ranunculaceae. In the simple flower of *Ranunculus ficaria* the nectaries are exposed to all visitors, but in the more specialised flower of *R. bulbosus* there is a flap over the nectary, and in *Aquilegia* the nectaries have developed into long tubes and only visitors with long probosces (tongues) can reach the nectar. In the related *Aconitum* the whole flower is structured so that the nectaries are accessible only to insects of the right shape and size that are forced to brush against the anthers and pick up pollen. Unprotected nectaries have the advantage of a ready supply of pollinators, but because these pollinators are unspecialised they transfer much of the pollen to the flowers of other species (though in practice, many generalists are actually ‘sequential specialists’, foraging preferentially on one plant species for hours or days). Protected nectaries have the advantage of efficient transfer of pollen by specialists to other flowers of the same species, but are reliant on there being sufficient numbers of these specialists.

Charles Darwin (1859) recognised that a long nectary, as in *Aquilegia* (see Section 3.7), forced a pollinating insect into close contact with the pollen at the nectary’s mouth. Natural selection may then favour even longer nectaries, and as an evolutionary reaction, the tongues of the pollinator would be selected for increasing length – in this case, Darwin’s hawkmoth, *Xanthopan morganii praedicta*, with its 20 cm long proboscis – a reciprocal and escalating process of specialisation. Nilsson (1988) deliberately shortened the nectary tubes of the long-tubed orchid *Platanthera* and showed that the flowers then produced many fewer seeds – presumably because the pollinator was not forced into a position that maximised the efficiency of pollination.

seasonality

Flowering is a seasonal event in most plants, imposing strict limits on the degree to which a pollinator can become an obligate specialist. A pollinator can only become completely dependent on specific flowers as a source of food if its life cycle matches the flowering season of the plant. This is feasible for many short-lived insects like butterflies and moths, but longer lived pollinators such as bats and rodents, or bees with their long-lived colonies, are more likely to be generalists, turning from one relatively unspecialised flower to another through the seasons or to quite different foods when nectar is unavailable.

APPLICATION 13.4 Restoration of pollination networks

Land degradation results in ecological disruption worldwide, leading to both loss of native biodiversity and disruption of processes such as the provisioning of pollination (which can be considered a critical ‘ecosystem service' – see Section 15.4.1). Vegetation restoration is often used to mitigate loss of biodiversity, but does it also help to restore the functioning of pollination networks? Kaiser-Bunbury *et al.* (2017) recorded the effects on pollination of vegetation restoration (the removal of exotic shrubs) in four isolated rocky mountaintop communities on the island of Mahé in the Seychelles, in comparison with four control sites where no restoration was performed.

Pollination networks were recorded at each site seven months after shrub removal, and this was repeated at eight monthly intervals through the flowering season (September 2012 to April 2013). Pollinators of all woody flowering plant species (38 species) were scored for a total of 1525 observation hours, during which 581 plant–pollinator interactions (links) and 12 235 pollinator visits to flowers were recorded. Pollinators included bees and wasps (Hymenoptera, 25 spp.), flies (Diptera, 59 spp.), beetles (Coleoptera, 38 spp.), moths and butterflies (Lepidoptera, 17 spp.), two bird species (Nectariniidae, Pycnonotidae), and three lizard species (Gekkonidae, Scincidae). Figure 13.10a shows the structure of the pollination networks for each restored and control site in November 2013 (late spring).

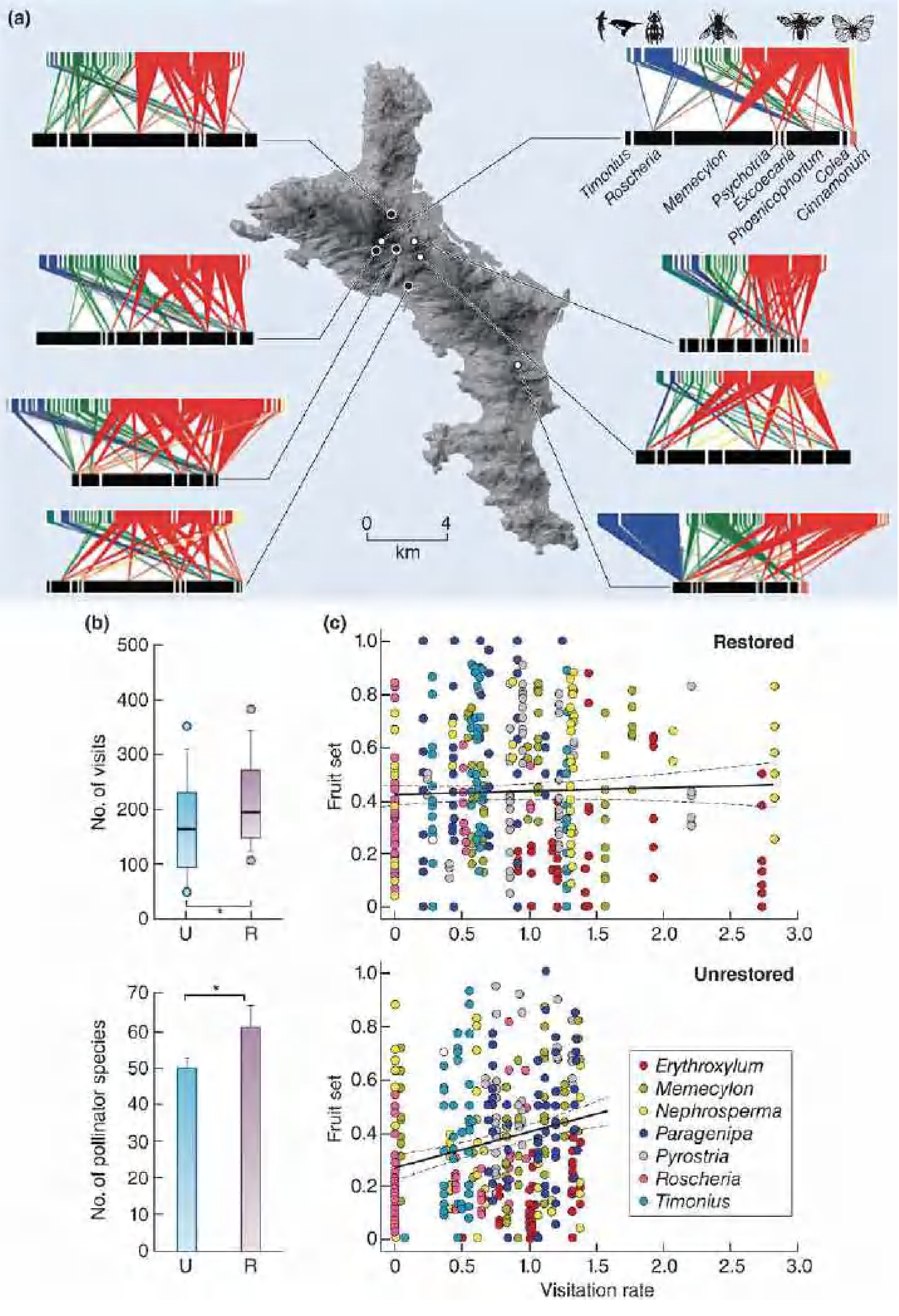

(a)
Timonius
Roscheria
Memecylon
Psychotria
Excoecaria
Phoenicophorium
Colea
Cinnamomum
0
4
km
(b)
No. of visits
500
400
300
200
100
0
U
R
*
No. of pollinator species
70
60
50
40
30
20
10
0
*
U
R
(c)
Restored
Fruit set
1.0
0.8
0.6
0.4
0.2
0
0
0.5
1.0
1.5
2.0
2.5
3.0
Unrestored
Fruit set
Visitation rate
Erythroxylum
Memecylon
Nephrosperma
Paragenipa
Pyrostria
Roscheria
Timonius

Figure 13.10 Restoration of vegetation increases the number of pollinator species, the number of pollinator visits and fruit production by native plants. (a) Pollination networks (links between plants (bottom bar) and pollinators (top bar)) recorded in late spring on the island of Mahéin in the Seychelles, seven months after exotic shrub removal from four mountain tops (on the left of figure, large black circles) compared with unrestored mountain top sites (on the right, small white dots). Each block represents a species, its width representing relative abundance, and the widths of the links represent the interaction frequency between pollinators and plants. Native and exotic plant species are shown in black and pink respectively (bottom bar). Pollinator groups are represented by colours: birds and lizards, light blue; beetles, dark blue; flies, green; wasps and bees, red; butterflies and moths, yellow. (b) Restoration (R) increases the number of pollinator visits ($P < 0.05$) and the number of pollinator species ($P < 0.05$), compared with unrestored sites (U). (c) Fruit set of the seven most common native species across all sites increased with visitation rate of pollinators at the unrestored site (i.e. the plants there were pollinator-limited), but not at restored sites ($P < 0.001$ for the visitation rate effect and $P < 0.001$ for the interaction between visitation rate and the restored/unrestored treatment). In (b) and (c) data for all eight months of the study were combined.

Source: From Kaiser-Bunbury *et al.* (2017).

Restoration resulted in a marked increase in the number of pollinator species, and in the number of visits per flower (Figure 13.10b). The restored pollination networks were also more generalised than the unrestored networks (having more pollinator species per plant and/or a more even distribution of visits per pollinator species), indicating a higher functional redundancy in restored communities with the implication that these networks would be more resilient to any future environmental disturbance. Of particular significance was the finding that plant reproductive performance (the proportion of flowers that set fruit) was limited by pollination visitor rate in the unrestored sites but not the restored sites (Figure 13.10c). Such results highlight the importance of understanding pollination interaction networks for effective restoration.

13.5.3 Brood site pollination: figs and yuccas

figs and fig wasps ...

Not every insect-pollinated plant provides its pollinator with just a take-away meal. In a number of cases, the plants also provide a home and sufficient food for the development of the insect larvae (Sakai, 2002). The best studied of these are the complex, largely species-specific interactions between figs (*Ficus*) and fig wasps. Figs bear many tiny flowers on a swollen receptacle with a narrow opening to the outside; the receptacle then becomes the fleshy fruit. The best-known species is the edible fig, *Ficus carica*. Some cultivated forms are entirely female and require no pollination for fruit to develop, but in wild *F. carica* three types of receptacle are produced at different times of the year. (Other species are less complicated, but the life cycle is similar.) In winter, the flowers are mostly neuter (sterile female) with a few male flowers near the opening. Tiny females of the wasp *Blastophaga psenes* invade the receptacle, lay eggs in the neuter flowers and then die. Each wasp larva then completes its development in the ovary of one flower, but the males hatch first and chew open the seeds occupied by the females and then mate with them. In early summer the females emerge, receiving pollen at the entrance from the male flowers, which have only just opened.

The fertilised females carry the pollen to a second type of receptacle, containing neuter and female flowers, where they lay their eggs. Neuter flowers, which cannot set seed, have a short style that the wasps can reach to lay their eggs in the ovaries where they develop. Female flowers, though, have long styles so the wasps cannot reach the ovaries and their eggs fail to develop, but in laying these eggs they fertilise the flowers, which set seed. Hence, these receptacles generate a combination of viable seeds (that benefit the fig) and adult fig wasps (that obviously benefit the wasps, but also benefit the figs since they are the figs' pollinators).

Following another round of wasp development, fertilised females emerge in the autumn, and a variety of other animals eat the fruit and disperse the seeds. The autumn-emerging wasps lay their eggs in a third kind of receptacle containing only neuter flowers, from which wasps emerge in winter to start the cycle again.

... show mutualism despite conflict

This, then, apart from being a fascinating piece of natural history, is a good example of a mutualism in which the interests of the two participants nonetheless appear not to coincide. Specifically, the optimal proportion of flowers that develop into fig seeds and fig wasps is different for the two parties, and we might reasonably expect to see a negative correlation between the two: seeds produced *at the expense* of wasps, and vice versa (Herre & West, 1997). In fact, detecting this negative correlation, and hence establishing the conflict of interest, has proved elusive for reasons that frequently apply in studies of evolutionary ecology. The two variables tend, rather, to be *positively* correlated, since both tend to increase with two 'confounding' variables: the overall size of fruit and the overall proportion of flowers in a fruit that are visited by wasps. Herre and West (1997), however, in analysing data from nine species of New World figs, were able to overcome this in a way that is generally applicable in such situations. They controlled statistically for variation in the confounding variables (asking, in effect, what the relationship between seed and wasp numbers would be in a fruit of constant size in which a constant proportion of flowers was visited) and then were able to uncover a negative correlation. The fig and fig wasp mutualists *do* appear to be involved in an on-going evolutionary battle.

yuccas and yucca moths

A similar, and similarly much studied, set of mutualisms occurs between the 35 and 50 species of *Yucca* plant that live in North and Central America and the 17 species of yucca moth (Pellmyr, 2002). A female moth uses specialised 'tentacles' to collect together pollen from several anthers in one flower, which she then takes to the flower of another inflorescence (promoting outbreeding) where she both lays eggs in the ovaries and carefully deposits the pollen, again using her tentacles. The development of the moth larvae requires successful pollination, since unpollinated flowers quickly die, but the larvae also consume seeds in their immediate vicinity, though many other seeds develop successfully. On completing their development, the larvae drop to the soil to pupate, emerging one or more years later during the yucca's flowering season. The reproductive success of an individual adult female moth is not, therefore, linked to that of an individual yucca plant in the same way as are those of female fig wasps and figs.

13.6 Mutualisms involving gut inhabitants

Most of the mutualisms discussed so far have depended on patterns of behaviour, where neither species lives entirely 'within' its partner. In many other mutualisms, one of the partners is a unicellular eukaryote or bacterium that is integrated more or less permanently into the body cavity or even the cells of its multicellular partner. The microbiota occupying parts of various animals' alimentary canals are the best-known extracellular symbionts.

13.6.1 Vertebrate guts

digestion of cellulose

The crucial role of microbes in the digestion of cellulose by vertebrate herbivores has long been appreciated, but it is now clear that the gastrointestinal tracts of all vertebrates are populated by a mutualistic microbiota. Protozoa and fungi are usually present but the major contributors to these ‘fermentation’ processes are bacteria. Their diversity is greatest in regions of the gut where the pH is relatively neutral and food retention times are relatively long. In small mammals (e.g. rodents, rabbits and hares) the caecum is the main fermentation chamber, whereas in larger non-ruminant mammals such as horses the colon is the main site, as it is in elephants, which, like rabbits, practice coprophagy (consume their own faeces) (Figure 13.11). In ruminants, such as cattle and sheep, and in kangaroos and other marsupials, fermentation occurs in specialised stomachs.

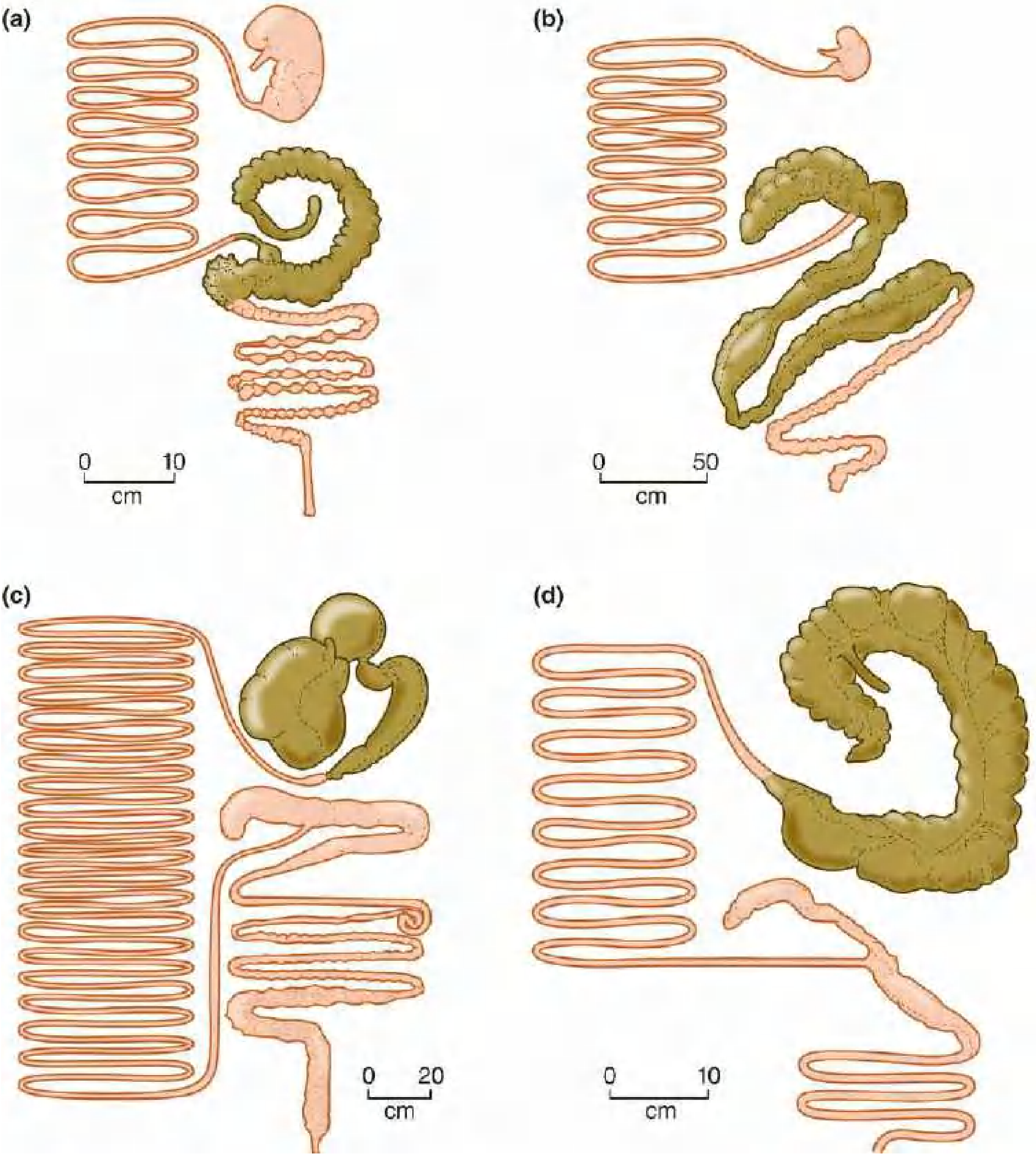


Figure 13.11 The digestive tracts of herbivorous mammals are commonly modified to provide fermentation chambers containing a diverse microbial community. (a) A rabbit, with a fermentation chamber (greeny brown colouration) in the expanded caecum. (b) A zebra, with fermentation chambers in both the caecum and colon. (c) A sheep, with foregut fermentation in an enlarged portion of the stomach, rumen and reticulum. (d) A kangaroo, with an elongate fermentation chamber in the proximal portion of the stomach.

Source: After Stevens & Hume (1998).

The basis of the mutualism is straightforward. The microbes receive a steady flow of substrates for growth in the form of food that has been eaten, chewed and partly homogenised. They live within a chamber in which pH and, in endotherms, temperature are regulated and anaerobic conditions are maintained. The vertebrate hosts, especially the herbivores, receive nutrition from food that they would otherwise find, literally, indigestible. The bacteria produce short-chain fatty acids (SCFAs) by fermentation of the host's dietary cellulose and starches and of the endogenous carbohydrates contained in host mucus and sloughed epithelial cells. SCFAs are often a major source of energy for the host; for example, they provide more than 60% of the maintenance energy requirements for cattle and 29–79% of those for sheep (Stevens & Hume, 1998). The microbes also convert nitrogenous compounds (amino acids that escape absorption in the midgut, urea that would otherwise be excreted by the host, mucus and sloughed cells) into ammonia and microbial protein, conserving nitrogen and water; and they synthesise B vitamins. The microbial protein is useful to the host if it can be digested – in the intestine by foregut fermenters and following coprophagy in hindgut fermenters – but ammonia is usually not useful and may even be toxic to the host.

ruminant guts

The stomach of ruminants comprises a three-part forestomach (rumen, reticulum and omasum) followed by an enzyme-secreting abomasum that is similar to the whole stomach of most other vertebrates. The rumen and reticulum are the main sites of fermentation, and the omasum serves largely to transfer material to the abomasum. Only particles with a volume of about 5 μl or less can pass from the reticulum into the omasum; the animal regurgitates and rechews the larger particles (the process of rumination). Dense populations of bacteria (10^{10}–10^{11} ml^{-1}) and protozoa (10^5–10^6 ml^{-1} but occupying a similar volume to the bacteria) are present in the rumen. The bacterial communities of the rumen are composed almost wholly of obligate anaerobes – many are killed instantly by exposure to oxygen – but they perform a wide variety of functions (subsist on a wide variety of substrates) and generate a wide range of products (Table 13.1). Cellulose and other fibres are the important constituents of the ruminant's diet, and the ruminant itself lacks the enzymes to digest these. The cellulolytic activities of the rumen microbiota are therefore of crucial importance. But not all the bacteria are cellulolytic: many subsist on substrates (lactate, hydrogen) generated by other bacteria in the rumen.

Table 13.1 Functions and products of rumen bacteria. A number of the bacterial species of the rumen, illustrating their wide range of functions and the wide range of products that they generate.

Source: After Allison (1984); Stevens and Hume (1998).

Species	**Functions**	**Products**
Bacteroides succinogenes	C, A	F, A, S
Ruminococcus albus	C, X	F, A, E, H, C
R. flavefaciens	C, X	F, A, S, H
Butyrivibrio fibrisolvens	C, X, PR	F, A, L, B, E, H, C
Clostridium lochheadii	C, PR	F, A, B, E, H, C
Streptococcus bovis	A, SS, PR	L, A, F
B. amylophilus	A, P, PR	F, A, S
B. ruminicola	A, X, P, PR	F, A, P, S
Succinimonas amylolytica	A, D	A, S
Selenomonas ruminantium	A, SS, GU, LU, PR	A, L, P, H, C
Lachnospira multiparus	P, PR, A	F, A, E, L, H, C
Succinivibrio dextrinosolvens	P, D	F, A, L, S
Methanobrevibacter ruminantium	M, HU	M
Methanosarcina barkeri	M, HU	M, C
Spirochete species	P, SS	F, A, L, S, E
Megasphaera elsdenii	SS, LU	A, P, B, V, CP, H, C
Lactobacillus sp.	SS	L
Anaerovibrio lipolytica	L, GU	A, P, S
Eubacterium ruminantium	SS	F, A, B, C

Functions: A, amylolytic; C, cellulolytic; D, dextrinolytic; GU, glycerol utilising; HU, hydrogen utiliser; L, lipolytic; LU, lactate utilising; M, methanogenic; P, pectinolytic; PR, proteolytic; SS, major soluble sugar fermenter; X, xylanolytic.

Products: A, acetate; B, butyrate; C, carbon dioxide; CP, caproate; E, ethanol; F, formate; H, hydrogen; L, lactate; M, methane; P, propionate; S, succinate; V, valerate.

The protozoa in the gut are also a complex mixture of specialists. Most are holotrich ciliates and entodiniomorphs. A few can digest cellulose. The cellulolytic ciliates have intrinsic cellulases, although some other protozoa may use their own bacterial symbionts. Some consume bacteria: in their absence the number of bacteria rise. Some of the entodiniomorphs prey on other protozoa. Thus, the diverse processes of competition, predation and mutualism, and the food chains that characterise terrestrial and aquatic communities in nature, are all present within the rumen microcosm.

refection and coprophagy

Eating faeces is taboo amongst humans, presumably through some combination of biological and cultural evolution in response to the health hazards posed by pathogenic microbes, including many that are relatively harmless in the hindgut but are pathogenic in more anterior regions. For many vertebrates, however, symbiotic microbes, living in the hindgut beyond the regions where effective nutrient absorption is possible, are a resource that is too good to waste. Thus coprophagy (eating faeces) or refection (eating one's own faeces) is a regular practice in many small, and some large, herbivorous mammals. This is developed to a fine art in species such as rabbits that have a 'colonic separation mechanism' that allows them to produce separate dry, non-nutritious faecal pellets and soft, more nutritious pellets that they consume selectively. These contain high levels of SCFAs, microbial protein and B vitamins, and can provide 30% of a rabbit's nitrogen requirements and more B vitamins than it requires (Björnhag, 1994 ; Stevens & Hume, 1998).

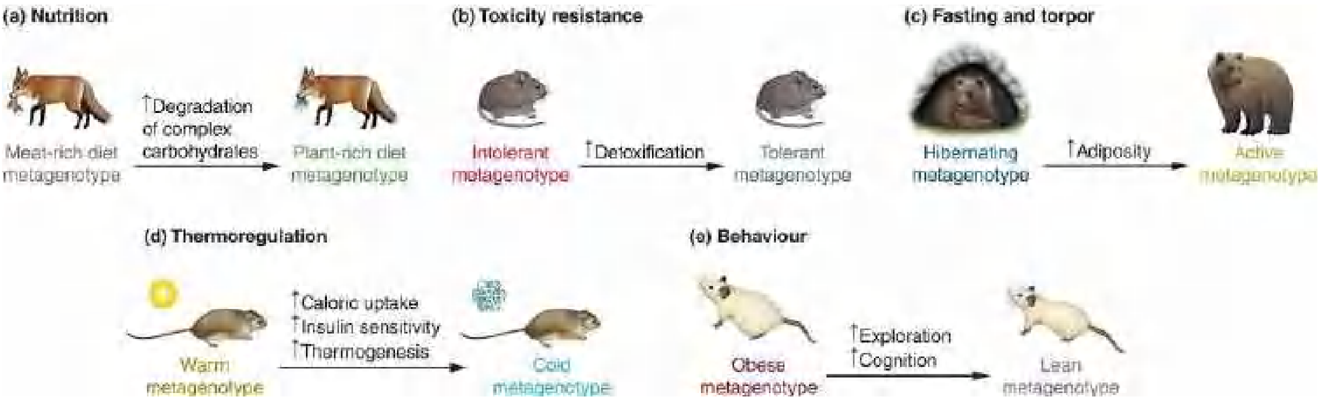
13.6.2 The vertebrate gut metagenome

The gut metagenome is defined as the complete set of DNA of symbiotic microorganisms in an animal's gut. The largest populations of gut microorganisms are found in mammals, with up to 10^{10}–10^{12} cells ml^{-1} comprising a multitude of species in the rumen and/or large intestine. The human gut, for example, contains about a thousand bacterial species with more than 3 million genes between them, in comparison to a mere 19–20 000 genes in the human genome (Alberdi *et al*., 2016).

The effects of the microorganisms on host vertebrate fitness can range from mutualistic, through commensalistic to parasitic but the full spectrum, and certainly the details, of these effects is not yet known (Walter *et al*., 2010). The net effect of the gut microbiota, however, is beneficial, and this is not due simply to improved digestion. They also confer benefits by competing with pathogenic bacteria, through their contribution to a gut-associated mucosal immune system that regulates and maintains a beneficial microbial community, and by modulating the host's immune system. In the digestive tracts of birds and mammals, the bacterium *Lactobacillus reuteri*, for example, produces an array of antimicrobial compounds that inhibit pathogenic bacteria. In addition, it plays an immunoregulatory role, being highly efficient in reducing inflammation and preventing or reducing diarrhoea and colitis (Walter *et al*., 2010).

the gut metagenome and adaptation to climate change

The gut metagenome is much more easily altered than the host's genome because the gut microbiota changes continuously in response to variations in both host condition and external environmental factors. And there are now many examples of environmentally induced changes to the gut metagenome associated with changed biochemical, physical or behavioural characteristics of the host (Figure 13.12). Alberdi *et al*. (2016) argue that the plasticity of the gut microbiota may be an essential factor in determining plasticity in vertebrate performance, and thus may play a critical role in the ability of vertebrates to acclimatise and adapt to fast environmental change, such as that expected from climate warming.



[**Figure 13.12**](#) **Alterations in host characteristics associated with changes to the gut metagenome.** (a) A switch from a meat-rich to a plant-rich diet may alter microbial community structure, improving the ability to digest complex carbohydrates. (b) Desert woodrats (*Neotoma lepida*) exposed to toxic plants develop a toxicity-tolerant gut metagenome that enhances detoxification ability. (c) The winter and summer gut metagenomes of brown bears (*Ursus arctos*) are different, the latter fostering energy storage and adiposity. (d) Cold exposure changes the gut metagenome, fostering enhanced caloric uptake, insulin sensitivity and thermogenesis. (e) Lean and obese mice have different gut metagenomes, the latter fostering more active behaviour and enhancing cognition.

Source: From Alberdi *et al*. (2016), where relevant references can be found.

13.6.3 Insect guts

The guts of insects (Figure 13.13) are also sites of microorganism activity, and insect gut microorganisms have been shown to upgrade nutrient-poor diets, aid digestion, protect from enemies and even influence interindividual communication and mating and reproductive systems (Engel & Moran, 2013). Compared with vertebrate guts, most insect guts contain far fewer microbial species, but some contain large communities of specialised bacteria. In comparison with vertebrates, an obstacle to the evolution of intimate mutualistic associations in insects is the lack of dependable transmission routes between host individuals, with, for example, both moulting and metamorphosis leading to the reduction or elimination of gut bacteria. However, social interactions in termite, bee and other social insect species provide opportunities for transfer of gut microorganisms (e.g. Martinson *et al*., 2012) and it is in these groups that the most distinctive gut communities have been found.

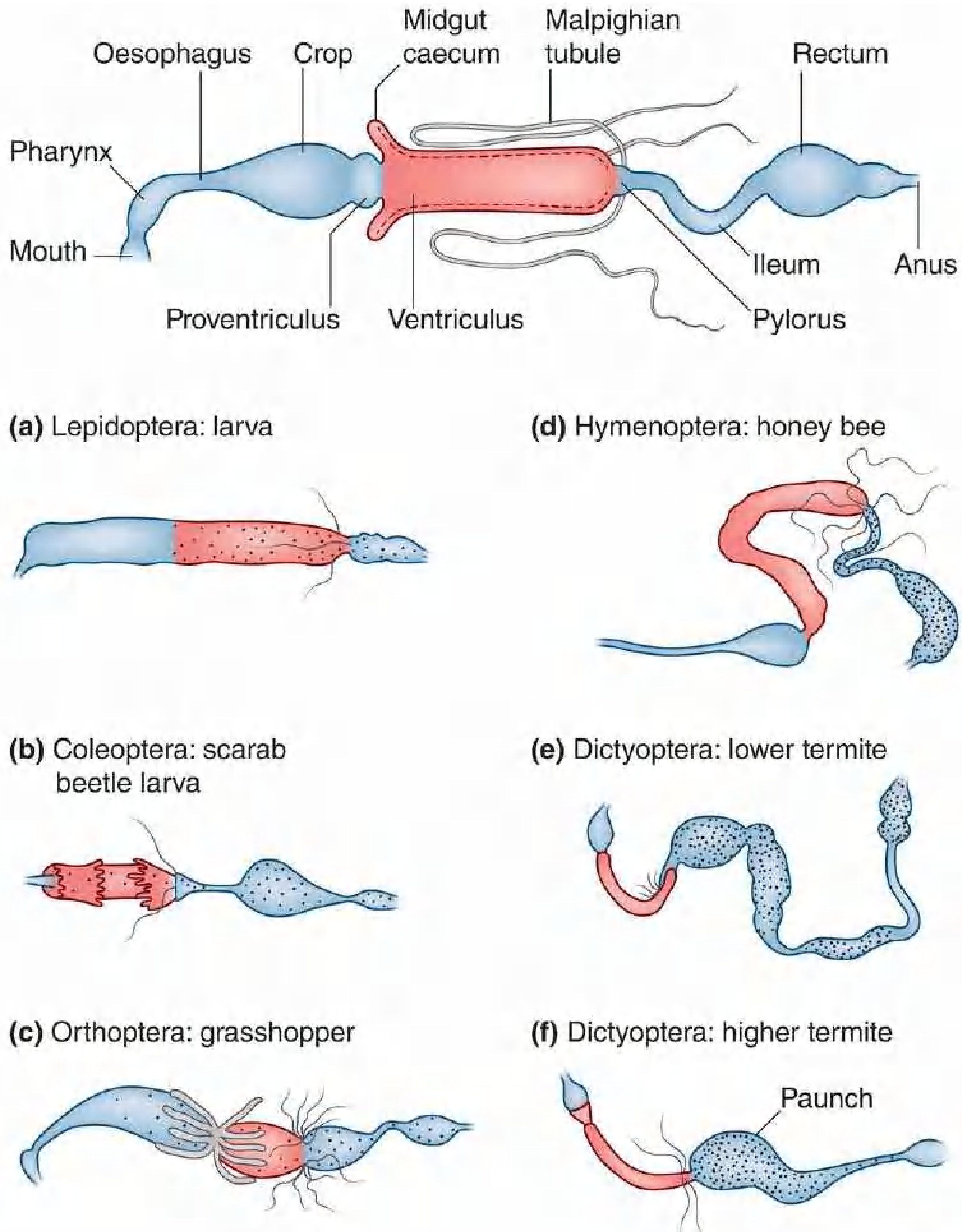


Figure 13.13 **Examples of insect guts showing the localisation of gut bacteria.** Top, generalised insect gut structure (foreguts and hindguts (blue) are lined by a cuticle layer and the midgut (pink) secretes a peritrophic matrix (dashed line)). (a–f) Specific examples from selected insect orders. Stippling shows predominant localisation of bacteria.

Source: From Engel & Moran (2012), who adapted the diagrams from various sources.

Termites are social insects in the superorder Dictyoptera, many of which depend on mutualists for the digestion of wood (Brune & Dietrich, 2015). Lower termites feed directly on wood, and most of the cellulose and hemicelluloses is digested by mutualists in the paunch (part of the segmented hindgut), which forms a microbial fermentation chamber (Figure 13.13e). Cellulolytic flagellates play a dominant role in wood digestion in the lower termites and the evolution of this mutualism was facilitated by the development of proctodeal trophyllaxis – where nestmates consume droplets of hindgut contents. The evolutionary loss of flagellates in the higher termites (75% of all species; Figure 13.13f) resulted in an entirely prokaryotic gut microbiota of bacteria and archaea, and has been associated with enormous dietary diversification (wood, plant litter, animal dung, humus, soil organic matter). Finally, a third group, the Macrotermitinae, cultivate wood-digesting fungi (see Section 13.4.3 for other examples of fungus farming) that the termites eat along with the wood itself, which the fungal cellulases assist in digesting.

APPLICATION 13.5 Disruption of the honey bee gut metagenome by a pesticide

If honey bees are deprived of their gut microbiota, normally dominated by eight bacterial species, their growth is reduced, their metabolism altered, they are more susceptible to pathogens and they suffer higher mortality in the hive (Motta *et al.*, 2018). Given the growing realisation of the importance of animal gut metagenomes, it is clear that attention needs to be paid to the effects of xenobiotics (alien or unnatural chemicals) not just on animals directly, but via changes to their gut microbiotas. Glyphosate, the world's primary herbicide, targets the enzyme 5-enolpyruvylshikimate-3-phosphate synthase and its biochemical pathway 'shikimate', which are found only in plants and some microorganisms. Given that animals lack the shikimate pathway, glyphosate has been considered one of the least toxic of agricultural pesticides. However, if such a pesticide were to adversely affect a honey bee's gut bacteria, it becomes a candidate in the search for understanding the alarming increase in mortality of bee hives, and the loss of honey bee pollination services. Motta *et al.* (2018) found that the absolute abundances of four of the eight dominant gut bacteria (including *Snodgrassella alvi*) of worker bees treated with glyphosate at levels occurring in the environment were significantly reduced in comparison to control bees treated with a sterile sucrose syrup. Moreover, when newly emerged worker bees in the process of acquiring their normal microbial community were challenged with the pathogenic bacterium *Serratia marcescens*, mortality was higher for bees previously exposed to glyphosate than unexposed controls. Understanding how xenobiotics affect the gut metagenomes of target animals will help identify their potential roles in population endangerment, including the colony declines reported for honey bees around the world.

13.7 Mutualism within animal cells: insect bacteriocyte symbioses

nutritional benefits

In bacteriocyte symbioses between microorganisms and insects, the maternally inherited microorganisms are found within the cytoplasm of specialised cells, bacteriocytes, and the interaction is unquestionably mutualistic. It is required by the insects for the nutritional benefits the microorganisms bring, as key providers of essential amino acids, lipids and vitamins, and is required by the microorganisms for their very existence. Bacteriocyte symbioses have no parallel in mammals, where all known intracellular bacteria are pathogens (Douglas, 2014). Bacteriocytes are found in a wide variety of insects including cockroaches, plant-sucking hemipterans, bed bugs, blood sucking lice, tsetse flies, lyctid beetles and camponotid ants. They have evolved independently in different groups of microorganisms and their insect partners, but in effectively all cases the insects live their lives on nutritionally poor or unbalanced diets: phloem sap, vertebrate blood, wood and so on. Mostly the symbionts are various sorts of bacteria, although in some insects yeasts are involved.

aphids and *Buchnera* ...

Amongst these symbioses, most is known by far about the interactions between aphids and bacteria in the genus *Buchnera*. The bacteriocytes are found in the haemocoel of the aphids and the bacteria occupy around 60% of the bacteriocyte cytoplasm. The bacteria cannot be brought into culture in the laboratory and have never been found other than in aphid bacteriocytes, but the extent and nature of the benefit they bring to the aphids can be studied by removing the *Buchnera* by treating the aphids with antibiotics. Such ‘aposymbiotic’ aphids grow very slowly and develop into adults that produce few or no offspring. The most fundamental function performed by the bacteria is to produce essential amino acids that are absent in phloem sap and antibiotic treatment confirms that the aphids cannot do this alone (Gündüz & Douglas, 2009). In addition, though, the *Buchnera* seem to provide other benefits, since symbiotic aphids still outperform aposymbiotic aphids when the latter are provided with all the essential amino acids, but establishing further nutritional functions has proved elusive.

... provide an ecological and evolutionary link

The aphid–*Buchnera* interaction also provides an excellent example of how an intimate association between mutualists may link them at both the ecological and the evolutionary level (Akman & Douglas, 2009). The bacterium *Buchnera aphidicola* is transmitted vertically by aphids such as *Acyrthosiphon pisum*, that is, they are passed by a mother to her offspring in her eggs. Hence, the aphid lineage supports a corresponding *Buchnera* lineage. Providing a final twist, the only other aphids without *Buchnera* (in the family Hormaphididae) appear to have lost them secondarily in their evolutionary history, but they do instead host symbiotic yeasts. It seems more likely that the yeasts competitively displaced the bacteria than that the bacteria were first lost and the yeasts subsequently acquired. Douglas (1998) also points out that whereas all plant-sucking hemipterans that feed on nutritionally deficient phloem sap have bacteriocyte symbioses, including the aphids described above, those that have switched secondarily in their evolutionary history to feeding on intact plant cells have lost the symbiosis. This, then, is an illustration from a comparative, evolutionary perspective that even in clearly mutualistic symbioses like these, the benefit is a *net* benefit. Once the insects’ requirements are reduced, as in a switch of diet, the balance of the costs and benefits of the symbionts is also changed. In this case, the costs clearly outweigh the benefits on a changed diet: those insects that *lost* their symbionts have been favoured by natural selection.

More generally, Baumann (2005) has described the phylogeny of four sap-sucking groups in the suborder Sternorrhyncha – aphids, psyllids, whiteflies and mealybugs – each of which has an obligatory endosymbiotic relationship with a different bacterium resulting from a single infection followed by maternal vertical transmission (Figure 13.14). Apparently related to *Buchnera*, the endosymbionts of three further groups diverged quite recently: *Baumannia* in the sharpshooters (which feed on xylem sap, of poorer quality than the phloem sap fed on by the sternorrhynchans), *Wigglesworthia* in tsetse flies (which feed on vertebrate blood) and *Biochmannia* in carpenter ants (generalist feeders). Thus, it seems that a single ancestor of this endosymbiont cluster gave rise to four types, each 'domesticated' by its host group to perform somewhat different functions.

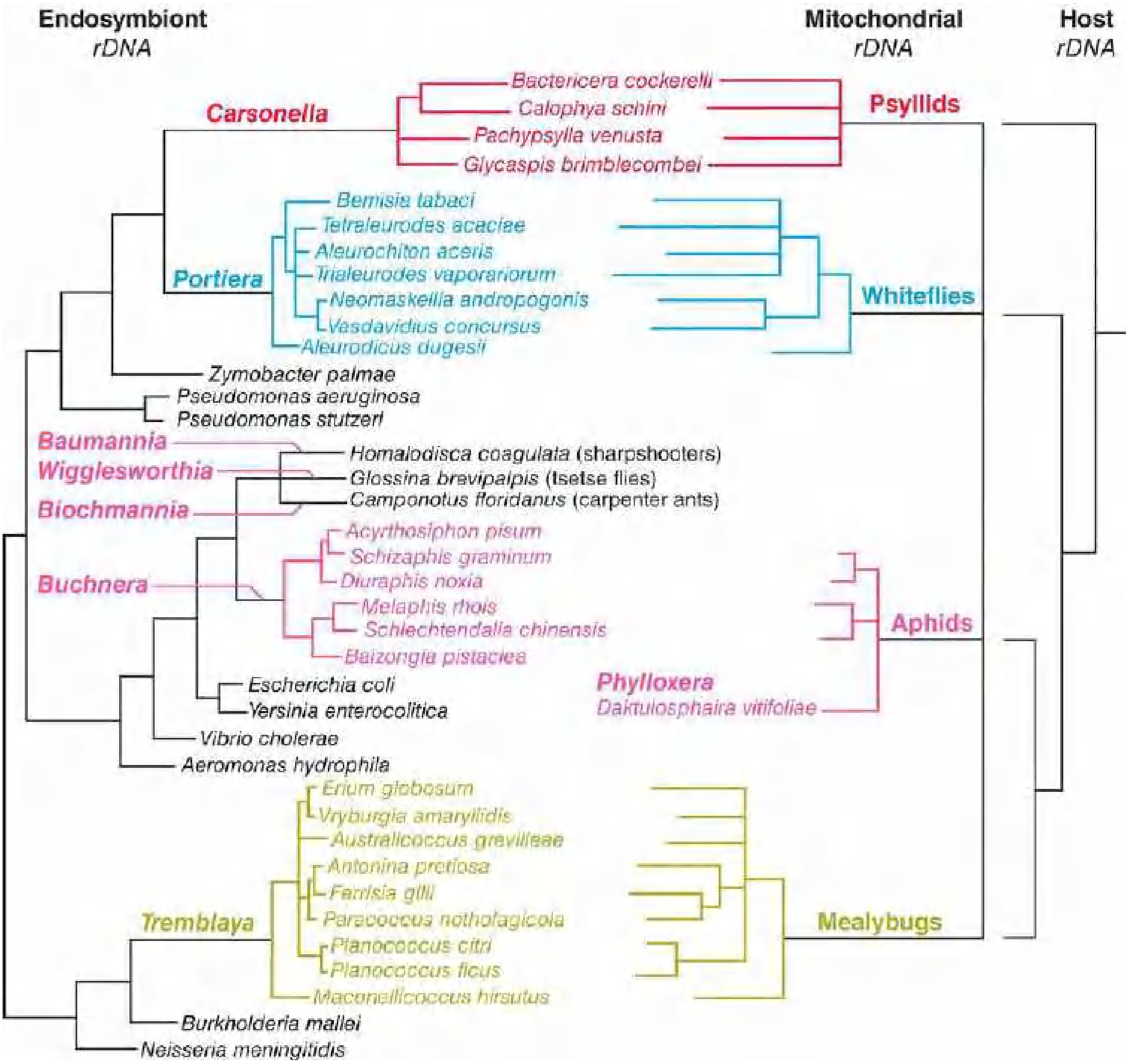


Figure 13.14 Each of several insect groups has evolved an obligatory relationship with a different endosymbiotic bacterium. Comparisons of evolutionary relationships of endosymbionts (*Carsonella*, *Portiera*, *Buchnera*, *Tremblaya*, *Baumannia*, *Wigglesworthia*, *Biochmannia*) of plant sap-sucking insects to the phylogeny of their hosts (psillids, whiteflies, aphids and mealybugs, based on rDNA (ribosomal DNA) sequence analysis), as well as to some free-living bacteria (*Zymobacter*, *Pseudomonas*, *Burkholderia*, *Neisseria*). Also shown are examples of other insects whose endosymbionts are closely related to those of aphids (sharpshooters, tsetse flies and carpenter ants). Note that *Phylloxera*, an aphid pest of grapes, has no endosymbionts. The relationships between the sap-sucking groups could not be resolved using their mitochondrial DNA, but were established on the basis of their nuclear DNA.

Source: From Baumann (2005).

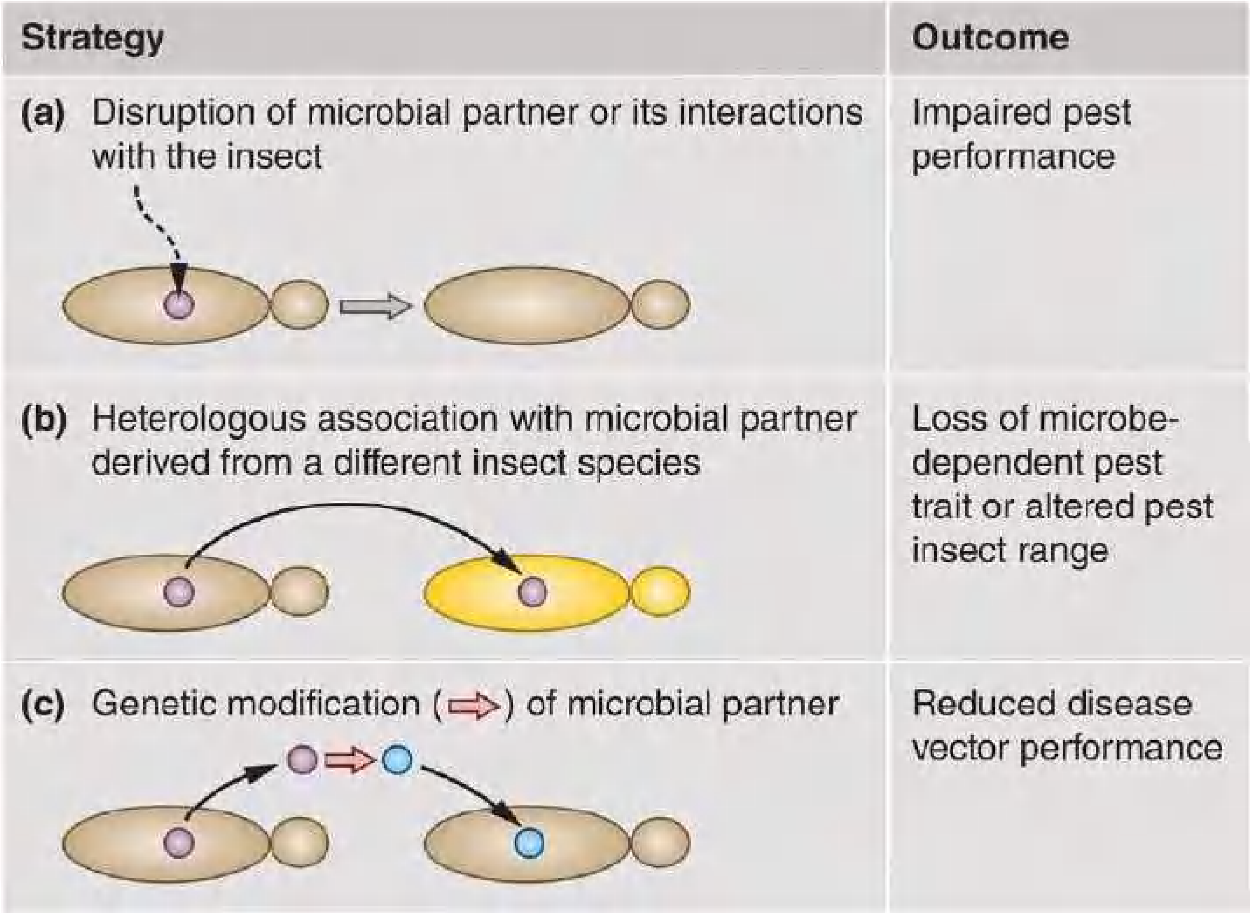
defensive symbioses

Some endosymbionts that live in specialised insect cells are now thought to serve a defence function (as do some counterparts in both vertebrate (Section 13.6.1) and insect guts). While much remains to be discovered about the molecular mechanisms, the toxic agents seem to be small molecules and protein toxins. For example, some rove beetles in the genus *Paederus*, as long as they harbour a bacterial endosymbiont in the genus *Pseudomonas*, carry pederin, a polyketide toxin, that appears to protect them from predation by wolf spiders (Van Arnam *et al.*, 2017). In like manner, individuals of the pea aphid, *Acyrthosiphon pisum,* that host the endosymbiont *Hamiltonella defensa* are protected against the specialist parasitoid wasp *Aphidius ervi.* It turns out that the protein toxins responsible for suppressing wasp development are produced by bacteriophages (viruses that infect bacteria) hosted by the endosymbiont (Oliver *et al.*, 2014).

APPLICATION 13.6 Novel uses of insect microbial symbionts to advance human welfare

The developing research area of insect defensive endosymbionts may prove important in the search for antibiotics effective in human health care. New antibiotics are sorely needed and underexplored niches such as the defensive microbes of insects hold considerable promise.

In another context, manipulation of endosymbiotic microbial partners can be expected to provide novel approaches for pest control by, for example, targeting the endosymbionts absolutely required by a pest insect in order to suppress growth or reproduction. An example is the introduction of an antimicrobial peptide isolated from the venom glands of a scorpion into the pea aphid: this was active against aphid bacterial symbionts, including *Buchnera*, and resulted in delayed reproduction and reduced aphid survival (Luna-Ramirez *et al.*, 2017). Strategies utilising heterologous microorganisms (derived from other insect species) are also under development, particularly in relation to insect vectors of human disease agents (Figure 13.15) (Arora & Douglas, 2017). For example, the mosquito *Aedes aegypti* does not naturally bear *Wolbachia*, a parasitic bacterium that infects many insects: its introduction to the mosquito from *Drosophila* fruit flies substantially reduced the acquisition of dengue virus by the mosquito (Hoffman *et al.*, 2011). An illustration of the use of genetic modification is transformation in the kissing bug (*Rhodnius prolixus*) of the gut bacterium *Rhodococcus rhodnii* to express an antimicrobial peptide (cecropin A) effective against the parasitic protozoan *Trypanosoma cruzi,* the agent of Chagas disease which is transmitted by kissing bugs (Hurwitz *et al.*, 2012).



Strategy	Outcome
(a) Disruption of microbial partner or its interactions with the insect	Impaired pest performance
(b) Heterologous association with microbial partner derived from a different insect species	Loss of microbe-dependent pest trait or altered pest insect range
(c) Genetic modification (⇨) of microbial partner	Reduced disease vector performance

Figure 13.15 Three strategies for novel insect pest control. Each is based on the manipulation of benign or beneficial bacterial endosymbionts, resulting in either suppression of pest traits or impaired insect performance.

Source: Modified from Arora & Douglas (2017).

13.8 Photosynthetic symbionts within aquatic invertebrates

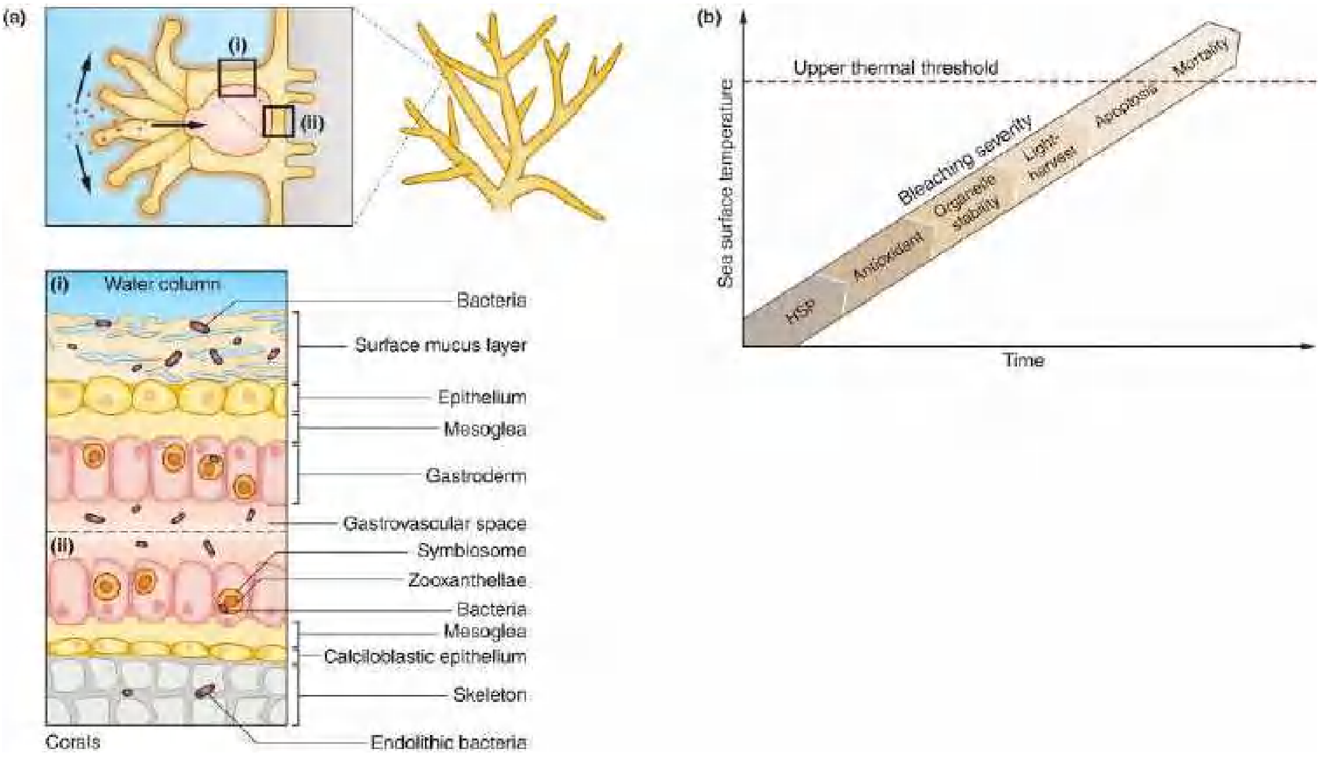
freshwater

Algae are found within the tissues of a variety of freshwater animals, including the cnidarian *Hydra viridis* and the protist *Paramecium bursaria*. In both these cases of facultative photosymbiosis, the host contains the algal endosymbiont *Chlorella* sp. in host-derived vacuoles called symbiosomes (analogous to bacteriocytes, Section 13.7). The host provides its symbionts with nitrogen compounds derived from heterotrophy (feeding on bacteria and other organic food sources), while the endosymbionts provide hosts with maltose and oxygen derived from autotrophy (photosynthesis). It is clear that there must be regulatory processes harmonising the growth of an endosymbiont and its host, otherwise the symbionts would either overgrow and kill the host or fail to keep pace and become diluted as the host grew.

The question of whether such endosymbioses are founded on mutualism or exploitation was addressed by Lowe *et al.* (2016), who compared the fitness effects of manipulating the supplies of light (influencing autotrophy) and bacteria (influencing heterotrophy) on *Paramecium* and *Chlorella* in the symbiotic state and the autonomous state (each living freely without the other). The hosts benefited from the symbiosis at high light intensities (increased intrinsic growth rate) but carrying symbionts was costly in the dark (reduced intrinsic growth rate compared with the autonomous state). The densities of *Chlorella* in the free-living autonomous state increased with light intensity, but, critically, their density did not increase with high light intensity in the symbiosomes. Host energetic requirements can be met by fewer symbionts at high light levels, and the results suggest that the host exerts strong control over endosymbionts, most likely by restraining algal cell division by limiting metabolite supply. Photosynthetic efficiency was consistently lower for symbiotic compared with autonomous *Chlorella*, again indicating nutritional constraints on them in symbiosis. In their experiments, Lowe *et al*. (2016) found no evidence that this nutritional endosymbiosis was mutually beneficial and argue that ‘controlled exploitation’ is an important evolutionary pathway toward stable endosymbiosis.

shallow tropical oceanic waters

Symbioses between invertebrates and algae are especially common in shallow tropical waters, where water clarity and low nutrient concentrations provide maximal benefit to the association. Many involve the dinoflagellate alga *Symbiodinium* and coral or giant clam hosts. The algae occur within symbiosomes in corals (Figure 13.16) and are present extracellularly in tridacnid clams. For algal autotrophy to contribute to host growth, the host must transport inorganic carbon (CO_2), and suitable sources of nitrogen (NH_3, NO_{3-} or N_2), phosphate and other inorganic nutrients through its tissues to the algae (Yellowlees *et al*., 2008). (In a sense, the algae might be considered to be providing a service to the host by removing some metabolic breakdown products.) The algal photosynthetic products are then exported back to the host to supplement its energy and nutrient requirements, providing up to 90% of the coral's energy needs (Iluz & Dubinsky, 2015).



[Figure 13.16](#) **The coral microbiome, consisting of algal and bacterial symbionts, responds to thermal stress by a process of coral bleaching.** (a) The coral microbiome, consisting of mutualistic algae (*Symbiodinium*) and up to 1000 bacterial species in three microhabitats: the surface mucus layer, coral tissue (gastroderm) and the coral skeleton. Bacteria also occur in the gastrovascular space. Regions i and ii above are expanded in the lower panel. (b) Generalised model of coral–*Symbiodinium* response to thermal stress. As the mutualistic association is pushed towards upper thermal thresholds (for the particular species concerned), heat shock proteins (HSP) and antioxidant functioning are first affected, then organelle stability, the ability to capture light and finally apoptosis (cell death), with bleaching becoming increasingly severe.

Source: (a) From Hernandez-Agreda *et al.* (2017). (b) From Suggett *et al.* (2017).

the coral holobiont

In fact, the coral holobiont (the host plus all its microbial symbionts) is not made up only of symbiotic algae. Each coral species also harbours around a thousand bacterial OTUs (operational taxonomic units, roughly equivalent to species) in three microhabitats – the surface mucus layer, the coral tissue (where the algal symbiosomes occur) and the coral skeleton – comprising whole communities of potential mutualists. It is likely that many of the bacteria contribute to the host's health and nutrition and to nutrient cycling in the coral. For example, bacteria in the coral's surface mucus layer are thought to be important in the provision of antibiotic activity and may help prevent pathogen invasion, while diazotrophic (nitrogen fixing) bacteria have also been consistently found in coral tissue, skeleton and mucus (order Rhizobiales, family Cyanobacterium), while members of the phylum Actinobacteria have been found within the algal symbiosomes themselves (Hernandez-Agreda *et al.*, 2016). Sponges provide a further example of a complex holobiont, comprising at least 39 microbial phyla, dominated by bacteria but including photoautotrophs (Pita *et al.*, 2018).

coral bleaching and global warming

The symbiosis between coral and *Symbiodinium* disassociates when surface seawater temperature exceeds the long-term maximum monthly mean for extended durations, leading to rapid loss of *Symbiodinium* cells and/or their pigmentation and conspicuous whitening of coral tissues (Figure 13.16). In the light of climate change (Section 2.9), such coral bleaching has become a global concern (Application 22.3). Susceptibility to high temperatures varies among coral and *Symbiodinium* species, resulting in complex bleaching mosaics across affected reefs. It is likely that the responses of corals to global warming will also be modulated by temperature effects on bacterial species in the coral microbiome.

another mutualism extending beyond two species

Meanwhile, another study of coral bleaching adds to the growing realisation that seemingly simple two-species mutualisms may be more complex and subtler than might be imagined. The ecologically dominant Caribbean corals *Montastraea annularis* and *M. faveolata* both host three quite separate 'species' or 'phylotypes' of *Symbiodinium* (denoted *A*, *B* and *C* and distinguishable only by genetic methods). Phylotypes *A* and *B* are common in shallower, high-irradiance habitats, whereas *C* predominates in deeper, lower irradiance sites – illustrated both by comparisons of colonies from different depths and of samples from different depths within a colony (Rowan *et al.*, 1997). In the autumn of 1995, following a prolonged period above the mean maximum summer temperature, bleaching occurred in *M. annularis* and *M. faveolata* in the reefs off Panama and elsewhere. Bleaching, however, was rare at the shallowest and the deepest sites, and resulted from the selective loss of *Symbiodinium C* at intermediate sites. At shaded deep-water sites, dominated by *C*, the high temperatures in 1995 were not sufficient to push *C* into bleaching conditions. The shallowest sites were occupied by the species *A* and *B*, which were not susceptible to bleaching at these temperatures. Bleaching occurred, however, where *C* was initially present but was pushed beyond its limit by the increased temperature. At these sites, the loss of *C* was typically close to 100%, *B* decreased by around 14%, but *A* more than doubled in three of five instances.

It seems, therefore, firstly, that the coral–*Symbiodinium* mutualism involves a range of endosymbionts that allows the corals to thrive in a wider range of habitats than would otherwise be possible. Second, looking at the mutualism from the algal side, the endosymbionts must constantly be engaged in a competitive battle, the balance of which alters over space and time. Finally, bleaching, and any subsequent recovery, may be seen as manifestations of this competitive battle: not breakdowns and reconstructions in a simple two-species association, but shifts in a complex symbiotic community.

13.9 Mutualisms involving higher plants and fungi

not roots but mycorrhizas

Most higher plants do not have roots, they have mycorrhizas – intimate mutualisms between root tissue and fungi, which live inside the cortex of plant roots, on the root surface or around the root's epidermal cells, with hyphae growing out into the soil where they forage for plant nutrients (van der Heijden *et al.*, 2015). About 50 000 fungal species form mycorrhizal associations with about 250 000 plant species, and plants of only a few families lack them, including the Brassicaceae and Proteaceae (Table 13.2). Broadly, the fungal networks in mycorrhizas capture nutrients from the soil, supplying up to 80% of plant nitrogen and phosphorus, which they transport to the plants in exchange for carbohydrates. Many plant species can live without their mycorrhizal fungi in soils where neither nutrients nor water are ever limiting, but in the harsh world of natural plant communities, the symbioses, if not strictly obligate, are nonetheless 'ecologically obligate'. That is, they are necessary if the individuals are to survive in nature. The fossil record suggests that the earliest land plants, too, were heavily infected. These species lacked root hairs, even roots in some cases, and the early colonisation of the land may have depended on the presence of the fungi to make the necessary intimate contact between plants and substrates.

Table 13.2 **Numbers of plant and fungal species forming arbuscular mycorrhiza, ectomycorrhizal, orchid mycorrhizal and ericoid mycorrhizal associations.** The distinction between mycorrhizal categories is not necessarily strict, with some plant species, including poplar and eucalypt trees forming dual associations with both arbuscular and ectomycorrhizal fungi.

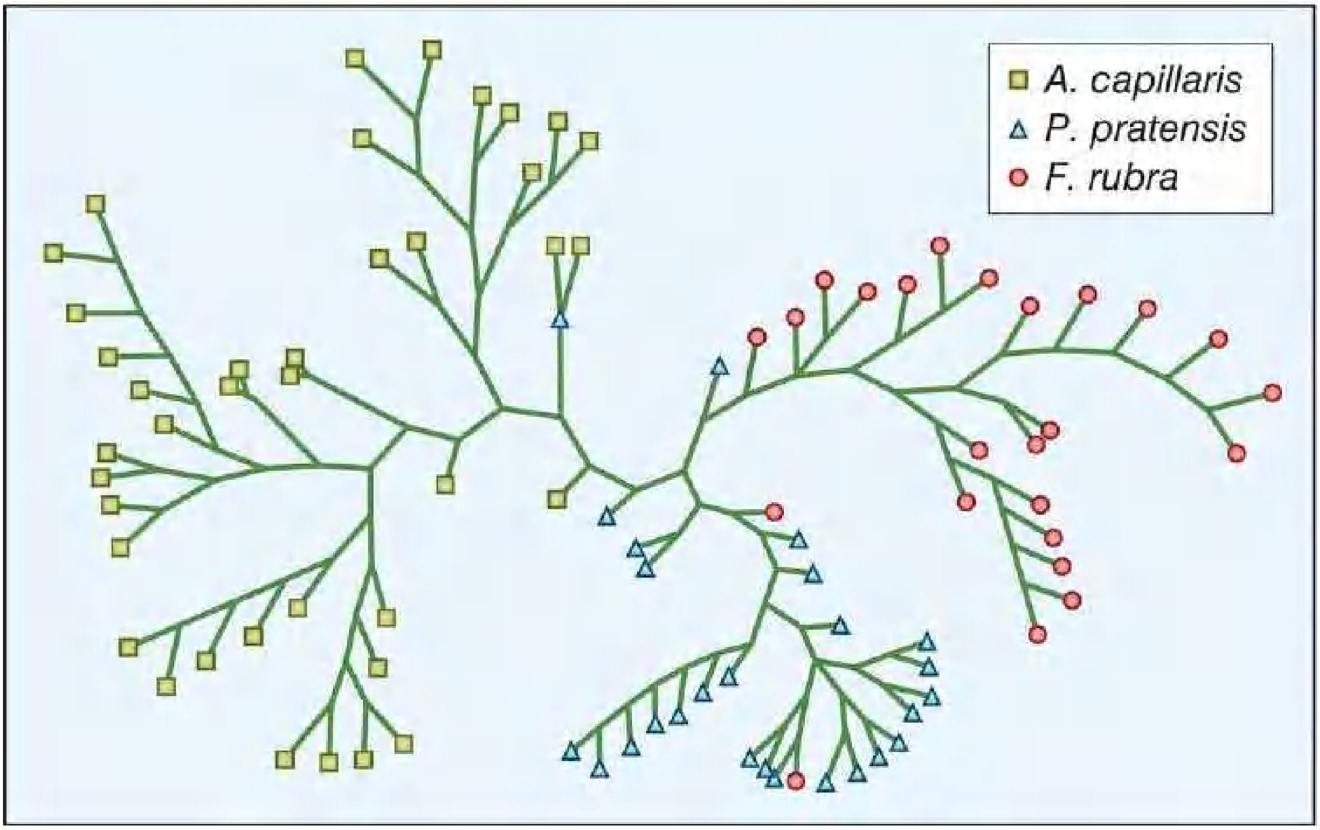
Source: From van der Hiejden *et al.* (2015).

Mycorrhizal type	Major groups of plants	Number of plant species	Fungal identity	Total estimated number of fungal taxa
Arbuscular mycorrhiza	Most herbs, grasses and many trees, many hornworts and liverworts	200 000	Glomeromycota	300–1600
Ectomycorrhiza	Pinaceae and angiosperms (mostly shrubs and trees, mostly temperate), some liverworts.	6000	Basidiomycota and Ascomycota	20 000
Ericoid mycorrhiza	Members of the Ericaceae, some liverworts	3900	Mainly Ascomycota, some Basidiomycota	>150
Orchid mycorrhiza	Orchids	20 000–35 000	Basidiomycota	25 000
Non-mycorrhizal plant species	Brassicaceae, Crassulaceae, Orobanchaceae, Proteaceae, etc.	51 500		0

Four major types of mycorrhiza are recognised. Arbuscular mycorrhizas are found in almost three quarters of all plant species, including most non-woody species and tropical trees. Ectomycorrhizal fungi form symbioses with many trees and shrubs, dominating boreal and temperate forests and also some tropical rainforests. Orchid mycorrhizas are associated with orchids and, finally, ericoid mycorrhizas are found in the dominant species of heathlands including the northern hemisphere heaths and heathers (Ericaceae) and the Australian heaths (Epacridaceae).

13.9.1 Arbuscular mycorrhizas

Arbuscular mycorrhizas (AMs) penetrate *within* the roots of the host, though they do not alter the host's root morphology. Roots become infected from mycelium present in the soil or from germ tubes that develop from asexual spores, which are very large and produced in small numbers. Initially, the fungus grows between host cells but then enters them and forms a finely branched intracellular 'arbuscule'. The fungi responsible comprise a distinct phylum, the Glomeromycota. This consists of only 300–1600 species, which might suggest a lack of host specificity (since there are vastly more species of hosts); however, genetic studies have provided evidence of niche differentiation amongst them. For instance, when 89 root samples were taken from three grass species that co-occurred in the same plots in a field experiment, and their AM fungi were characterised using such a method – terminal restriction fragment length polymorphism – there was clear separation amongst the AM strains found on the different hosts (Figure 13.17).



[Figure 13.17](#) **There is a clear separation of arbuscular mycorrhizal fungal strains according to the plants they infect.** The similarity among 89 arbuscular mycorrhiza (AM) fungal communities taken from the roots of three coexisting grass species, *Agrostis capillaris*, *Poa pratensis* and *Festuca rubra*, assessed by terminal restriction fragment length polymorphism. Each terminal on the 'tree' is a different sample, with the grass species from which it originated shown. More similar samples are closer together on the tree. The similarity within, and the differentiation between, the AM fungal communities associated with different hosts are plainly apparent.

Source: After Vandenkoornhuyse *et al*. (2003).

a range of benefits?

There has been a tendency to emphasise facilitation of the uptake of phosphorus as the main benefit to plants from AM symbioses (phosphorus is a highly immobile element in the soil, which is therefore frequently limiting to plant growth), but the truth appears to be more complex than this. Benefits have been demonstrated, too, in nitrogen uptake (Smith & Smith, 2011), herbivore protection (Sharma *et al*., 2017) and resistance to toxic metals (Ferrol *et al*., 2016).

13.9.2 Ectomycorrhizas

An estimated 20 000–25 000 species of Basidiomycete and Ascomycete fungi form ectomycorrhizas (ECMs) on the roots of trees and woody shrubs (Kumar & Atri, 2018). Infected roots are usually concentrated in the litter layer of the soil. Fungi form a sheath or mantle of varying thickness around the roots. From there, hyphae radiate into the litter layer, extracting nutrients and water and also producing large fruiting bodies (toadstools) that release enormous numbers of wind-borne spores – a striking contrast with the arbuscular fungi. The fungal mycelium also extends inwards from the sheath, penetrating between the cells of the root cortex to give intimate cell-to-cell contact with the host and establishing an interface with a large surface area for the exchange of photo-assimilates, soil water and nutrients between the host plant and its fungal partner. The fungus usually induces morphogenetic changes in the host roots, which cease to grow apically and remain stubby. Host roots that penetrate into the deeper, less organically rich layers of the soil continue to elongate.

The ECM fungi enhance water absorption and are effective in extracting the sparse and patchy supplies of phosphorus and especially nitrogen from the forest litter layer. A few fungi are host-specific, but most are associated with multiple hosts: the plants are also usually associated with multiple fungi. Carbon flows from the plant to the fungus, very largely in the form of the simple hexose sugars: glucose and fructose. Fungal consumption of these may represent up to 30% of the plants' net rate of photosynthate production. The plants, though, are often nitrogen-limited, since in the forest litter there are low rates of nitrogen mineralisation (conversion from organic to inorganic forms), and inorganic nitrogen is itself mostly available as ammonia. It is therefore crucial for forest trees that ECM fungi can access organic nitrogen directly through enzymic degradation, utilise ammonium as a preferred source of inorganic nitrogen, and circumvent ammonium depletion zones through extensive hyphal growth. Nonetheless, the idea that this relationship between the fungi and their host plants is mutually exploitative is emphasised by its responsiveness to changing circumstances. ECM growth is directly related to the rate of flow of hexose sugars from the plant. But when the direct availability of nitrate to the plants is high, either naturally or through artificial supplementation, plant metabolism is directed away from hexose production (and export) and towards amino acid synthesis. As a result, the ECM degrades; the plants seem to support only as much ECM as they appear to need.

13.9.3 Ericoid mycorrhizas

Heathlands exist in environments characterised by soils with low levels of available plant nutrients, often as a result of regular fires in which, for example, up to 80% of the nitrogen that has accumulated between fires may be lost. It is unsurprising, therefore, that heathlands are dominated by many plants that have evolved an association with ericoid mycorrhizal fungi. This enables them to facilitate the extraction of nitrogen and phosphorus from the superficial layers of detrital material generated by the plants. Indeed, the conservation of natural heathlands is threatened now by nitrogen supplementation and fire control, which allow colonisation and domination by grasses that would otherwise be unable to exist in these impoverished environments.

The ericoid mycorrhizal root itself is anatomically simple compared to other mycorrhizas, characterised by a loose net of external hyphae that extend along, between and inside the epidermal cells, without penetrating the membrane, to form coils at different densities within the cells of fine 'hair roots' that have an ephemeral lifespan (Vohnik *et al.*, 2012). As a result, the individual hair roots are delicate structures that collectively form a dense fibrous root system, the bulk of which is concentrated towards the surface of the soil profile. The fungi are effective, unlike the plants alone, in absorbing nitrate, ammonium and phosphate ions that have been mobilised by other decomposers in the soil, but crucially they are also 'saprotrophic'. They are therefore able to compete directly with the other decomposers in liberating nitrogen and phosphorus from the organic residues in which most of these elements are locked up in heathland ecosystems. A mutualism can thus be seen, again, to be woven into a larger web of interactions: the symbiont enhances its contribution to the host by making a pre-emptive competitive strike for scarce inorganic resources, and its own competitive ability is presumably enhanced in turn by the physiological support provided by its host.

13.9.4 Orchid mycorrhizas

Fungi that form mycorrhizas with orchids typically live as saprotrophs in the soil or form ECMs with nearby trees (Dearnaley *et al.*, 2013). Orchid seedlings are devoid of chlorophyll and depend entirely for their nutrients and carbohydrates on colonisation by mycorrhizal fungi. Many orchids have between one and 10 host-specific fungal associates, and more than 20 000 fungal species have been identified (Table 13.2).

13.9.5 Mycorrhizal networks

Because most plant roots are colonised by several mycorrhizal fungi and most of these fungi are not host specific, colonising various host plants at the same time, plants are usually interconnected by mycorrhizal mycelium networks (Figure 13.18). Intriguingly, the existence of mycorrhizal networks implies that nutrients and photosynthate may be transferred from plant to plant via the fungal hyphae, and some research suggests that chemical signals could be transmitted via the network, perhaps prompting neighbouring plants to respond to threats from herbivores or pathogens (Babikova *et al.*, 2013). But this question is far from resolved.

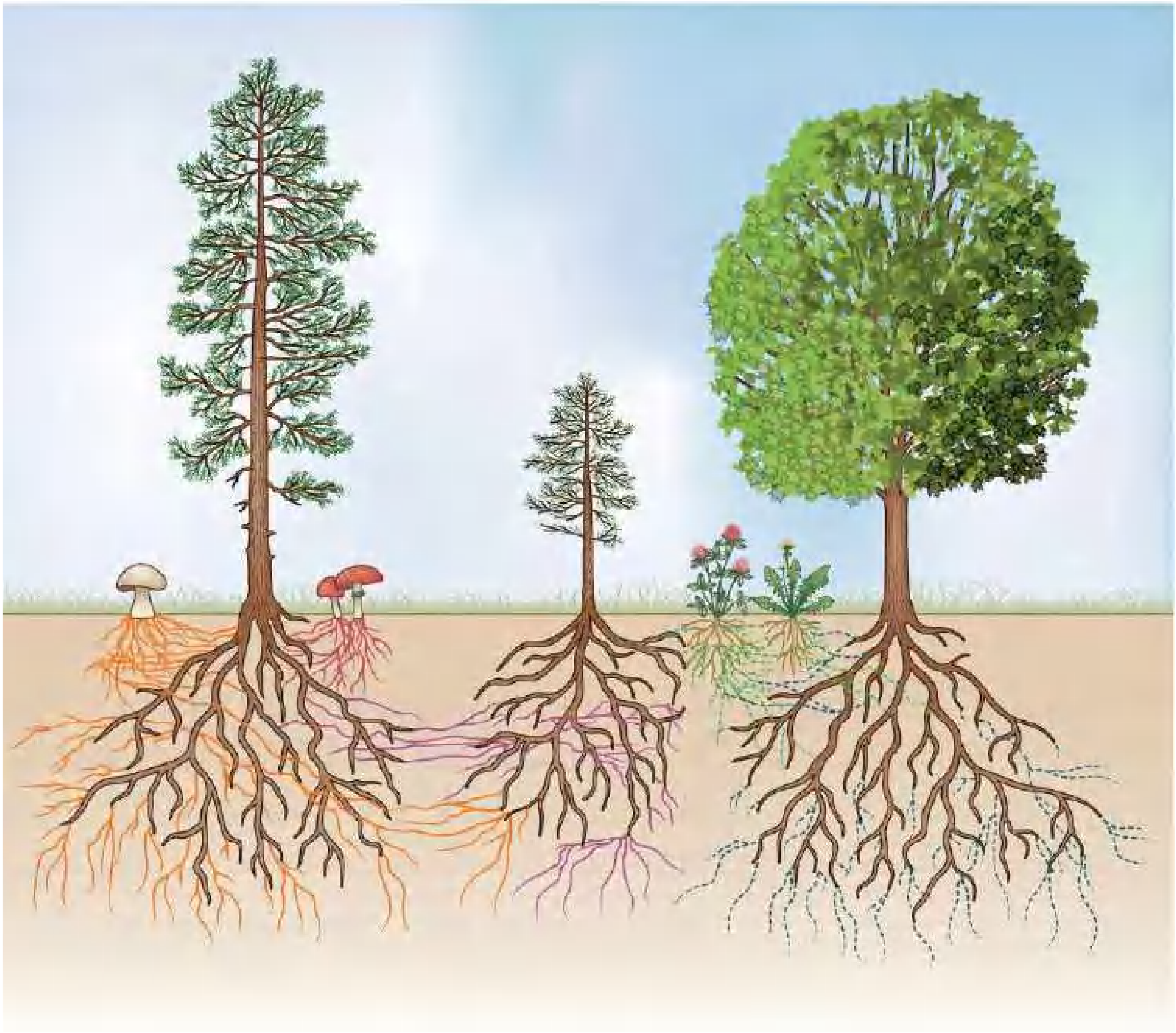


Figure 13.18 Plants will often be connected to each other by mycorrhizal networks. To the left, a pair of trees form a network (solid lines) with various species of ectomycorrhizal fungi (different colours). Typical fungal fruiting bodies are also shown (toadstools). To the right, herbaceous species and a tree form a network (dotted lines) with different species of arbuscular mycorrhizal fungi. Other potential combinations may also occur, involving ericoid or orchid species with neighbouring trees.

Source: Modified from van der Hiejden *et al.* (2015).

13.10 Fungi with algae: the lichens

mycobionts and photobionts

Of the nearly 100 000 species of fungus that are known (of the many million that undoubtedly exist; Blackwell, 2011), approximately 20 000 are ‘lichenised’. Lichens are nutritionally specialised fungi (the so-called ‘mycobiont’ component) that have escaped from their normal way of life into a mutualistic association with a ‘photobiont’. In around 90% of lichen species the photobiont is an alga, which provides carbon compounds to the mycobiont through photosynthesis. In some cases, the photobiont is a cyanobacterium, which may also provide fixed nitrogen to the association. In a relatively few, ‘tripartite’ lichen species both an alga and a cyanobacterium are involved. Lichenised fungi belong to diverse taxonomic groups (predominantly in the Ascomycota, with six main classes, and a few from the Basidiomycota) and the mutualistic algae to about 120 species (plus a few cyanobacteria) (Grube & Wedin, 2016).

The photobionts are located extracellularly between the fungal hyphae, in a thin layer near the upper surface. Together, the two components form an integrated ‘thallus’ but the photobiont makes up only about 3–10% by weight. The advantage to the photobiont in the association, if any, has not been established clearly. All lichenised algal species, for example, can also occur free-living outside their association with their mycobiont. It may be that they are ‘captured’ by the fungus and exploited without any recompense. However, some of the species (e.g. in the algal genus *Trebouxia*) are rare in their free-living form but very common in lichens, suggesting that there is something special about life in their mycobiont that they need. Moreover, since minerals, including nitrogen, are largely ‘captured’ from what is deposited directly onto the lichen, often in rainwater and from the flow and drip down the branches of trees, and since the surface and biomass are largely fungal, the mycobiont must contribute the vast bulk of these minerals.

parallels with higher plants

Hence, the mutualistic pairs (and threesomes) in lichens provide two striking parallels with higher plants. There is a structural parallel: in plants, the photosynthetic chloroplasts are similarly concentrated close to light-facing surfaces. There is also a functional parallel. The economy of a plant relies on carbon produced largely in the leaves and nitrogen absorbed principally through the roots, with a relative shortage of carbon giving rise to shoot growth at the expense of roots, and a shortage of nitrogen leading to root growth at the expense of shoots. Likewise, in lichens, the synthesis of carbon-fixing photobiont cells is inhibited by a relative shortage of nitrogen in the mycobiont, but synthesis is stimulated when the carbon supply is limiting (Palmqvist, 2000).

Lichenisation, then, gives the mycobiont and the photobiont, between them, the functional role of higher plants, but in so doing it extends the ecological range of both partners onto substrata (rock surfaces, tree trunks) and into regions (arid, arctic and alpine) that are largely barred to higher plants. Indeed, it has been claimed that lichens dominate 8–10% of terrestrial land area, both in terms of abundance and species diversity, and their biomass on tree trunks is equivalent to a few percent, but in some cases up to 60%, of tree leaf biomass (Ellis, 2012). However, all lichens grow slowly: the colonisers of rock surfaces rarely extend faster than 1–5 mm year^{-1}. They are, though, very efficient accumulators of the mineral cations that fall or drip onto them, and this makes them particularly sensitive to environmental contamination by, for example, heavy metals and fluoride. Hence, they are amongst the most sensitive indicators of environmental pollution. The ‘quality’ of an environment in humid regions can be judged rather accurately from the presence or absence of lichen growth on tombstones and tree trunks.

remarkable morphological responses on the fungi

One remarkable feature in the life of the lichenised fungi is that the growth form of the fungus is usually profoundly changed when the alga is present. When the fungi are cultured in isolation from the algae, they grow slowly in compact colonies, much like related free-living fungi; but in the presence of the algal symbionts they take on a variety of morphologies that are characteristic of specific algal–fungal partnerships, including fruticose (shrub-like), foliose (leaf-like) and crustose forms (Figure 13.19). In fact, the algae stimulate morphological responses in the fungi that are so precise that the lichens have been classified as distinct species, and a cyanobacterium and an alga, for example, may elicit quite different morphologies from the same fungus.

Figure 13.19 Contrasting shapes of lichen thalli. (a) Coral-like fruticose (shrub-like) thallus of *Cladia retipora*. (b) Crustose (crust-like) thallus of *Acarospora fuscata* growing on a siliceous rock. (c) Foliose (leaf-like) thalli of *Lobaria pulmonaria* and *Lobarina scrobiculata* covering the trunk of a *Salix caprea* in an old-growth spruce-dominated forest.

Source: From Grube & Wedin (2016). Photos by: (a) Birgitta Strömbäck; (b) Martin Westberg; and (c) Mats Wedin.

diverse communities within lichens

Just as other intimate symbioses turn out to involve more than just the focal species (e.g. the coral holobiont in Section 13.8), the lichen thallus is now known to contain a range of other fungal species (more than 1800 species of lichenicolous fungi – not the mycobiont itself – have been described) and bacterial species, whose influence on the lichen is not fully understood (Grube & Wedin, 2016).

APPLICATION 13.7 A role for lichens in medicine

Lichens are very slow growing and live in low-resource habitats, two circumstances that may favour the production of higher levels of defence chemicals than faster growing counterparts (Coley, 1988). Certainly, they are known to produce more than 1050 secondary metabolites whose roles in nature include defence against bacteria, non-lichenised fungi, animal predators such as nematodes and plant competitors, as well as protection against environmental stresses such as UV radiation and dessication. These secondary metabolites are produced exclusively by the fungal partner (the mycobiont) and they have generated considerable interest for potential roles in medicine.

lichens as a source of antibiotics

Just like the insect defensive endosymbionts discussed in Application 13.6, lichens provide considerable promise in the search for new antibiotics. Lichen substances, including usnic acid, lichesterinic acid and depsidones, have demonstrable antibiotic properties, showing potency even against various multidrug resistant strains (Shrestha & St Clair, 2013). More than half of the lichens so far tested have shown antibiotic activity but the group remains largely unexplored.

lichens as a source of anticancer drug therapies?

Furthermore, a variety of studies have shown that crude lichen extracts and certain isolated lichen compounds demonstrate significant inhibitory activity against various cancer cell lines at very low concentrations (Table 13.3). The mechanisms have not been well characterised but the effects of lichen compounds include cell-cycle arrest, cell death and inhibition of blood vessel formation. Shrestha and St Clair (2013) highlight the urgent need for further research, including identification of specific mechanisms of action, clinical trials of the most promising lichen-based drug therapies, and development of large-scale production processes.

Table 13.3 Selected examples of anticancer activity of lichen secondary metabolites. *Source*: From Strestha & St Clair (2013), where original references are provided.

Lichen species	**Human cancer cell line**	**Major finding**
Hypogymnia physodes	Breast cancer (MCF-7)	Methanolic extract reduced cell viability in a dose-dependent manner
Lethariella zahlbruckneri	Colon cancer (HT-29)	Acetone extract induced cell death in a dose-dependent manner
Myelochroa aurulenta (leucotylic acid)	Leukaemia (HL-60)	Potent antiproliferative activity
Sphaerophorus globosus (pherophorin)	Melanoma (M-14)	Inhibitory effect by inducing apoptotic cell death

13.11 Fixation of atmospheric nitrogen in mutualistic plants

The inability of most plants and animals to fix atmospheric nitrogen is one of the great puzzles in the process of evolution, since nitrogen is in limiting supply in many habitats. However, the ability to fix nitrogen is widely though irregularly distributed amongst most phyla of bacteria and the methanogenic archaea (Figure 13.20), and many of these have been caught up in tight mutualisms with systematically quite different groups of eukaryotes. Presumably such symbioses have evolved a number of times independently. These symbioses can be of enormous ecological importance through their effects on individual plants and on primary productivity overall.

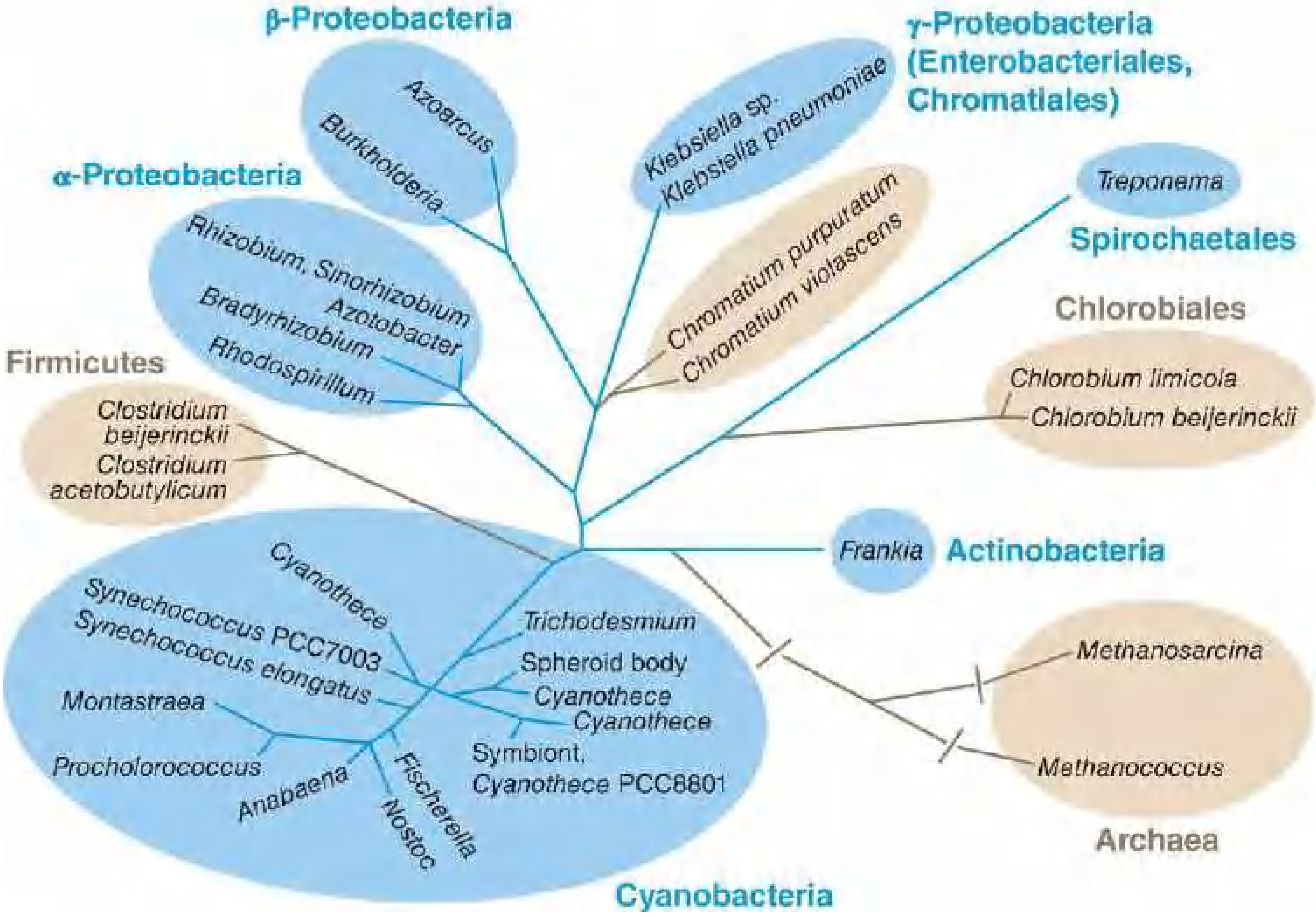


Figure 13.20 Phylogenetic affinities of nitrogen-fixing bacteria and archaea. Affinities of symbiotic (blue elipses) and undetermined or non-symbiotic (pale brown elipses) nitrogen-fixing bacteria and archaea.

Source: From Kneip *et al.* (2007).

the range of nitrogen-fixing bacteria

The nitrogen-fixing bacteria that have been found in symbioses (not necessarily mutualistic) are members of the following taxa (Franche *et al.*, 2009).

1. Rhizobia, which fix nitrogen in the root nodules of most leguminous plants (plus just one non-legume, *Parasponia*, a member of the family Ulmaceae, the elms). These are members of the alpha-Proteobacteria (Figure 13.20).
2. Azotobacteriaceae (beta-Proteobacteria), which can fix nitrogen aerobically and are commonly found on leaf and root surfaces, such as *Azotobacter vinelandii*, and Spirillaceae (beta-Proteobacteria), such as *Azospirillum lipoferum*, which is an obligate aerobe found on grass roots. In their interactions with plants, these are referred to as associative nitrogen fixers.

3. Enterobacteriaceae (gamma-Proteobacteria), such as *Enterobacter*, which occur regularly in gut microbiotas (e.g. of termites) and occasionally as associative nitrogen fixers on leaf surfaces and on root nodules.
4. Actinomycetes (Actinobacteria) of the genus *Frankia*, which fix nitrogen in the nodules (actinorhiza) of a number of non-leguminous and mainly woody plants, such as alder (*Alnus*) and sweet gale (*Myrica*).
5. Cyanobacteria of the family Nostocaceae, which are found in association with a remarkable range (though rather few species) of flowering and non-flowering plants and also are photobionts in some lichens (Section 13.10).

Of these, the association of the rhizobia with legumes is the most thoroughly studied, because of the huge agricultural importance of legume crops.

13.11.1 Mutualisms of rhizobia and leguminous plants

several steps to a liaison

The establishment of a liaison between rhizobia and legume plants proceeds by a series of reciprocating steps. The bacteria occur in a free-living state in the soil and are stimulated to multiply by root exudates – a cocktail of phenolic molecules, predominantly flavonoids. These are responsible for switching on a complex set of genes in the rhizobia (*nod* genes) that control the process that induces nodulation in the roots of the host. Each rhizobial species/strain interacts with a specific group of legumes and vice versa – a specificity that is partly determined because the *nod* genes of particular rhizobial species are switched on by particular sets of legume flavonoids, and partly by the differential capacity of particular *nod* factors and bacterial surface polysaccharides to suppress the defence responses of the legume (Wang *et al*., 2012). In a typical case, a bacterial colony develops on the root hair, which then begins to curl and is penetrated by the bacteria. The host responds by laying down a wall that encloses the bacteria and forms an ‘infection thread’, within which the rhizobia proliferate extracellularly. This grows within the host root cortex, and the host cells divide in advance of it, beginning to form a nodule. Rhizobia in the infection thread cannot fix nitrogen, but some are released into the host meristem cells. There, surrounded by a host-derived peribacteroid membrane, they differentiate into ‘bacteroids’ that can fix nitrogen. In some species, those with ‘indeterminate’ growth like the rhizobia of the pea (*Pisum sativum*), the bacteroids themselves are unable to reproduce further. Only undifferentiated rhizobia are released back into the soil to associate with another root when the original root senesces. By contrast, in species with ‘determinate’ growth like those of the soybean (*Glycine max*), bacteroids survive root senescence and can then invade other roots (Kiers *et al*., 2003).

A special vascular system develops in the host, supplying the products of photosynthesis to the nodule tissue and carrying away fixed nitrogen compounds (very often the amino acid asparagine) to other parts of the plant (Figure 13.21). The nitrogen-fixing nitrogenase enzyme accounts for up to 40% of the protein in the nodules and depends for its activity on a very low oxygen tension. A boundary layer of tightly packed cells within the nodule serves as a barrier to oxygen diffusion. A haemoglobin (leghaemoglobin) is formed within the nodules, giving the active nodules a pink colour. It has a high affinity for oxygen and allows the symbiotic bacteria to respire aerobically in the virtually anaerobic environment of the nodule. Indeed, wherever nitrogen-fixing symbioses occur, at least one of the partners has special structural (and usually also biochemical) properties that protect the anaerobic nitrogenase enzyme from oxygen, yet allow normal aerobic respiration to occur around it.

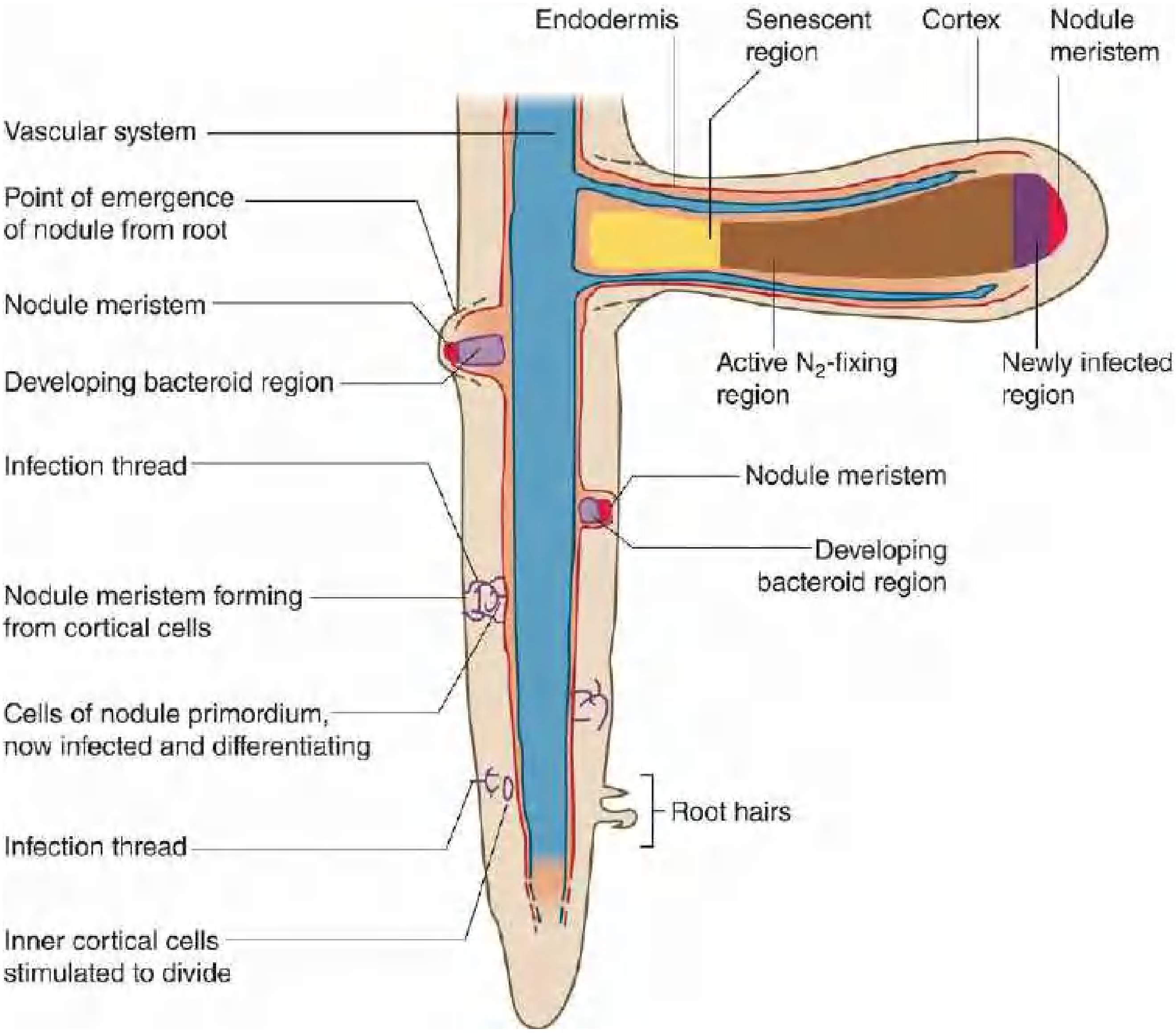


Figure 13.21 **Development of the root nodule**. Structural changes during the course of development of infection of a legume root by *Rhizobium*.

Source: After Sprent (1979).

costs and benefits of rhizobial mutualisms

The costs and benefits of this mutualism need to be considered carefully. From the plant's point of view, we need to compare the energetic costs of alternative processes by which supplies of fixed nitrogen might be obtained. The route for most plants is direct from the soil as nitrate or ammonium ions. The metabolically cheapest route is the use of ammonium ions, but in most soils ammonium ions are rapidly converted to nitrates by microbial activity (nitrification). The energetic cost of reducing nitrate from the soil to ammonia is about 12 mol of adenosine triphosphate (ATP) per mole of ammonia formed. The mutualistic process (including the maintenance costs of the bacteroids) is energetically slightly more expensive to the plant: about 13.5 mol of ATP. However, to the costs of nitrogen fixation itself we must also add the costs of forming and maintaining the nodules, which may be about 12% of the plant's total photosynthetic output. It is this that makes nitrogen fixation energetically inefficient. Energy, though, may be much more readily available for green plants than nitrogen. A rare and valuable commodity (fixed nitrogen) bought with a cheap currency (energy) may be no bad bargain. On the other hand, when a nodulated legume is provided with nitrates (i.e. when nitrate is not a rare commodity) nitrogen fixation declines rapidly. And in previously N-limited ecosystems whose soil nitrate has been enhanced by nitrogen pollution (acid rain; Section 2.8), the numbers and sizes of root nodules, and the benefits of rhizobial symbiosis decline dramatically (Regus *et al.*, 2017).

The benefits to the rhizobia are more problematic from an evolutionary point of view, especially for those with indeterminate growth, where the rhizobia that have become bacteroids can fix nitrogen but cannot reproduce. Hence, they cannot themselves benefit from the symbiosis, since 'benefit' must express itself, ultimately, as an increased reproductive rate (fitness). The rhizobia in the infection thread are capable of reproduction (and are therefore able to benefit), but they cannot fix nitrogen and are therefore not themselves involved in a mutualistic interaction. However, since the rhizobia are clonal, the bacteroids and the cells in the infection thread are all part of the same, single genetic entity. The bacteroids, therefore, by supporting the plant and generating a flow of photosynthates, can benefit the cells of the infection thread, and hence benefit the clone as a whole, in much the same way as the cells in a bird's wing can bring benefit, ultimately, to the cells that produce its eggs – and hence to the bird as a whole.

why no cheating?

One puzzle, though, since the rhizobia associated with a particular plant are typically a mixture of clones, is why individual clones do not 'cheat': that is, derive benefits from the plant, which itself derives benefit from the rhizobia in general, without themselves entering fully into the costly enterprise of fixing nitrogen. Indeed, we can see that this question of *cheating* applies to many mutualisms, once we recognise that they are, in essence, cases of mutual exploitation (Jones *et al.*, 2015). There would be evolutionary advantage in exploiting without being exploited. Perhaps the most obvious answer is for the plant (in this case) to monitor the performance of the rhizobia and apply 'sanctions' if they cheat (Westhoek *et al.*, 2017). This, clearly, will provide evolutionary stability to the mutualism by preventing cheats from escaping the interaction, and evidence for such sanctioning has indeed been found for a legume–rhizobium mutualism (Kiers *et al.*, 2003). A normally mutualistic rhizobium strain was prevented from cooperating (fixing nitrogen) by growing its soybean host in an atmosphere in which air (80% nitrogen, 20% oxygen) was replaced with approximately 80% argon, 20% oxygen and only around 0.03% nitrogen, reducing the rate of nitrogen fixation to around 1% of normal levels. Thus, the rhizobium strain was forced to cheat. In experiments at the whole plant, the part-root and the individual nodule level, the reproductive success of the non-cooperating rhizobia was decreased by around 50% (Figure 13.22). Non-invasive monitoring of the plants indicated that they were applying sanctions by withholding oxygen from the rhizobia. Cheating did not pay.

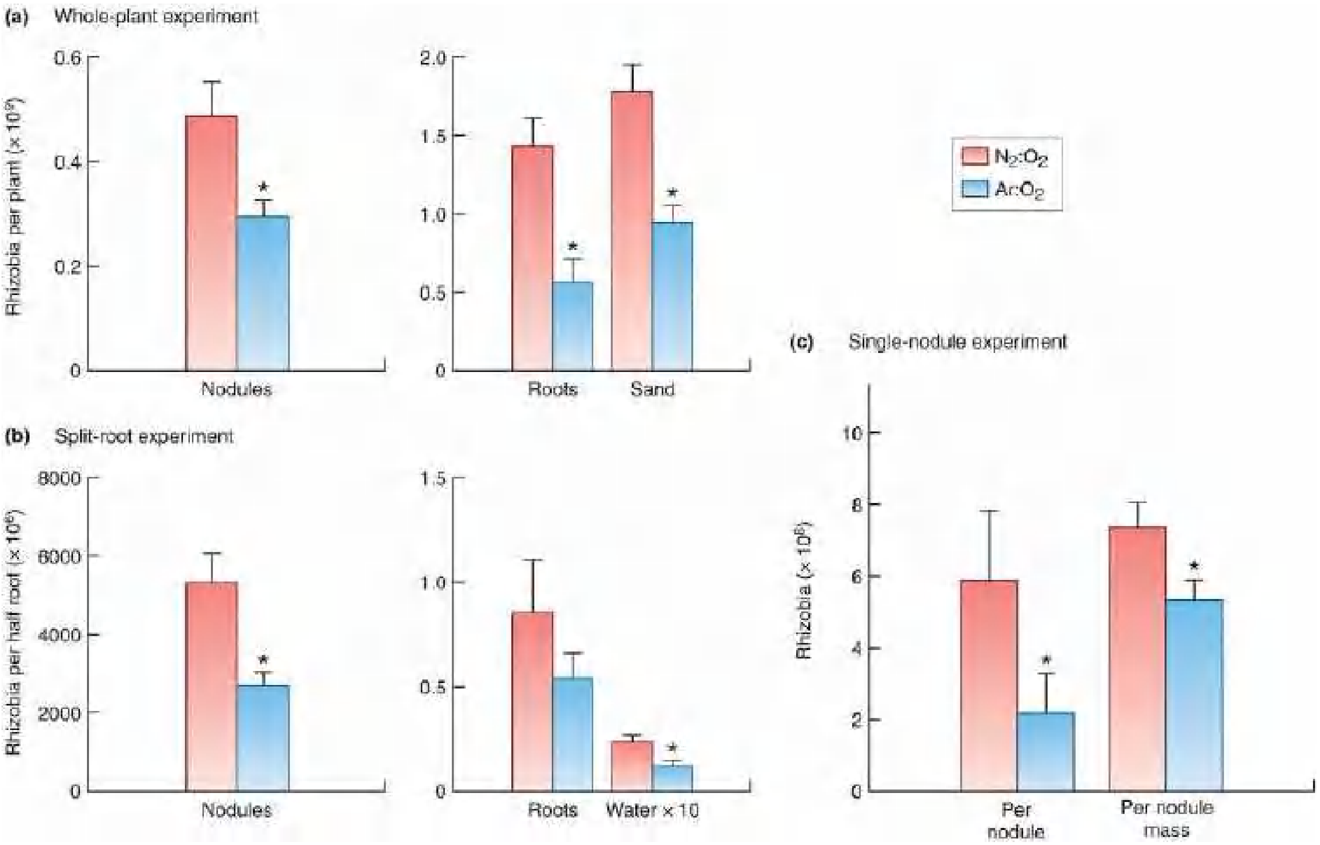


Figure 13.22 To prevent cheating by rhizobia, the soybean plant applies sanctions by withholding oxygen. The number of rhizobia grew to much larger numbers when allowed to fix nitrogen in normal air ($N_2 : O_2$) than when prevented from doing so by manipulation of the atmosphere ($Ar : O_2$). (a) When the different treatments were applied at the whole plant level, there were greater numbers within the nodules (left; $P < 0.005$) and on the root surface and in the surrounding sand (right; $P < 0.01$). $n = 11$ pairs; bars are SEs. (b) When the different treatments were applied to different parts of the same root system, there were greater numbers within the nodules (left; $P < 0.001$) and for those in the surrounding water (right; $P < 0.01$), but not significantly so for those on the root surface. $n = 12$ plants; bars are SEs. (c) When the different treatments were applied to individual nodules from the same root system, there were greater numbers on a per nodule basis ($P < 0.05$) and a per nodule mass basis ($P < 0.01$). $n = 6$ experiments; bars are SEs.

Source: After Keirs *et al.* (2003).

rhizobia–legume mutualisms in a plant community context

The mutualisms of rhizobia and legumes (and other nitrogen-fixing mutualisms) must not be seen as isolated interactions between bacteria and their own host plants. In nature, legumes normally form mixed stands in association with non-legumes. These are potential competitors with the legumes for fixed nitrogen (nitrates or ammonium ions in the soil). The nodulated legume sidesteps this competition by its access to a unique source of nitrogen. It is in this ecological context that nitrogen-fixing mutualisms gain their main advantage. Where nitrogen is plentiful, however, the energetic costs of nitrogen fixation often put the plants at a competitive *dis*advantage.

a classic ‘replacement series’

Figure 13.23, for example, shows the results of a classic experiment in which soybeans (*Glycine soja*, a legume) were grown in mixtures with *Paspalum*, a grass. The mixtures either received mineral nitrogen, or were inoculated with *Rhizobium*, or received both. The experiment was designed as a 'replacement series' (see Section 8.9), which allows us to compare the growth of pure populations of the grass and legume with their performances in the presence of each other. In the pure stands of soybean, yield was increased very substantially *either* by inoculation with *Rhizobium* or by application of fertiliser nitrogen, or by receiving both. The legumes can use either source of nitrogen as a substitute for the other. The grass, however, responded only to the fertiliser. Hence, when the species competed in the presence of *Rhizobium* alone, the legume contributed far more to the overall yield than did the grass: over a succession of generations, the legume would have outcompeted the grass. When they competed in soils supplemented with fertiliser nitrogen, however, whether or not *Rhizobium* was also present, it was the grass that made the major contribution: long term, it would have outcompeted the legume.

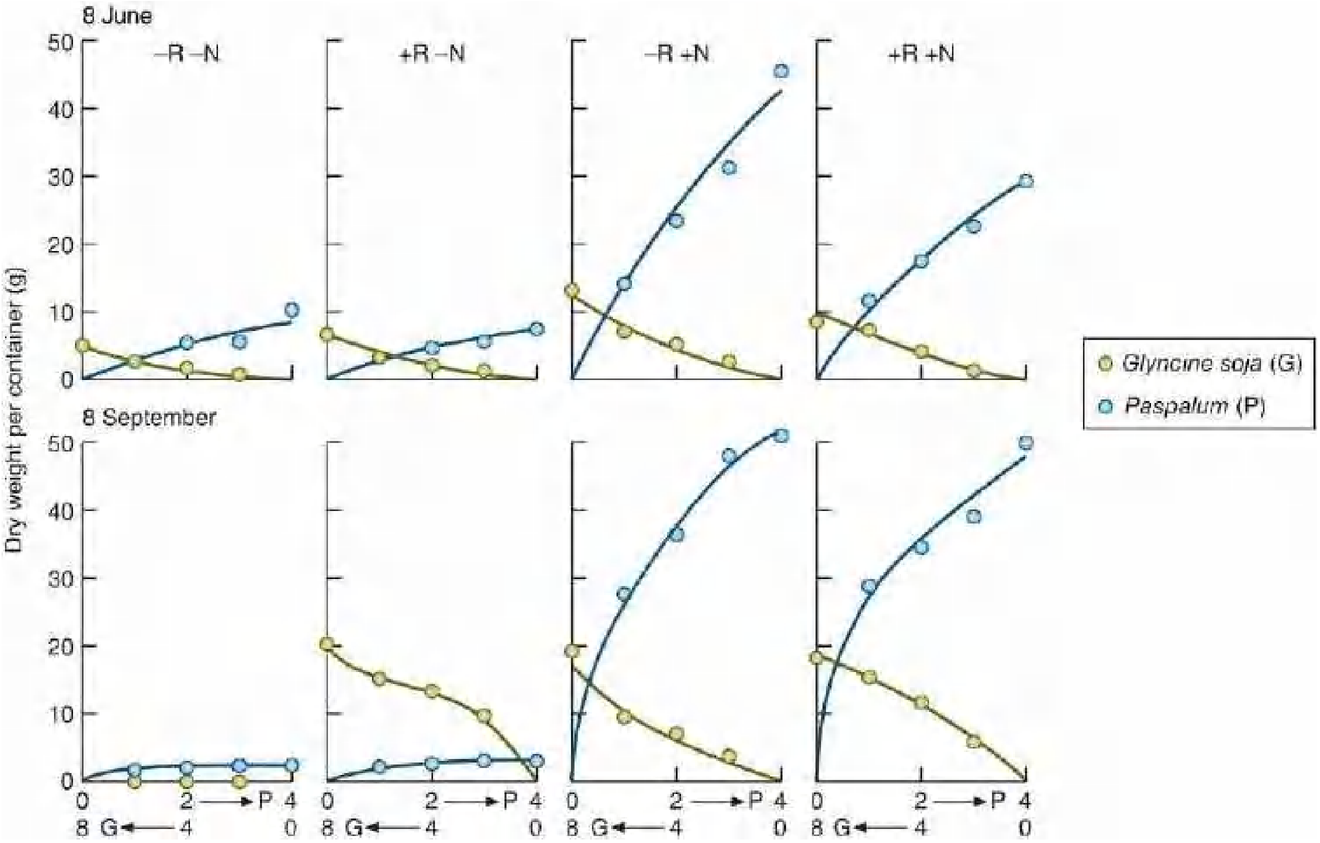


Figure 13.23 Relative importance of rhizobia and nitrogen fertiliser in the outcome of competition between a legume (soybean) and a grass. The growth of soybeans (*Glycine soja*, G) and a grass (*Paspalum*, P) grown alone and in mixtures with and without nitrogen fertiliser and with and without inoculation with nitrogen-fixing *Rhizobium*. The plants were grown in pots containing 0–4 plants of the grass and 0–8 plants of *Glycine*. The horizontal scale on each figure shows the mass of plants of the two species in each container. −R −N, no *Rhizobium*, no fertiliser; +R −N, inoculated with *Rhizobium* but no fertiliser; −R +N, no *Rhizobium* but nitrate fertiliser was applied; +R +N, inoculated with *Rhizobium* and nitrate fertiliser was supplied.

Source: After de Wit *et al.* (1966).

Quite clearly, then, it is in environments deficient in nitrogen that nodulated legumes have a great advantage over other species. But their activity raises the level of fixed nitrogen in the environment. After death, legumes augment the level of soil nitrogen on a very local scale with a 6–12-month delay as they decompose. Thus, their advantage is lost – they have improved the environment of their competitors, and the growth of associated grasses will be favoured in these local patches. Hence, organisms that can fix atmospheric nitrogen can be thought of as locally suicidal. This is one reason why it is very difficult to grow repeated crops of pure legumes in agricultural practice without aggressive grass weeds invading the nitrogen-enriched environment. It may also explain why leguminous herbs or trees usually fail to form dominant stands in nature.

Grazing animals, on the other hand, continually remove grass foliage, and the nitrogen status of a grass patch may again decline to a level at which the legume may once more be at a competitive advantage. In a stoloniferous legume, such as white clover, the plant is continually ‘wandering’ through the sward, leaving behind it local grass-dominated patches, while invading and enriching with nitrogen new patches where the nitrogen status has become low. The symbiotic legume in such a community not only drives its nitrogen economy but also some of the cycles that occur within its patchwork (Cain *et al.*, 1995).

13.11.2 Nitrogen-fixing mutualisms in non-leguminous plants

The distribution of nitrogen-fixing symbionts in non-leguminous higher plants is patchy. A genus of actinomycete, *Frankia* (200 strains identified), forms symbioses (actinorhiza) with members of 20 genera of flowering plants, almost all of which are shrubs or trees (Franche *et al.*, 2009). The nodules are usually hard and woody. The best-known hosts are the alder (*Alnus*), sea buckthorn (*Hippophaë*), sweet gale (*Myrica*), she-oak (*Casuarina*) and the arctic/alpine shrubs *Arctostaphylos* and *Dryas*. *Ceonothus*, which forms extensive stands in Californian chaparral, also develops *Frankia* nodules. Unlike rhizobia, the species of *Frankia* are filamentous and produce specialised vesicles and sporangia that release spores. Whilst the rhizobia rely on their host plant to protect their nitrogenase from oxygen, *Frankia* provides its own protection in the walls of the vesicles, which are massively thickened with as many as 50 monolayers of lipids.

Cyanobacteria form symbioses with three genera of liverwort (*Anthoceros*, *Blasia* and *Clavicularia*), with one fern (the free-floating aquatic *Azolla*), with many cycads (e.g. *Encephalartos*) and with all 40 species of the flowering plant genus *Gunnera*, but with no other flowering plants. In the liverworts, the cyanobacteria *Nostoc* live in mucilaginous cavities and the plant reacts to their presence by developing fine filaments that maximise contact with it. *Nostoc* is found at the base of the leaves of *Gunnera*, in the lateral roots of many cycads, and in pouches in the leaves of *Azolla* (Franche *et al.*, 2009).

13.11.3 Nitrogen-fixing plants and succession

An ecological succession (treated in much more detail in Chapter 18) is the directional replacement of species by other species at a site. A shortage of fixed nitrogen commonly hinders the earliest stages of the colonisation of land by vegetation: the initial stages of a succession on open land. Some fixed nitrogen will be contributed in rain after thunderstorms, and some may be blown in from other more established areas, but nitrogen-fixing organisms such as bacteria, cyanobacteria and lichens are important pioneer colonisers. Higher plants with nitrogen-fixing symbionts, however, are rarely pioneers. The reason appears to be that open land is usually colonised first by plants with light, dispersible seeds. A legume seedling, however, depends on fixed nitrogen in its seed reserves and the soil before it can grow to a stage where it can nodulate and fix nitrogen for itself. It is likely, therefore, that only large-seeded legumes carry enough fixed nitrogen to carry them through the establishment phase, and species with such large seeds will not have the dispersibility needed to be pioneers (Grubb, 1986).

Finally, note that since symbiotic nitrogen fixation is energetically demanding, it is not surprising that most of the higher plant species that support nitrogen-fixing mutualists are intolerant of the shade that is characteristic of the late stages of successions. Higher plants with nitrogen-fixing mutualists are seldom in at the beginning of a succession and they seldom persist to the end.

13.12 Models of mutualisms

Several of the previous chapters on interactions have included a section on mathematical models. They did so because the models, by separating essence from detail, were able to provide insights that would not be apparent from a simple catalogue of examples. For modelling to be a success, then, it is imperative that the 'essence' is correctly identified. If we assume that the essence of a mutualism is simply that each partner has a positive influence on the fitness of the other partner, then we might imagine that an appropriate model for a mutualistic interaction would simply replace the negative contributions in models of two-species competition (see Chapter 8) with positive contributions. However, such a model leads to absurd solutions in which both populations explode to unlimited size (May, 1981), because it places no limits on the carrying capacity of either species, which would therefore increase indefinitely. In practice, intraspecific competition must eventually limit the size of any mutualist population, even if the population of the partner mutualist is present in excess. For example, a plant whose growth is limited by a shortage of fixed nitrogen may be released into faster growth by mutualism with a nitrogen-fixing partner, but its faster growth must ultimately become constrained by a shortage of some other limiting resource (e.g. water, phosphate, radiant energy).

Moreover, as we have noted previously, the essence of mutualism is not simply 'mutual benefit', but that each partner exploits the other, with benefits to be gained but also costs to pay. We should recognise, too, that the balance of benefits and costs can alter – with changing conditions, changing resource levels, the abundance of either partner, or the presence or abundance of other species. Even the simplest of models, therefore, should include terms signifying both benefits and costs, varying, ideally, according to the state of some other part of the model community – not simple at all, compared with the models described, and found useful, in previous chapters. Cropp and Norbury (2018) trace the history of models of mutualism and emphasise the importance of including costs. Indeed they propose a model of their own that includes not only benefits and costs, but also variations in 'obligation' (the extent to which one or other partner can exist in the absence of its partner mutualist), and explicit dynamics of another element in the interaction: a non-living resource that the two mutualists share, for example nitrogen. Through appropriate manipulation of the parameters of their model, a wide variety of realistic outcomes are possible. This in turn emphasises a point we have made before in this chapter – that mutualism is, in terms of population dynamics, *essentially* an interaction that should only be viewed within the broader context of a larger community. We saw this, for example, for ants and aphids in the presence and absence of aphid predators, for *Symbiodinium* taxa coexisting in coral, and for the legume–*Rhizobium* mutualism, which brings its great advantage to the legume when it is competing with some other plant (e.g. a grass) for limited nitrogen from the soil.

a bird, a bee and two plants

This same point has been captured in a model of a two-species bee–plant pollination mutualism, initially considered in isolation (Figure 13.24a) (Ringel *et al.*, 1996). The bees could either take nectar and pollen from the plants but fail to pollinate them (predator–prey) or could pollinate them successfully (mutualistic). The model was, as described above, intrinsically unstable. The pair could only persist if the strength of intraspecific competition exceeded that of mutualism: the more mutualistic the interaction became, the more unstable it was.

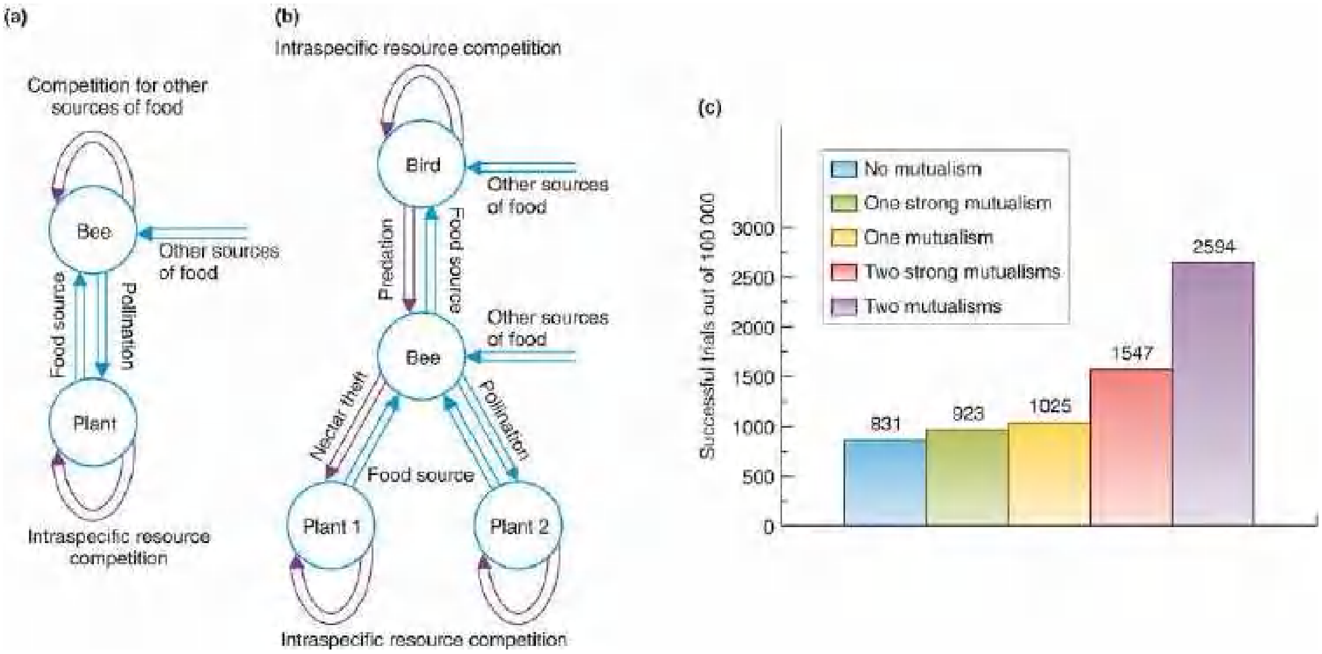


Figure 13.24 A model of bee–plant mutualisms is intrinsically unstable unless the mutualistic pair is embedded in a larger species assemblage. (a) A model two-species bee–plant mutualism. Both species are also subject to intraspecific competition. Blue arrows indicate a positive interaction, either a resource–consumer or pollination; purple arrows indicate negative interactions, either consumer–resource or intraspecific competition. The model of this simple mutualistic pair is intrinsically unstable. (b) The bee and plant embedded in a community with another plant and a bird predator of the bees. The plants suffer intraspecific competition but do not compete with one another. The birds suffer intraspecific competition but the bees do not. The bees take pollen and nectar from both plant species and either fail to pollinate them (predator–prey) or pollinate them successfully (mutualistic). In the figure, the interaction with plant 1 is predator–prey and that with plant 2 is mutualistic, but cases were examined in which neither, one or both were mutualistic. (c) Comparison of persistence in the possible assemblages in (b). Persistence of an assemblage is the maintenance of all species at positive population densities. The bars indicate the number that persisted when the dynamics of each assemblage were simulated 100 000 times, with the strengths of each interaction given by values generated randomly within defined bounds. In a 'strong mutualism', the strength of interaction could be up to twice that in a 'mutualism'. Mutualisms greatly increased the chances of persistence; two-tailed t-test of persistence versus no mutualism: one mutualism ($t = 4.52$, $P < 0.001$), one strong mutualism ($t = 2.21$, $P < 0.05$), two mutualisms ($t = 30.46$, $P < 0.001$), two strong mutualisms ($t = 14.78$, $P < 0.001$).

Source: After Ringel *et al.* (1996).

A quite different picture emerges, however, once the pair is embedded in a larger assemblage that also contained another species of plant and a species of bird that preyed on the bees (Figure 13.24b). Cases were examined in which either or both of the bee-plant interactions were mutualistic or predator–prey, and in which mutualisms were either 'standard' or strong (Figure 13.24c). It was apparent that mutualisms tended to *in*crease the chances of the assemblage's persistence. Clearly, there is no necessary paradox between the widespread occurrence of mutualistic interactions in nature and their effects in models. Equally clearly, though, while models are inevitably simplifications (in this case, just five species), the effects of mutualistic interactions in nature can easily be misjudged if they are *too* simple (the mutualistic pair alone).

Finally, and in a similar vein, Takimoto and Suzuki (2017) modelled the interaction between two obligate mutualists with and without a third species that acts as a facultative mutualist. As an illustration of such a three-way interaction in nature, the authors point to the obligate mutualism between yuccas and yucca moths (Section 13.5.3) in marginal habitats with low moth abundance which may be made persistent by the presence of *Pseudocalliope* flies that can subsist on resources other than the yucca plant and act as its facultative pollinator. They found that the addition of the third, facultative mutualistic species turned the fragile two-species interaction into one that is permanently persistent.

Chapter 14

Abundance

14.1 Introduction

Throughout most of this book, we incorporate applications of fundamental principles at appropriate places in the text. In this and the following chapter, we take a different approach. Firstly, in this chapter, we draw together threads from previous chapters and address some fundamental questions about abundance. Why are some species rare and others common? Why does a species occur at low population densities in some places and at high densities in others? What factors cause fluctuations in a species' abundance? However, there are three major areas of application of the determination of abundance – the control of pests, the harvesting of natural resources, and conservation ecology – that are so important and coherent in their own right that whole sections are devoted to them. Together, they constitute the next chapter.

numbers are not enough

The raw material for the study of abundance is usually some estimate of population size. In its crudest form, this consists of a simple count. But this can hide vital information. Picture three human populations containing identical numbers of individuals. One is an old people's residential area, the second is a population of young children, and the third is a population of mixed age and sex. No amount of attempted correlation with factors outside the population would reveal that the first was doomed to extinction, the second would grow fast but only after a delay, and the third would continue to grow steadily. More detailed studies, therefore, involve recognising individuals of different age, sex, size and dominance and even distinguishing genetic variants.

estimates may be misleading

In practice, ecologists usually have to estimate abundance rather than measure it (see Section 4.3). We must try our best, and it is important to ensure that our estimates are as accurate as possible, but these may be deficient in a number of ways. Firstly, data may be misleading unless there is sufficient commitment of time and money to ensure that sampling is adequate over both space and time. The lifetime of investigators, the hurry to produce publishable work, and the short tenure of most research programmes all deter individuals from even starting to conduct studies over extended periods of time. It is also usually a technically formidable task to follow individuals in a population throughout their lives – baby rabbits within their warrens or seeds in the soil. For these reasons, a large part of population theory depends on the relatively few exceptions where logistical difficulties have been overcome or the researchers' commitments have been exceptional (Taylor, 1987). In fact, many of the really long-term or geographically extensive studies of abundance have been made of organisms of economic importance such as fur-bearing animals, game birds and pests, or the furry and feathered favourites of amateur naturalists. Insofar as generalisations emerge, they should be treated with great caution.

correlation, causation and experimentation

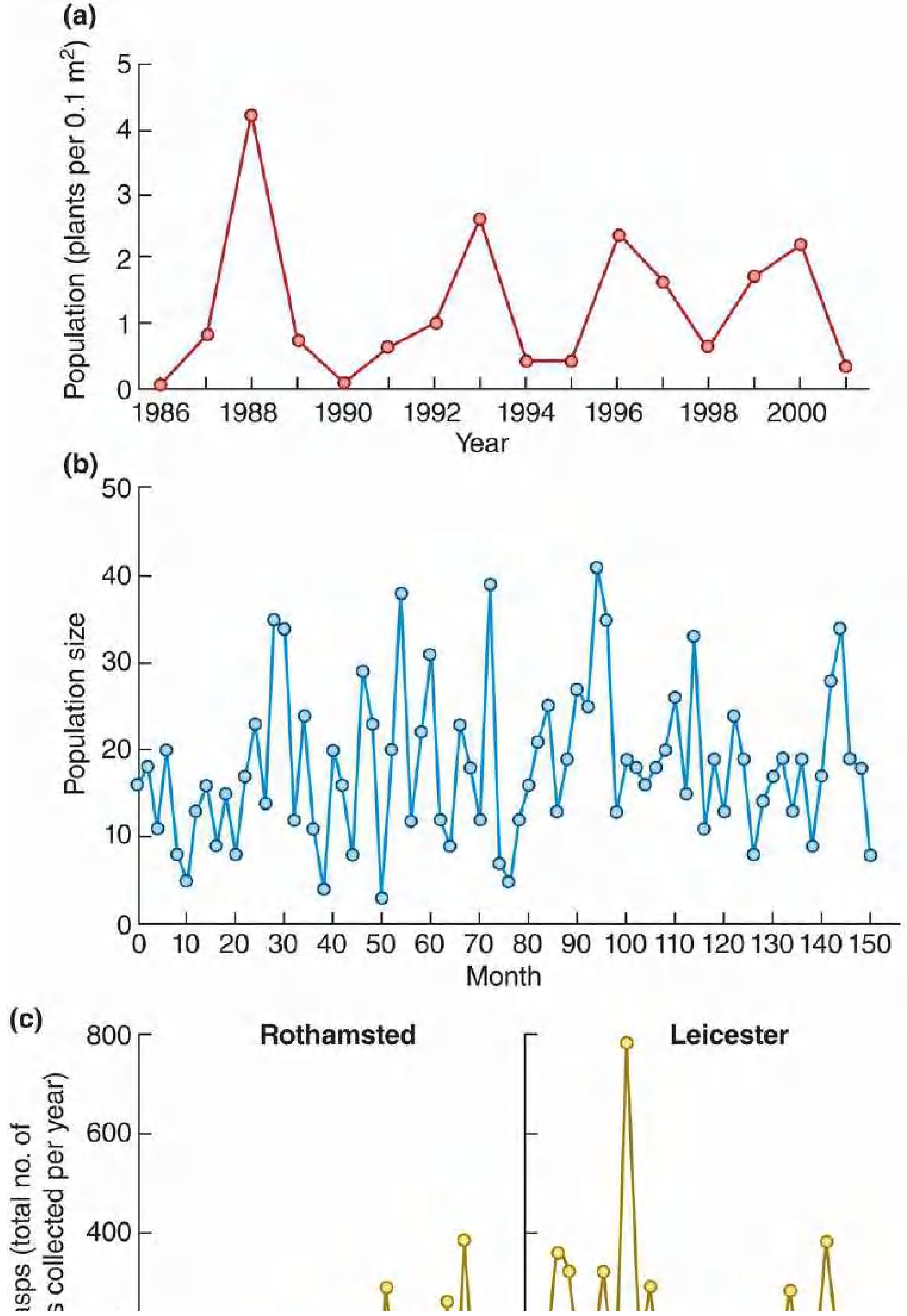
Abundance data may be used to establish correlations with external factors (e.g. the weather) or correlations 'within' the abundance data themselves (e.g. correlating numbers one year with those in the previous year). Correlations may be used to predict the future: following rain, plant abundance increases. They may also be used to suggest causal relationships, although they can usually neither prove nor identify them. For example, a negative correlation may be demonstrated between the size of a population and its growth rate. This suggests a causal connection – but not what that cause is. As we noted in Chapter 5, populations do not respond to density: organisms respond to one of its consequences. It may be that when population density is high, many individuals starve to death, or fail to reproduce, or become aggressive and drive out the weaker members. The correlation could not tell us which.

Incorporating observations on individuals into mathematical models of populations, and finding that the model population behaves like the real population, may also provide strong support for a particular hypothesis. But often, the acid test comes when it is possible to carry out a field experiment or manipulation. If we suspect that predators or competitors determine the size of a population, we can ask what happens if they are removed. If it is suspected that a resource limits the size of a population, more of it can be added. Besides indicating the adequacy of hypotheses, the results of such experiments may show that ecologists have the power to determine a population's size – to reduce the density of a pest or weed, or to increase the density of an endangered species. Ecology becomes a predictive science when it can forecast the future: it becomes a management science when it can determine the future.

14.2 Fluctuation or stability?

14.2.1 Determination and regulation of abundance

When we examine time series of abundances, some exhibit apparently regular patterns of rises and falls (Figure 14.1a), others exhibit much less regular fluctuations, though typically within some upper and lower bound (Figure 14.1b), while still others seem to flip between quite different patterns (Figure 14.1c). Looking at studies like these, some ecologists have emphasised the apparent constancy of population sizes and stressed the need to look for stabilising forces within populations to explain why they do not increase without bounds or decline to extinction. Others have emphasised the fluctuations and have looked to external factors, for example the weather, to explain the changes. Disagreements between the two camps dominated much of ecology in the middle third of the 20th century. By considering some of these arguments, it will be easier to appreciate the details of the modern consensus (see also Turchin, 2003).


(a)
Population (plants per 0.1 m²)
0
1
2
3
4
5
1986
1988
1990
1992
1994
1996
1998
2000
Year
(b)
Population size
0
10
20
30
40
50
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
Month
(c)
asps (total no. of
s collected per year)
800
600
400
Rothamsted
Leicester

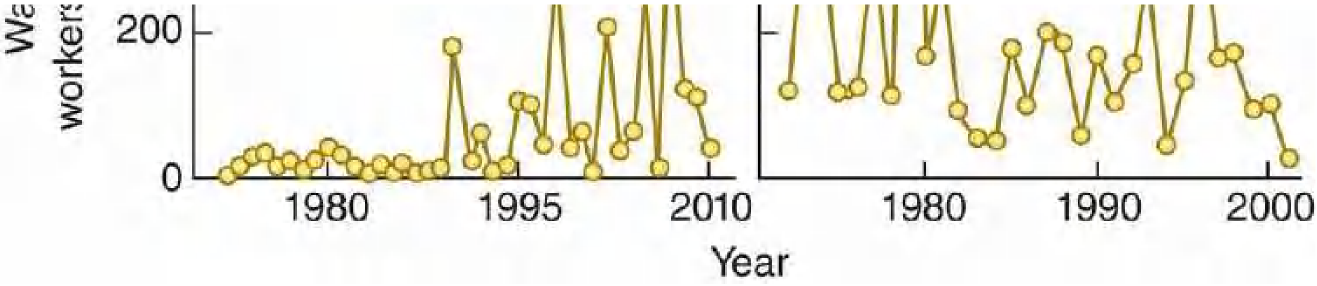


Figure 14.1 **Population time series showing a range of patterns of abundance.** (a) Fluctuations in the abundance of flixweed, *Descurainia sophia*, in Spain, with an apparent regularity, peaking every four or five years. (b) Irregular fluctuations in the abundance of the black-eared opossum, *Didelphis aurita*, in the state of Rio de Janeiro, Brazil. (c) Varied patterns of fluctuation, even within populations, for two English populations of the common wasp, *Vespula vulgaris*.

Source: (a) After Gonzalez-Andujar *et al.* (2006). (b) After Brigatti *et al.* (2016). (c) After Lester *et al.* (2017).

theories of abundance: Nicholson, Andrewartha and Birch

The 'stability' viewpoint usually traces its roots back to A. J. Nicholson (e.g. Nicholson, 1954), who believed that density-dependent, biotic interactions play the main role in determining population size, holding populations in a state of balance in their environments. He recognised that 'factors which are uninfluenced by density may produce profound effects upon density', but he considered that density dependence 'is merely relaxed from time to time and subsequently resumed, and it remains the influence which adjusts population densities in relation to environmental favourability'. The other point of view is usually traced back to Andrewartha and Birch (1954) who worked mostly on insect pests. They believed that populations could be viewed as passing through a repeated sequence of setbacks and recovery – a view that can certainly be applied to insect pests that are sensitive to unfavourable environmental conditions but are able to bounce back rapidly. They also rejected any subdivision of the environment into density-dependent and density-independent 'factors', preferring instead to see populations as sitting at the centre of an ecological web, where the essence was that various factors and processes interacted in their effects on the population.

the determination and regulation of abundance

To reconcile the two schools, we should first understand clearly the difference between questions about the *determination* of abundance and about how abundance is *regulated*. Regulation is the tendency of a population to decrease in size when it is above a particular level, but to increase when below that level. Hence, by definition, regulation can occur only as a result of one or more of the density-dependent processes discussed in previous chapters. We must look at regulation, therefore, to understand how it is that a population tends to remain within defined upper and lower limits.

On the other hand, the precise abundance of individuals will be determined by the combined effects of all the processes that affect a population, whether dependent or independent of density. Figure 14.2 shows this diagrammatically and very simply. Here, the birth rate is density dependent, whilst the death rate is density independent but varies with physical conditions that differ between three locations. Hence, there are three equilibrium populations (N_1, N_2, N_3), which correspond to the three death rates in the three environments. We must look at the determination of abundance, therefore, to understand how it is that a particular population exhibits a particular abundance at a particular time, and not some other abundance.

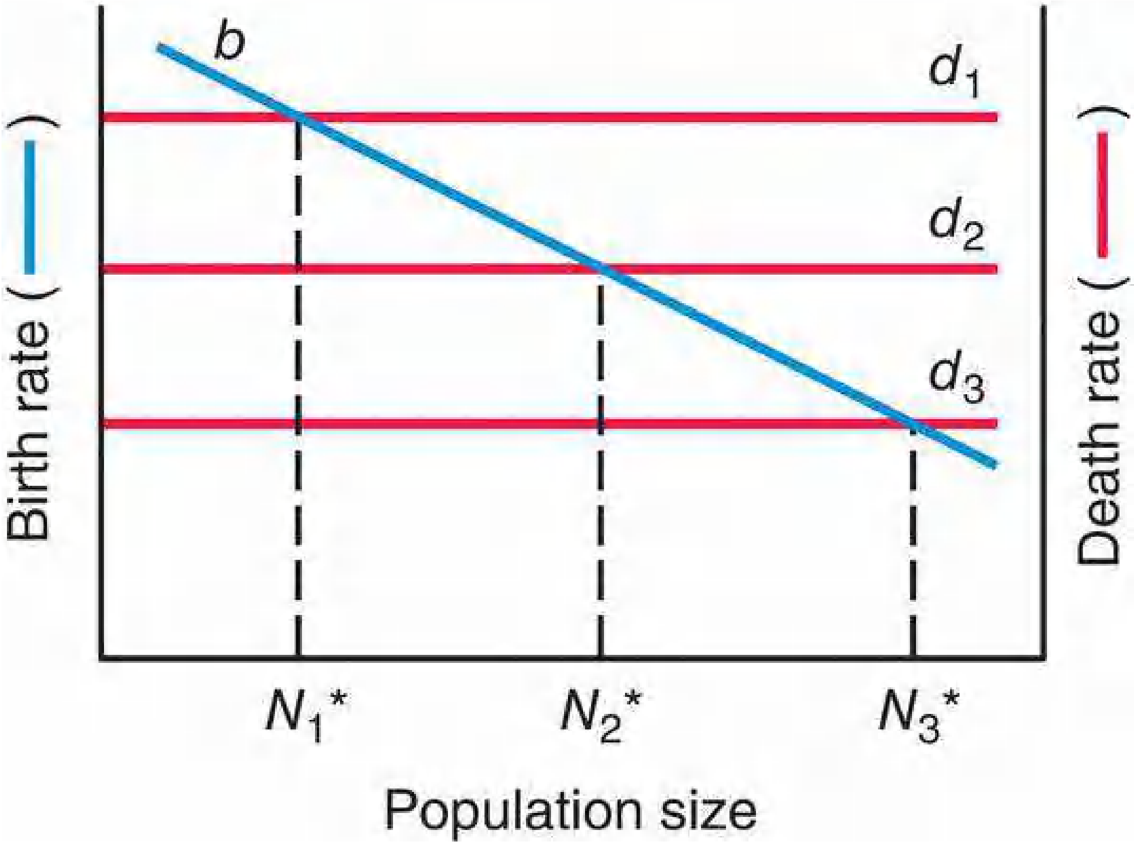


Figure 14.2 Density-dependent and -independent processes combine to determine abundance. Population regulation with density-dependent birth, *b*, and density-independent death, *d*. Death rates are determined by physical conditions that differ in three sites (death rates d_1, d_2 and d_3). Equilibrium population size varies as a result (N_1^*, N_2^*, N_3^*).

With this in mind, it seems clear that the Nicholson camp was preoccupied with what regulates population size, and the Andrewartha and Birch camp with what determines population size – and both are perfectly valid interests. The disagreements perhaps arose because of some feeling within the first camp that whatever regulates *also* determines; and some feeling in the second camp that the determination of abundance is, for practical purposes, all that really matters. It is indisputable, however, that no population can be absolutely free of regulation: long-term unrestrained population growth is unknown, and unrestrained declines to extinction are rare. Also, though, as we shall see, it is not necessarily the case that whatever regulates the size of a population also determines its size for most of the time. It would be wrong, therefore, to give regulation or density dependence some kind of pre-eminence. It may be occurring only infrequently or intermittently. And even when regulation is occurring, it may be drawing abundance towards a level that is itself changing in response to changing levels of resources. The most common and obvious example of this is seasonality, where regulatory forces may be attracting a population towards a cyclically varying target, often effectively captured in population models by making vital rates or carrying capacities subject to a sine wave or similarly cyclical mathematical function (see Section 14.6.3 for an example). It is likely that no natural population is ever truly at equilibrium.

14.2.2 Approaches to the investigation of abundance

demographic, mechanistic and density approaches

There are three broad approaches that have been used to address questions about the determination and regulation of abundance (Sibly & Hone, 2002). In each case, population growth rate is placed centre stage, since this summarises the combined effects on abundance of birth, death and movement. The *demographic* approach (Section 14.3) seeks to partition variations in the overall population growth rate amongst the phases of survival, birth and movement occurring at different stages in the life cycle, and is therefore closely linked to the topics discussed in Section 4.7. The aim is to identify the most important phases, and to determine what those phases are important for. The *mechanistic* approach (Section 14.4) seeks to relate variations in growth rate directly to variations in specified factors that might influence it – food, temperature, and so on. The approach itself can range from establishing correlations to carrying out field experiments. Finally, the *time series* approach (Section 14.5) applies analyses to those time series that seek to relate variations in growth rate to variations in density. As Sibly and Hone's (2002) survey also makes clear, many studies are hybrids of two, or even all three, of these approaches. Lack of space will prevent us from looking at all of the different variants.

14.3 The demographic approach

14.3.1 Key factor analysis

key factors? or key phases?

For many years, the demographic approach was represented by a technique called *key factor analysis*, though in fact it identifies key *phases* (rather than factors) in the life of the organism concerned. Here, we deal with it briefly, as a means of explaining important general principles, and for historical completeness, before turning to more recent alternatives.

For a key factor analysis, data are required in the form of a series of life tables (see Section 4.6) from a number of different cohorts of the population concerned. Thus, since its initial development (Morris, 1959 ; Varley & Gradwell, 1968) it has most commonly been used for species with discrete generations, or where cohorts can otherwise be readily distinguished. It is an approach based on the use of k values as a way of measuring the strength of mortality during a particular phase of the life cycle, or of a shortfall in mortality (see Sections 4.6.1 and 5.3). What are believed to be the main causes of death or reduced fecundity in each phase may then be listed, converting what is essentially a demographic technique (dealing with phases) into a mechanistic approach (by associating each phase with a proposed 'factor'). Key factor analysis has been applied to a great many insect populations, but to far fewer vertebrate or plant populations. Examples of these, though, are shown in Table 14.1 for populations of the wood frog (*Rana sylvatica*) in three regions of the USA, and in Figure 14.3 for a Polish population of the sand dune annual plant *Androsace septentrionalis*.

Table 14.1 Key factor (or key phase) analysis for wood frog populations from three areas in the USA: Maryland (two ponds, 1977–82), Virginia (seven ponds, 1976–82) and Michigan (one pond, 1980–93). In each area, the phase with the highest mean *k* value (second column), the key phase (third column) and any phase showing density dependence (fourth column) are highlighted in bold.

Source: After Berven (1995).

Age interval	Mean k value	Coefficient of regression on k_{total}	Coefficient of regression on log (population size)
Maryland			
Larval period	1.94	**0.85**	**Pond 1: 1.03 ($P = 0.04$)**
			Pond 2: 0.39 ($P = 0.50$)
Juvenile: up to 1 year	0.49	0.05	0.12 ($P = 0.50$)
Adult: 1–3 years	**2.35**	0.10	0.11 ($P = 0.46$)
Total	4.78		
Virginia			
Larval period	**2.35**	**0.73**	0.58 ($P = 0.09$)
Juvenile: up to 1 year	1.10	0.05	−0.20 ($P = 0.46$)
Adult: 1–3 years	1.14	0.22	**0.26 ($P < 0.05$)**
Total	4.59		
Michigan			
Larval period	1.12	**1.40**	1.18 ($P = 0.33$)
Juvenile: up to 1 year	0.64	1.02	0.01 ($P = 0.96$)
Adult: 1–3 years	**3.45**	−1.42	**0.18 ($P < 0.005$)**
Total	5.21		

Table 14.2 **Key factor analysis and λ-contribution analysis for red deer.** Columns 1–4 contain life table data for the females of a population of red deer, *Cervus elaphus*, on the island of Rhum, Scotland, using data collected between 1971 and 1983 (Clutton-Brock *et al.*, 1985): x is age, l_x is the proportion surviving at the start of an age class, k_x, killing power, has been calculated using natural logarithms, and m_x, fecundity, refers to the birth of female calves. These data represent averages calculated over the period, the raw data having been collected both by following individually recognisable animals from birth and ageing animals at death. The next two columns contain the sensitivities of λ, the population growth rate, to k_x and m_x in each age class. In the final two columns, the contributions of the various age classes have been grouped as shown. These columns show the contrasting results of a key factor analysis and a λ-contribution analysis as the regression coefficients of k_x and m_x on k_{total} and λ_{total}, respectively, where λ_{total} is the deviation each year from the long-term average value of λ.

Source: After Sibly and Smith (1998), where details of the calculations may also be found.

Age (years) at start of class, x	l_x	k_x	m_x	Sensitivity of λ to k_x	Sensitivity of λ to m_x	Regression coefficients of k_x, left, and m_x, right, on k_{total}	Regression coefficients of k_x, left, and m_x, right, on λ_{total}
1	0.64	0.08	0.00	−0.14	0.09	0.01, –	0.14, –
2	0.59	0.08	0.00	−0.14	0.08		
3	0.54	0.03	0.22	−0.13	0.07		
0	1.00	0.45	0.00	−0.14	0.16	0.01, –	0.32, –
4	0.53	0.03	0.22	−0.11	0.06	0.00, 0.05	0.03, 0.04
5	0.51	0.04	0.35	−0.10	0.05		
6	0.49	0.04	0.35	−0.08	0.05	−0.00, 0.03	0.08, 0.16
7	0.47	0.04	0.35	−0.07	0.04		
8	0.45	0.06	0.37	−0.05	0.04		
9	0.42	0.06	0.37	−0.04	0.03		
10	0.40	0.06	0.37	−0.03	0.03	0.01, 0.15	0.09, 0.12
11	0.38	0.06	0.37	−0.02	0.02		
12	0.35	0.06	0.37	−0.02	0.02		
13	0.33	0.30	0.30	−0.01	0.02		
14	0.25	0.30	0.30	−0.006	0.01		
15	0.18	0.30	0.30	−0.004	0.008		
16	0.14	0.30	0.30	−0.002	0.005	−0.05, 0.80	0.01, −0.00
17	0.10	0.30	0.30	−0.001	0.004		
18	0.07	0.30	0.30	−0.001	0.002		

Age (years) at start of class, x	l_x	k_x	m_x	Sensitivity of λ to k_x	Sensitivity of λ to m_x	Regression coefficients of k_x, left, and m_x, right, on k_{total}	Regression coefficients of k_x, left, and m_x, right, on λ_{total}
19	0.06	0.30	0.30	−0.000	0.002		

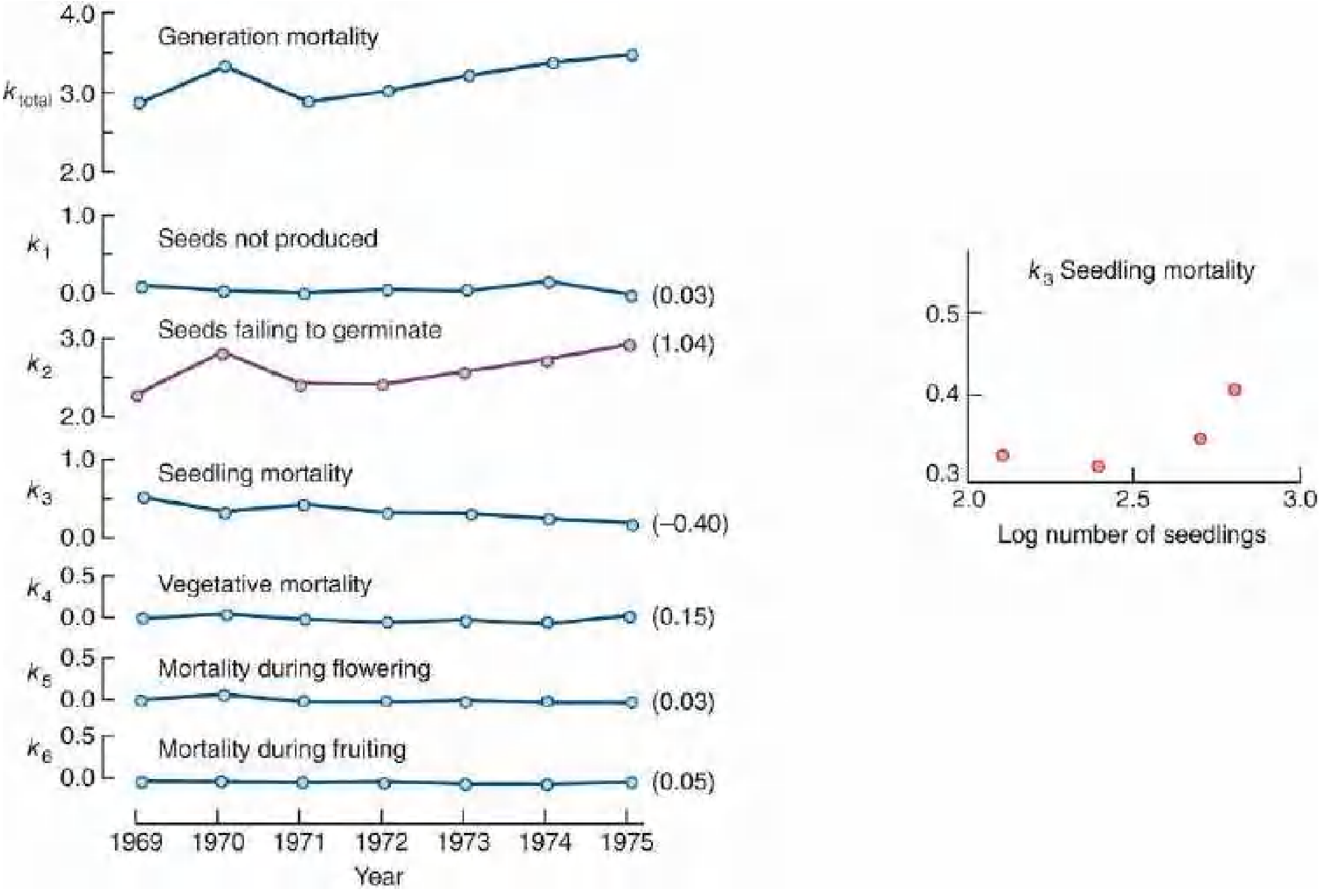


Figure 14.3 Key factor analysis of the sand dune annual plant *Androsace septentrionalis*. A graph of total generation mortality (k_{total}) and of various k factors is presented. The values of the regression coefficients of each individual k value on k_{total} are given in brackets. The largest regression coefficient signifies the key phase and is shown as a purple line. The one k value that varies in a density-dependent manner is shown to the right.

Source: After Symonides (1979); analysis in Silvertown (1982).

The first step is to determine the mean k value for each phase, indicating the relative strengths of the different phases (or the factors within them) as contributors to the total rate of mortality within a generation. What these mean values do not tell us, however, is the relative importance of each phase or factor as a determinant of the year-to-year *fluctuations* in mortality. A factor may repeatedly take a significant toll from a population, but, by remaining constant in its effects, play little part in determining the particular rate of mortality (and thus, the particular population size) in any one year. This can be assessed, however, by calculating the regression coefficient of each individual k value on the total generation value, k_{total}. A mortality factor that is important in determining population changes will have a regression coefficient close to unity, because its k value will tend to fluctuate in line with k_{total} in terms of both size and direction (Podoler & Rogers, 1975). A mortality factor with a k value that varies quite randomly with respect to k_{total}, however, will have a regression coefficient close to zero. Moreover, the sum of all the regression coefficients within a generation will always be unity. The values of the regression coefficients will, therefore, indicate the relative strength of the association between different factors and the fluctuations in mortality. The largest regression coefficient will be associated with the *key phase* or *key factor* causing population change. Thus, whilst mean k values indicate the average strengths of various factors as causes of mortality in each generation, key factor analysis indicates their relative contribution to the yearly *changes* in generation mortality, and thus measures their importance as determinants of population size.

For the frogs (Table 14.1), the highest mean value was associated with the larval phase in one of the populations, but with the adult phase in the other two. However, the larval period was consistently the key phase determining abundance in all three, largely as a result of year-to-year variations in rainfall during the larval period. In low rainfall years, the ponds could dry out, reducing larval survival to catastrophic levels. For *A. septentrionalis* (Figure 14.3), the seeds in the soil (and their possible failure to germinate) were the key phase determining abundance but also had the highest mean *k* value and so made the greatest contribution to overall mortality.

a role for factors in regulation?

What, though, of population regulation? We can address this by examining the density dependence of each factor by plotting *k* values against $\log_{10}$ of the numbers present before the factor acted (see Section 5.3). For the wood frogs, the key, larval phase was inconsistently related to the size of the larval population (one pond in Maryland, and only approaching significance in Virginia) and hence played an inconsistent part in regulating the sizes of the populations. Rather, in two of the regions it was during the adult phase that mortality was clearly density dependent and hence regulatory (apparently as a result of competition for food). For *A. septentrionalis*, again, mortality did not operate in a density-dependent manner during the key phase. In this case, it was the mortality of seedlings, which were not the key phase, that was found to be density dependent. Seedlings that emerge first in the season, at low density, stand a much greater chance of surviving. Overall, therefore, key factor analysis (its rather misleading name apart) is useful in identifying important phases in the life cycles of study organisms. It is useful too in distinguishing the variety of ways in which phases may be important: in contributing significantly to the overall sum of mortality; in contributing significantly to variations in mortality, and hence in *determining* abundance; and in contributing significantly to the *regulation* of abundance by virtue of the density dependence of the mortality.

14.3.2 λ-contribution analysis

Although key factor analysis was widely used in the past, it has been subject to persistent and valid criticisms, some technical (i.e. statistical) and some conceptual – for example, ‘importance’ may be inappropriately ascribed to different phases, because equal weight is given to all phases of the life history, even though they may differ in their power to influence abundance (Sibly & Smith, 1998). Sibly and Smith’s (1998) alternative to key factor analysis, λ-contribution analysis, overcomes these problems. λ is the population growth rate (e^{r}), which we referred to as *R* in Chapter 4, for example, though here we retain Sibly and Smith’s notation. The method makes use of a weighting of life cycle phases taken from sensitivity and elasticity analysis discussed in some detail in Section 4.8.3. Indeed, the population projection matrices and integral projection models discussed there are themselves very much focused on deconstructing the overall population growth rate into its component processes (phases, factors) in order to determine their relative importance, and in the case of integral projection models, the incorporation of density-dependent effects (and hence the potential for regulation) was identified as a valuable feature.

elasticity may say little about *variations* in abundance, but λ-contribution analysis does

Thus, such analyses are valuable in identifying phases and processes that may be important in determining abundance, but they do so by focusing on typical or average values, and in that sense they seek to account for the typical size of a population. However, a process with a high elasticity may still play little part, in practice, in accounting for variations in abundance from year to year or site to site if that process (mortality or fecundity) shows little temporal or spatial variation. In contrast, as we have seen, key factor analysis seeks specifically to understand temporal and spatial variations in abundance. The same is true of λ-contribution analysis. It deals with the contributions of the different phases not to an overall k value (as in key factor analysis) but to λ, a much more obvious determinant of abundance. It makes use of k values to quantify mortality, but can use fecundities directly rather than converting them into 'deaths of unborn offspring'. And crucially, the contributions of all mortalities and fecundities are weighted by their sensitivities. Hence, quite properly, where generations overlap, the chances of later phases being identified with a key factor are correspondingly lower in λ-contribution than in key factor analysis. As a result, λ-contribution analysis can be used with far more confidence when generations overlap. Subsequent investigation of density dependences proceeds in exactly the same way in λ-contribution analysis as in key factor analysis.

Table 14.2 contrasts the results of the two analyses applied to life table data collected on the Scottish island of Rhum between 1971 and 1983 for the red deer, *Cervus elaphus* (Clutton-Brock *et al.*, 1985). Over the 19-year lifespan of the deer, survival and birth rates were estimated in the following 'blocks': year 0, years 1 and 2, years 3 and 4, years 5–7, years 8–12 and years 13–19. This accounts for the limited number of different values in the k_x and m_x columns of the table, but the sensitivities of λ to these values are of course different for different ages (early influences on λ are more powerful), with the exception that λ is equally sensitive to mortality in each phase prior to first reproduction (since it is all 'death before reproduction'). The consequences of these differential sensitivities are apparent in the final two columns of the table, which summarise the results of the two analyses by presenting the regression coefficients of each of the phases against k_{total} and λ_{total}, respectively. Key factor analysis identifies reproduction in the final years of life as the key factor and even identifies reproduction in the preceding years as the next most important phase. In stark contrast, in λ-contribution analysis, the low sensitivities of λ to birth in these late phases relegate them to relative insignificance – especially the last phase. Instead, survival in the earliest phase of life, where sensitivity is greatest, becomes the key factor, followed by fecundity in the 'middle years' where fecundity itself is highest. Thus, λ-contribution analysis combines the virtues of key factor and elasticity analyses: distinguishing the regulation and determination of abundance, identifying key phases or factors, while taking account of the differential sensitivities of growth rate (and hence abundance) to the different phases.

14.4 The mechanistic approach

Key factor analyses are, as we have seen, directed at phases in the life cycle but often ascribe the effects occurring in particular phases to factors or processes known to operate during those phases – food, predation, etc. An alternative has been to study the role of particular factors in the determination of abundance directly, by relating the level or presence of the factor (the amount of food, the presence of predators) either to abundance itself or to population growth rate, the proximate determinant of abundance. This mechanistic approach has the advantage of focusing clearly on the particular factor, but in so doing it is easy to lose sight of the relative importance of that factor compared to others. The most basic way of doing this is through simple correlations. It is relatively easy, for example, to find cases in which population growth rate increases with the availability of food, though these suggest, in general, that such relationships are likely to level off at the highest food levels where some other factor or factors place an upper limit on abundance (Figure 14.4).

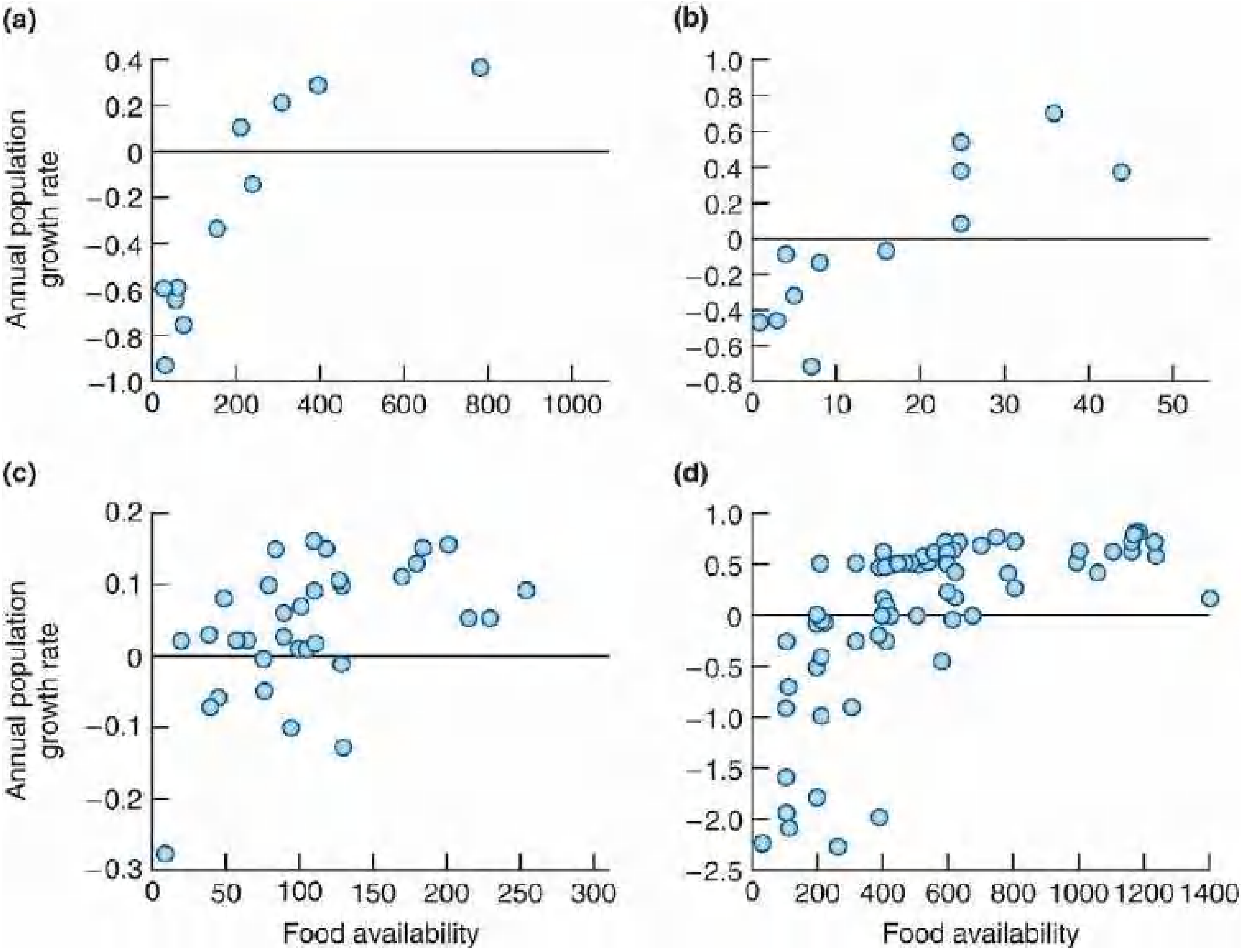


Figure 14.4 **Increases in annual population growth rate ($r = \ln \lambda$) with the availability of food**, measured as pasture biomass (in kg ha^{-1}), except in (b) where it is field vole abundance and in (c) where it is availability per capita. (a) Red kangaroo (from Bayliss, 1987). (b) Barn owl (after Taylor, 1994). (c) Wildebeest (from Krebs *et al.*, 1999). (d) Feral pig (from Choquenot, 1998).

Source: After Sibly & Hone (2002).

14.4.1 Experimental perturbation of populations

However, as previously noted, correlations can be suggestive, but a much more powerful test of the importance of a particular factor is to manipulate that factor and monitor the population's response. Predators, competitors or food can be added or removed, and if they are important in determining abundance, this should be apparent in subsequent comparisons of control and manipulated populations. Note, though, that field-scale experiments require major investments in time and effort (and money), especially as, for vertebrates especially, the spatial scale required may be very large, and a clear distinction between controls and experimental treatments is inevitably much more difficult to achieve than in the laboratory or greenhouse.

Both the power and the problems of field-scale experiments were illustrated by an example discussed in Section 12.8.1, in which Hudson *et al.* (1998) treated cyclic populations of the red grouse, *Lagopus lagopus scoticus*, against the nematode *Trichostrongylus tenuis*, while leaving other, control populations untreated. In the manipulated populations, the extent of the grouse 'crash' was very substantially reduced, proving the importance of the nematodes, normally, in reducing grouse abundance, and justifying the effort that had gone into the manipulation. But as we have seen, in spite of this effort, controversy remained about whether the nematodes had been proved to be the cause of the cycles (in which case, the residual smaller crashes were dying echoes) or whether, instead, the experiment had only proved a role for the nematodes in determining a cycle's amplitude, leaving their role in cyclicity itself uncertain. Experiments are better than correlations, but when they involve ecological systems in the field, eliminating ambiguity can never be guaranteed.

Another, related example was discussed in Section 12.8.2 in which Pedersen and Grieves (2008) treated populations of deer mice and white-footed mice, *Peromyscus maniculatus and P. leucopus*, in the USA with either supplementary food, or with a drug to kill parasites, or with both. Each was effective alone, but it was when both were applied together that the greatest effect on abundance was observed, emphasising the interactive effects of a multiplicity of factors in determining abundance generally.

Further examples of field-scale manipulations in which either food is supplemented or predators removed are discussed below, in an examination of what may drive the regular cycles of abundance exhibited by some species (see Section 14.6). We will meet examples in which the addition of nutrients leads to increased plant productivity, for example when iron is added to oceans causing phytoplankton blooms (see Section 20.4.3). And whenever there are successful examples of biological control, that is, the introduction of a pest's natural enemies to control that pest (see Section 15.2), this is experimental proof of the power of predators to reduce the abundance of their prey.

14.5 The time series approach

Associations with density have played a part in several of the approaches considered so far, and indeed, density dependence played a central role in our discussions of the determinants of abundance (birth, death and movement) in earlier chapters. Some studies, however, have focused much more on density dependences in their own right, and have been designed, especially, to seek evidence for both direct and *delayed* density dependence (see Section 10.1.2), since it is the combination of the two that is often so critical in determining the pattern of a population's dynamics.

abundance determination expressed as a time-lag equation

In particular, a number of related approaches have sought to dissect the density-dependent 'structure' of species' population dynamics by a statistical analysis of time series of abundance. We can think of abundance at a given point in time as a reflection of abundances at various times in the past. It reflects abundance in the immediate past in the obvious sense that the past abundance gave rise directly to the present abundance. But it may also reflect abundance in the more distant past if, for example, that past abundance gave rise to an increased abundance of a predator, which in due course (after a delay) affected the present abundance (that is, a delayed density dependence). One common and powerful approach to studying these relationships is through the use of linear autoregressive models (Royama, 1992 ; Berryman & Turchin, 2001). Without going into technical details, we saw in Section 4.7.1 that the intrinsic rate of natural increase of a population between times *t-1* and *t*, r_t, is the difference between the log-abundances at those times, that we refer to here at X_{t-1} and X_t. The idea that population growth is determined by past abundances is captured in the following equation:

$$r_t = X_t - X_{t-1} = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots + \beta_d X_{t-d} + u_t. \qquad (14.1)$$

Here, β_1 reflects the strength of direct density dependence, and subsequent βs reflect the strengths of delayed density dependences with various time lags up to a maximum *d*. (β_0 sets the average growth rate but not the dynamics of the population.) The final term, u_t, represents fluctuations from time-point to time-point imposed from outside the population, independent of density. Any density-dependent regulatory tendencies will therefore be reflected in negative β-values. In fact, analyses are often based on the log-abundances themselves (not population growth rate), requiring a simple rearrangement of Equation 14.1:

$$X_t = \beta_0 + (1 + \beta_1)X_{t-1} + \beta_2 X_{t-2} + \ldots + \beta_d X_{t-d} + u_t, \tag{14.2}$$

this time reflecting directly the dependence of present abundance on past abundances.

autocorrelation function analysis

The most common first step in analysing such time series is to estimate the autocorrelation function (ACF), particularly in cases where regular cycles in abundance are suspected, a topic to which we turn in Section 14.6. The ACF determines the correlations between pairs of abundances one time interval apart, two time intervals apart, and so on (Figure 14.5a, b). The correlation between abundances just one time interval apart is usually not very informative, since it can be high simply because one abundance has led directly to the next. Thereafter, a high positive correlation between pairs, for example, four years apart, would indicate a regular cycle with a period of four years; while a further high *negative* correlation between pairs two years apart would indicate a degree of symmetry in the cycle: peaks and troughs typically four years apart, with peaks typically two years from troughs. An example of what appears to be a regular cycle is shown in Figure 14.5a, for the larch budmoth, *Zeiraphera diniana*, in Switzerland, and this is confirmed by the ACF, with a peak positive correlation at nine years and a peak negative correlation at around half this (Figure 14.5b). This can be called a ‘nine-year cycle’, but this does not mean that the population will peak every nine years exactly and unfailingly. Some variation around this pattern is to be expected, and indeed is implied by the other correlations in the ACF.

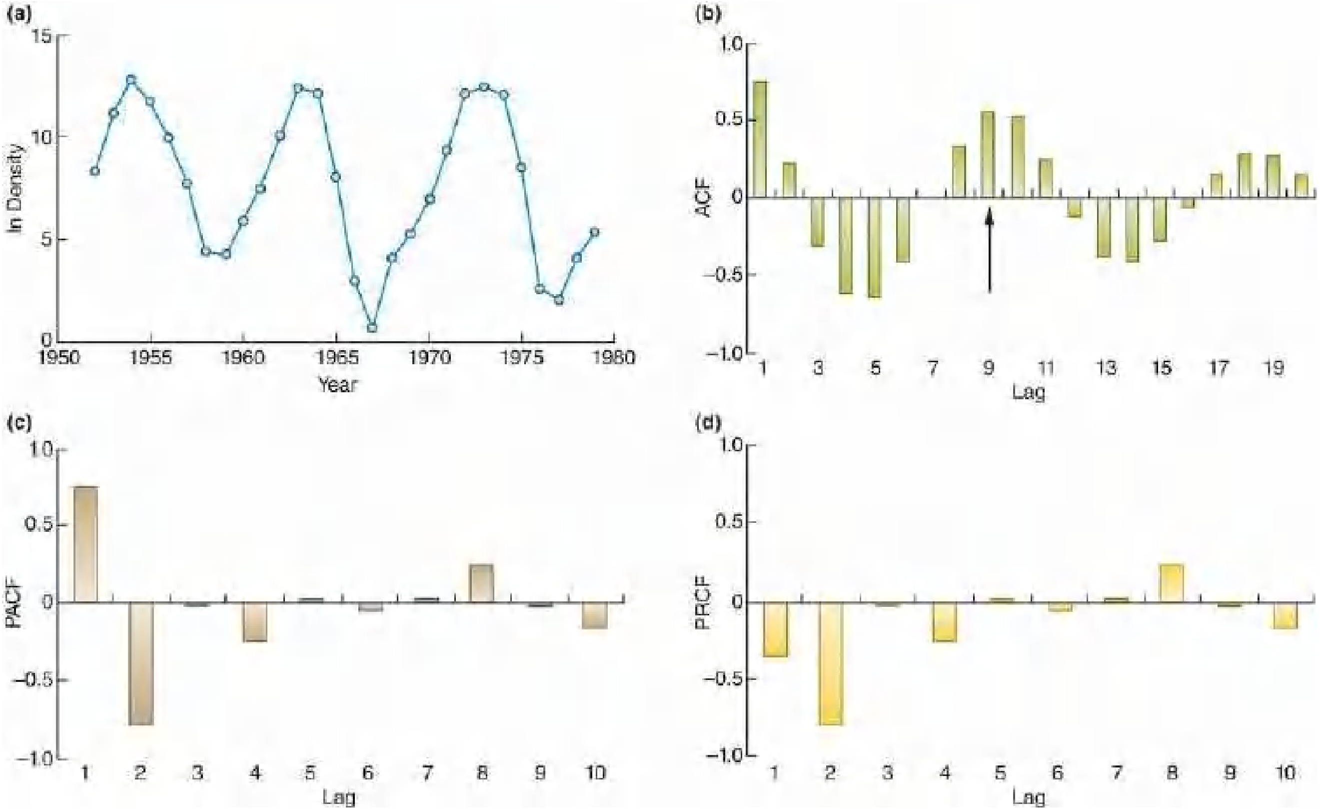


Figure 14.5 Larch budmoth time series and their analysis through correlation functions. (a) Fluctuations in the abundance (log larvae per 1000 kg of larch foliage) of the larch budmoth, *Zeiraphera diniana*, in Switzerland. (b) The autocorrelation function (ACF) of the time series. The peak positive correlation at lag 9 (arrow) suggests a 'nine year cycle'. (c) The partial autocorrelation function (PACF) of the time series. (d) The partial rate correlation function (PRCF) of the time series.

Source: After Berryman & Turchin (2001).

PACFs and PRCFs

A key question in such analyses is to determine the number of time delays that need to be retained in the equation; the value of d in Equations 14.1 and 14.2, often called the order or the dimension of the dynamics. The aim is to find the value that strikes the best balance between accounting for the variations in X_t and not including too many lags. Essentially, additional lags are included as long as they account for a significant additional element of the variation. This is valuable in accurately forecasting abundance on the basis of past abundance, and in identifying the combination of density-dependent processes that are instrumental in driving the population's dynamics, but also, as we shall see, in shedding light on the number and nature of those underlying processes. An ACF cannot help us with this. It tells us about patterns in the dynamics but not about underlying processes. To investigate these, a frequent next step is to estimate the partial autocorrelation function (PACF): the correlation between present log abundance and log abundance at successive lags, having accounted, statistically, for all correlations at shorter lags. This is shown for the larch budmoth in Figure 14.5c. The correlation at lag 1 is of course the same, since there is no shorter lag to take into account, but thereafter the pattern is very different. Firstly, there is a large negative partial correlation at lag 2, indicating delayed density dependence acting with a delay of two years: high current densities will have a large negative impact on densities in two years' time. Secondly, and without going into statistical details, only the first two lags are necessary to model this population – the next is close to zero.

However, PACFs also have their shortcomings in this context (see Berryman & Turchin (2001) for a full discussion). In particular, the correlation at lag 1 is positive only because of the almost inevitable tendency for adjacent values in a population time series to be similar to one another. That correlation therefore says nothing about the influence of density on an individual's contribution to abundance in one year's time, which is the essence of density dependence. This though can be seen through the use of a partial rate correlation function (PRCF), which regresses the population growth rate, r_t, against log abundance at successive lags (Equation 14.1). The pattern for the larch budmoth (Figure 14.5d) confirms the appropriateness of a model with only two lags, but now we can see that there is both direct (lag 1) and delayed density dependence.

Fennoscandian microtines

An example of the use of this approach is illustrated in Figure 14.6, which summarises analyses of 19 time series of microtine rodents (lemmings and voles) from various latitudes in Fennoscandia (Finland, Sweden and Norway) sampled once per year (Bjørnstad *et al.*, 1995). In almost all cases, as with the larch budmoth, the optimum number of lags was two, and so the analysis proceeded on the basis of these two lags: (i) direct density dependence; and (ii) density dependence with a delay of one year.

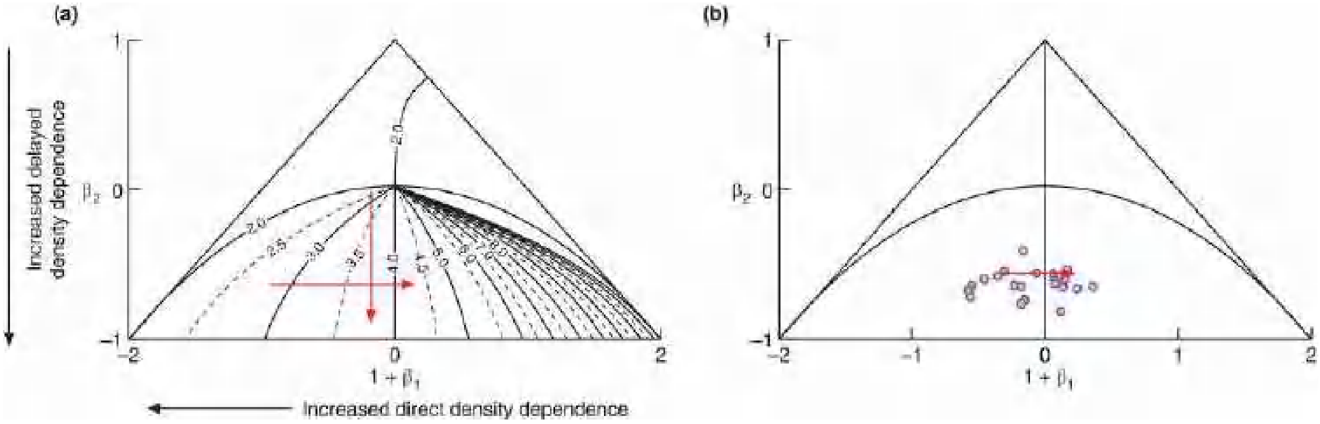


Figure 14.6 Analysis of microtine rodent population time series summarised as linear autoregressive models of order two. (a) The type of population dynamics generated by a linear autoregressive model (see Equation 14.2) incorporating direct density dependence, β_1, and delayed density dependence, β_2. Parameter values outside the triangle lead to population extinction. Within the triangle, the dynamics are either stable or cyclic and are always cyclic within the semicircle, with a period (length of cycle) as shown by the contour lines. Hence, as indicated by the red arrows, the cycle period may increase as β_2 decreases (more intense delayed density dependence) and especially as β_1 increases (less intense direct density dependence). (b) The locations of the pairs of β_1- and β_2-values, estimated from 19 microtine rodent time series from Fennoscandia. The red arrow indicates the trend of increasing latitude in the geographic origin of the time series, suggesting that a trend in cycle period with latitude, from around three to around five years, is the result of a decreased intensity of direct density dependence.

Source: After Bjørnstad *et al.* (1995).

Having established that this is the structure of the density dependence for these populations, Figure 14.6a gives a general picture of the dynamics of a population with this structure (Royama, 1992). Remember that delayed density dependence is reflected in a value of β_2 less than 0, while direct density dependence is reflected in a value of $(1 + \beta_1)$ less than 1. Thus, we can see that populations not subject to delayed density dependence tend not to exhibit cycles, but β_2-values less than 0 generate cycles, the period (length) of which tends to increase both as delayed density dependence becomes more intense (down the vertical axis) and especially as direct density dependence becomes less intense (left to right on the horizontal axis).

The results for the rodent populations are then set out in Figure 14.6b. The estimated values of β_2 for the 19 time series showed no trend as latitude increased, but the β_1-values increased significantly. It was known prior to the analysis, from the data themselves, that the rodents exhibited cycles in Fennoscandia and that the cycle length increased with latitude (see Section 14.6.3, below). The data in Figure 14.6b point to precisely the same pattern, and crucially, they suggest that the reason lies in the structure of the density dependences: on the one hand, a strong delayed density dependence throughout the region, and on the other hand, a significant decline with latitude in the intensity of direct density dependence. As we shall see in Section 14.6.3, this is consistent with the 'specialist predation' hypothesis for microtine cycles. Remember, however, that the method in itself tells us nothing about the biological bases of the forces acting on the population, only the likely structure of the density dependences. No doubt other underlying mechanisms are also consistent with this structure.

hares and lynx display three and two dimensions, respectively

In other, related cases, the emphasis has been on deriving the optimal statistical model because the number of lags in that model may provide clues as to how abundance is being determined – the number of lags reflecting the number of important interacting elements in the system itself. One example is the study by Stenseth *et al*. (1997) of the hare–lynx system in Canada, already mentioned briefly in Section 10.1.4. In this case, the optimal model for the hare time series included direct density dependence (self-regulation amongst the hares) plus further, delayed density dependence at one year (only very weak) and two years (much stronger) – hence two lags and three interacting elements overall. For the lynx, the model was simpler – just one additional lag and hence two interacting elements – including only negligible direct density dependence but strong density dependence with a delay of one year. Combined with what is known of the whole community of which the hare and lynx are part (Figure 14.7), this led Stenseth *et al*. to propose, for the lynx dynamics, a model comprising only the lynx themselves and the hares, since the hares are by far the lynx's most important prey. However, for hare dynamics they proposed a model comprising the hares themselves, 'vegetation' (since hares feed relatively indiscriminately on a wide range of vegetation), and 'predators' (since a wide range of predators feed on the hares). We return to the hares and lynx in Section 14.6.2, below.

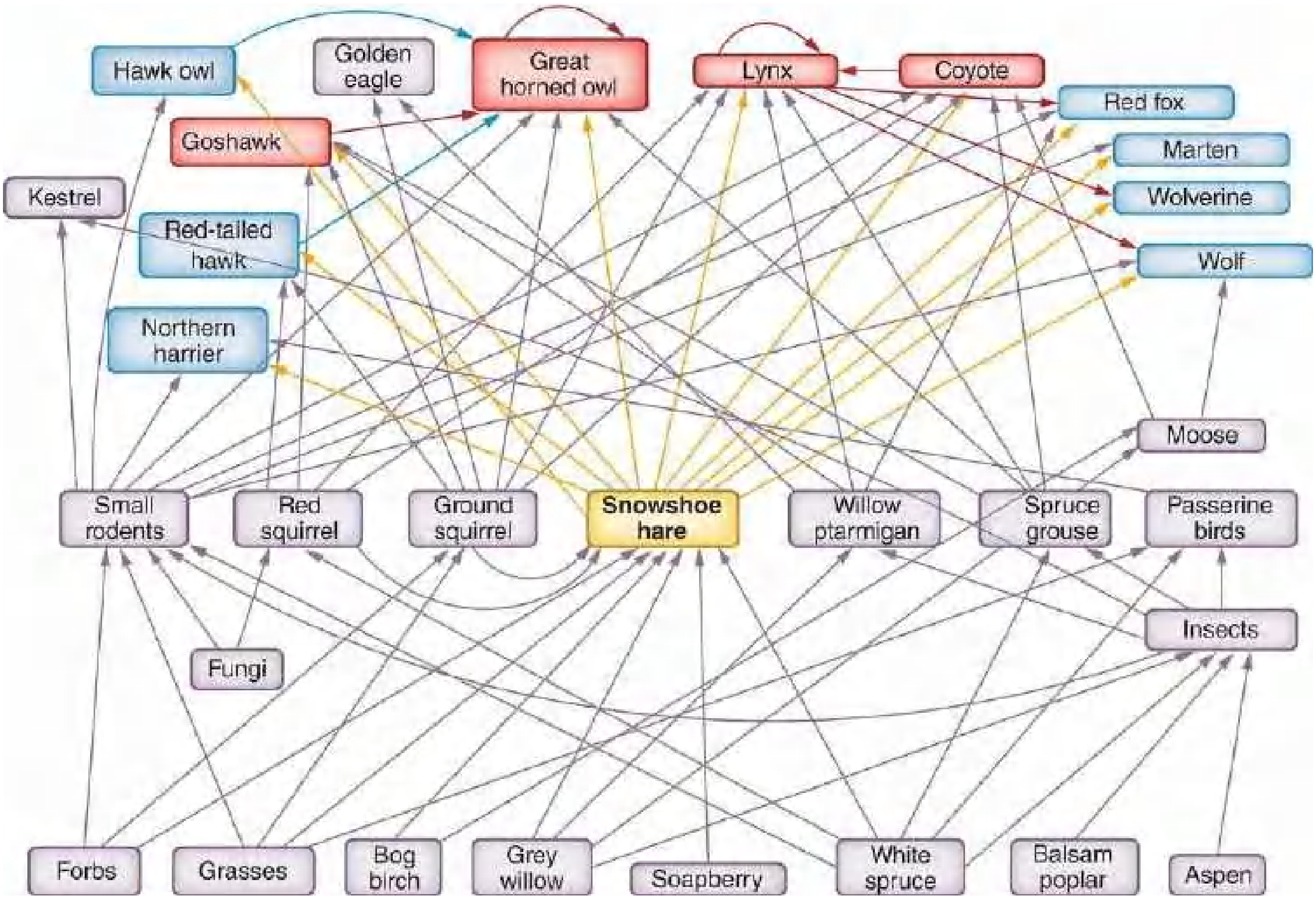


Figure 14.7 The food web around the snowshoe hare. The major components of the food web – plants and fungi, herbivores, carnivores – for the forests of the south-western Yukon near Kluane Lake, Canada. Arrows point from consumed to consumer. Major predators of the snowshoe hare are indicated in red, minor predators in blue.

Source: After Krebs (2011).

Similar analyses have been applied to time series we examined in Figure 14.1. A four-year cycle was supported for the annual flixweed, *Descurainia sophia*, in Spain, by the pattern in the ACF, driven by a two-year delayed density dependence indicated by the PACF (Figure 14.8a). The authors speculate that this delay may reflect interactions between the plant and litter and nutrient accumulation (Gonzalez-Andujar *et al.*, 2006). And for the common wasps, *Vespula vulgaris*, in England, the absence of any regular patterns was confirmed by the ACF, but the PRCFs were remarkably similar for the two populations (Figure 14.8b) – an underlying consistency in the density-dependence structure despite the strikingly different patterns in the time series – indicating powerful direct density dependence, which the authors were able to combine with weather variables, especially in the spring, to account for the detailed patterns observed (Lester *et al.*, 2017).

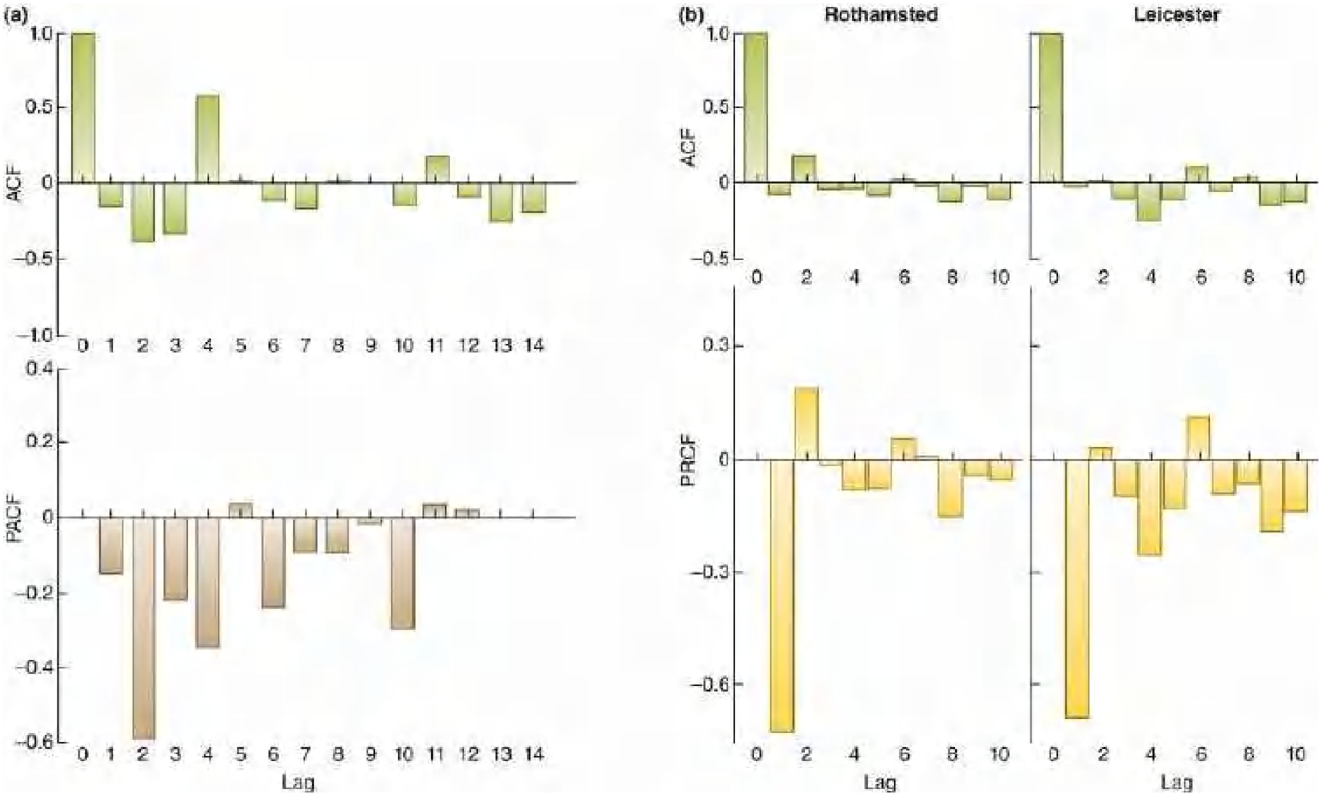
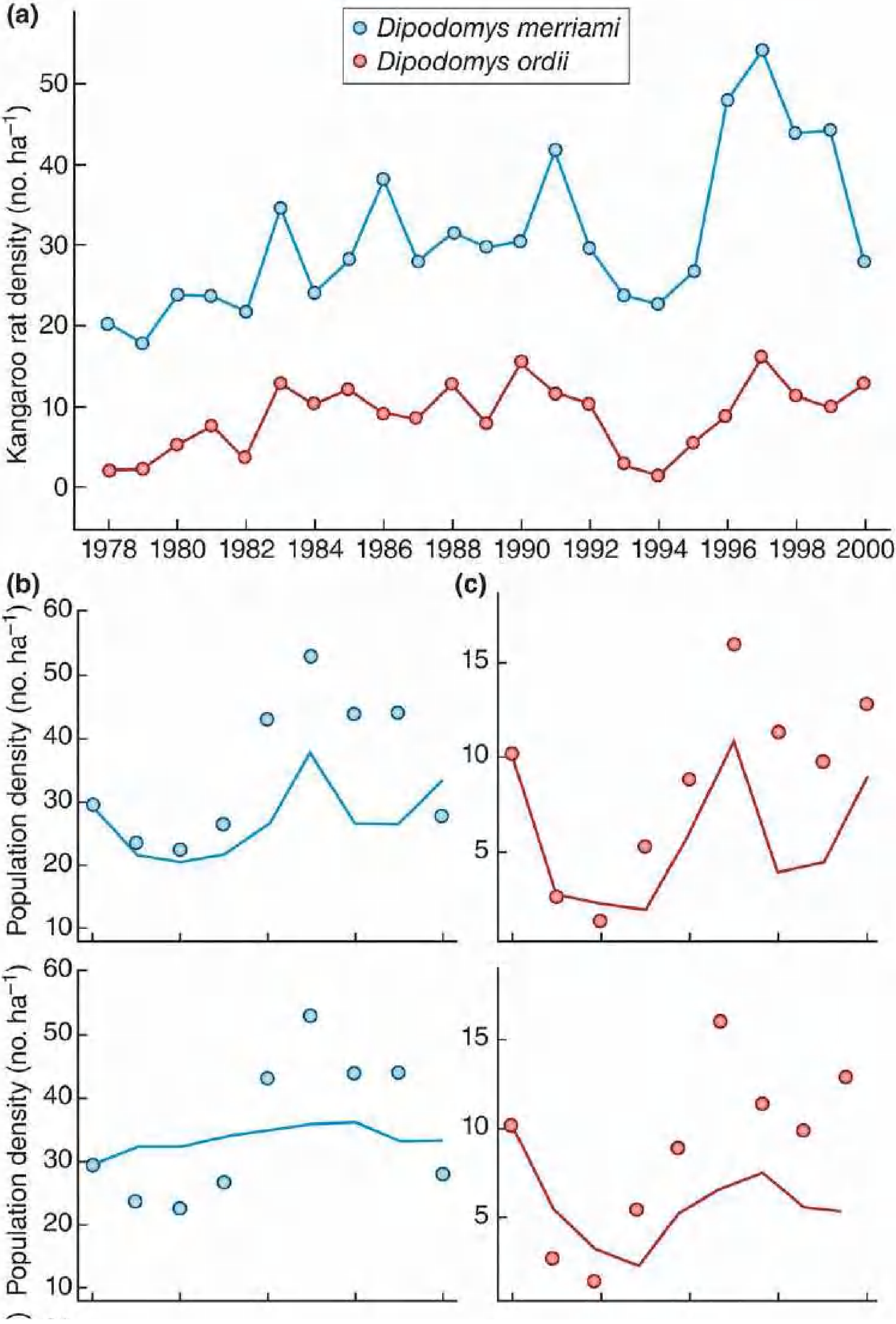


Figure 14.8 Analysis of flixweed and wasp population time series using correlation functions. (a) The autocorrelation function (ACF) and partial rate correlation function (PACF) for the time series of *Descurainia sophia* abundance shown in Figure 14.1a. (b) The ACF and PRCF for the two time series of *Vespula vulgaris* abundance shown in Figure 14.1b.

Source: (a) After Gonzalez-Andujar *et al.* (2006). (b) After Lester *et al.* (2017).

combining density dependence and weather

To take a closer look at how this autoregressive approach to the study of density dependence can be combined with other factors, such as weather, we turn to a study of kangaroo rats in the Chihuahuan desert of Arizona, USA (Lima *et al.*, 2008). Time series data for two species, *Dipodomys merriami* and *D. ordii*, are shown in Figure 14.9a. Calculations of PRCF values for these time series showed that the correlations were significant (and negative) with a one-year lag in both cases (−0.57 and −0.54, respectively, both $P < 0.05$), but there were no significant correlations with longer lags. Based on this, Lima *et al.* used as a starting model a version of Equation 14.1 with only the first two terms included, though rather than having a simple coefficient to represent the direct density dependence (β_1 in Equation 14.1) they used a more complex formulation closely related to the model of intraspecific competition described in Equation 5.18 (Section 5.6.4). This allowed them to estimate separately the different components of self-regulated population growth: the population growth rate, the intensity of competition and the type of compensation of the density dependence (Section 5.6.4). Crucially, though, they were able to compare the capacity of this equation to account for the time series in Figure 14.9a with the capacities of more complex equations that included effects of winter and summer rainfall on the various components already in the model, and of competition from another kangaroo rat, *D. spectabilis*. To add rigour to the comparisons between the different models, they used the data up to 1992 to estimate their parameters, and then tested the models against the data from 1993 onwards. The results are shown in Figure 14.9b and c.


(a)
Dipodomys merriami
Dipodomys ordii
Kangaroo rat density (no. ha^{-1})
50
40
30
20
10
0
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
(b)
Population density (no. ha^{-1})
60
50
40
30
20
10
(c)
15
10
5
Population density (no. ha^{-1})
60
50
40
30
20
10
15
10
5

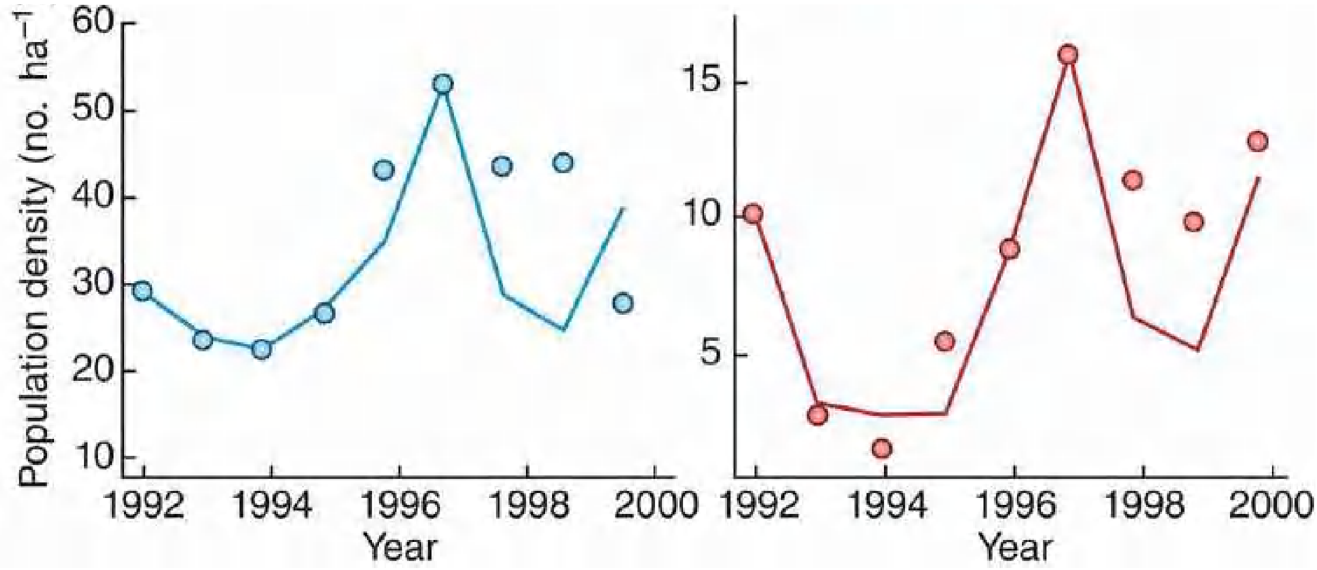


Figure 14.9 **Analysis of population time series of kangaroo rats using linear autoregressive models.** (a) Time series of the abundance of kangaroo rats, *Dipodomys merriami* and *D. ordii* in the Chihuahuan desert, Arizona, USA. (b) The fit, for *D. merriami*, of models based on data from 1978 to 1992 to the time series from 1993 to 2000: top, the best model including self-regulation and summer rainfall effects; middle, the best model including self-regulation and interspecific competition with *D. spectabilis*; bottom, the best model including self-regulation and both of these. (c) Equivalent results for *D. ordii*.

Source: After Lima *et al.* (2008).

For both species, models that included rainfall or interspecific competition performed better that those that included only intraspecific effects, and winter rainfall led to very little improvement in performance compared to summer rainfall. However, as the figures show, models that included both summer rainfall and interspecific competition performed best of all, though not perfectly of course. The autoregressive analysis of time series, therefore, allows us not only to identify and then explore density dependencies that appear to be driving a population's dynamics, but also to integrate these with other effects, both biotic and abiotic, as long as they can be measured (the amount of summer rainfall, or the abundance of *D. spectabilis*).

14.6 Population cycles and their analysis

Regular cycles in animal abundance were first documented in the long-term records of fur-trading companies and gamekeepers. Cycles have also been reported from many studies of small rodents, especially voles and lemmings, in certain forest Lepidoptera (Figure 14.5), and in some plants (Figure 14.1a). Population ecologists have been fascinated by cycles at least since Elton drew attention to them in 1924. In part, this is attributable simply to some sense of a striking pattern crying out for an explanation. But there are also sound scientific reasons for the preoccupation. Firstly, cyclic populations, by definition, exist at a range of densities. They therefore offer good opportunities (high statistical power) for detecting any density-dependent effects that exist, and integrating these with density-independent effects in an overall analysis of abundance. Also, regular cycles constitute a pattern with a high ratio of 'signal' to 'noise' (compared, say, with totally erratic fluctuations, which may appear to be mostly noise). Since any analysis of abundance is likely to seek ecological explanations for the signal and to attribute noise to stochastic perturbations, it is obviously helpful to know clearly which *is* signal and which is noise.

extrinsic and intrinsic factors

Explanations for cycles are often classified as emphasising either extrinsic or intrinsic factors. The former, acting from outside the population, may be food, predators or parasites, or some periodic fluctuation in the environment itself. Intrinsic factors are changes in the organisms themselves: changes in aggressiveness, in the propensity to disperse, in reproductive output, and so on. Below we examine studies on population cycles in three systems, all touched on previously: the red grouse (Section 14.6.1), the snowshoe hare (and lynx) (Section 14.6.2) and microtine rodents (Section 14.6.3). In each case, it is important to bear in mind the problems of disentangling cause from effect; that is, of distinguishing factors that *affect* density from those that merely vary *with* density. Equally, we must try to distinguish the factors that affect density (albeit in a cyclic population) from those that actually impose a pattern of cycles (see also Berryman, 2002 ; Turchin, 2003).

14.6.1 Red grouse

The explanation for cycles in the dynamics of the red grouse (*Lagopus lagopus scoticus*) in the UK has been a matter of disagreement for decades (Martinez-Padilla *et al*., 2014). Some have emphasised an extrinsic factor, the parasitic nematode *Trichostrongylus tenuis* (Dobson & Hudson, 1992 ; Hudson *et al*., 1998). Others have emphasised an intrinsic process through which increased density leads to more aggressive interactions, which in turn leads to wider territorial spacing, and, with a delay because this is maintained into the next year, to reduced recruitment (Watson & Moss, 1980 ; Moss & Watson, 2001). Both viewpoints, therefore, rely on a delayed density dependence to generate the cyclic dynamics (see Section 10.1.2), though these are arrived at by very different means (see Figure 14.10).

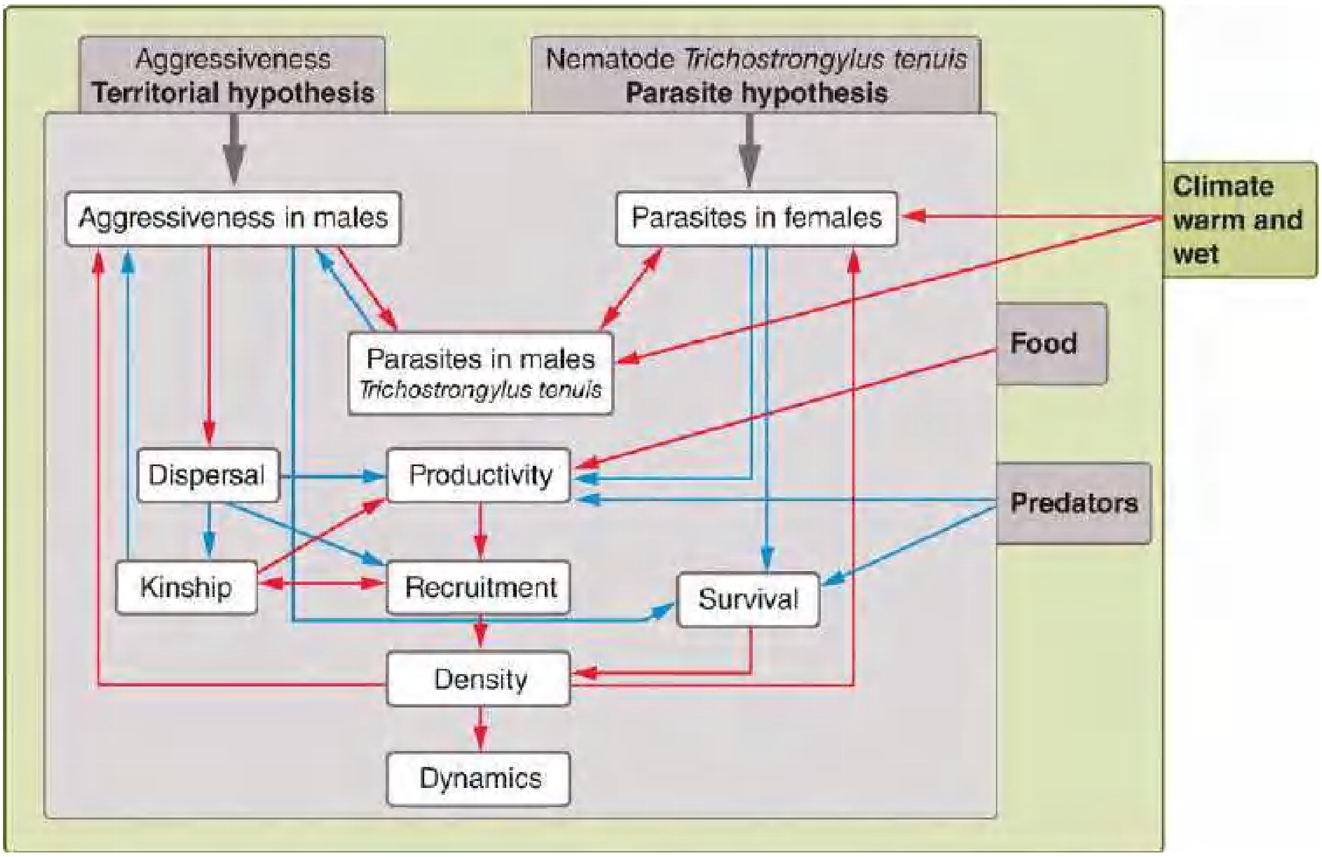


Figure 14.10 The two 'competing' theories for the generation of population cycles in red grouse, *Lagopus lagopus scoticus*, territoriality and parasitism, and the ways in which they act and may interact. Red arrows denote positive effects, blue arrows negative effects.

Source: After Martinez-Padilla *et al*. (2014).

We have already seen in Sections 12.8.1 and 14.4.1 that even field-scale experiments have been unable to determine the role of the nematodes with certainty. There seems little doubt that they can reduce density, and the results of field experiments are consistent with their determining the amplitude of the cycles. But they do not seem to be necessary for cycles to occur, and so they cannot, at least universally, be necessary for generating them in the first place.

In another field experiment, aspects of the alternative, 'territorial behaviour' hypothesis were tested (Mougeot *et al.*, 2003b). In experimental areas, established males were given testosterone implants at the beginning of the autumn, when territorial contests take place. This increased their aggressiveness (and hence the size of their territories) at densities that would not normally generate such aggression. By the end of the autumn, it was clear that, relative to the control areas, the increased aggression of the older males had reduced the recruitment of the younger males: testosterone treatment had significantly reduced male densities and had particularly reduced the ratio of young (newly recruited) to old males. Moreover, in the following year, even though the direct effects of the testosterone had worn off, the young males had not returned, and the per capita recruitment was therefore lower both in the treatment year and the following year. Thus, these results establish, at least, the potential for intrinsic processes to have (delayed) density-dependent effects on recruitment, and thus to generate cycles in the grouse. In a companion paper, Matthiopoulos *et al.* (2003) demonstrate how changes in aggressiveness can indeed cause population cycles.

What remains unanswered by the Mougeot experiment, however, is what mechanism normally drives the variations in aggression with density. One popular suggestion has been a 'kinship hypothesis' in which at low densities, male grouse that hold territories show little aggression toward their own kin (especially offspring) and so tolerate their establishing territories close to their own, promoting clusters of small territories from which recruitment is high. The 'rich' (the territory holders) keep their wealth within the family. However, as density increases and space becomes limited, tolerance declines, relatedness between neighbours (kinship) therefore declines, and aggressiveness increases further, driving density down as territory sizes are increased and recruitment decreased. Support for the kinship hypothesis has come from observations of kin clusters (groups of related males) building up during the increase phase of the cycle but then breaking down before the peak, and from recruitment being positively related to the size of these kin clusters (Piertney *et al.*, 2008), and from levels of kinship being reduced, too, in the spring following the experimental boosting of aggressiveness (Mougeot *et al.*, 2005).

However, there is also mounting evidence (reviewed in Martinez-Padilla *et al.*, 2014) that parasitism and aggressiveness interact (see Figure 14.10). For example, birds treated to remove parasites are more aggressive and more successful at winning territorial contests; and males treated with testosterone implants are more susceptible to the acquisition of parasites. Furthermore, there are both observational and experimental links between the parasite burdens in the male and female of a pair, which is important given the central role of males in the case of kinship/aggressiveness, and of females (through fecundity) in the case of parasites. Hence, it seems increasingly likely that both parasites and kinship have key roles to play, together, in an overall explanation for the red grouse cycles.

14.6.2 Snowshoe hares

The '10-year' hare and lynx cycles of northern Canada have also been examined in previous sections. The earliest data, based on records of fur pelts, go back to the 17th century, and the cycles continue today (Figure 14.11). Stenseth *et al.*'s (1997) time series analyses (Section 14.5), suggested that despite becoming a 'textbook' example of coupled predator–prey oscillations, the hare cycle may be generated by interactions with both its food and its predators, while the lynx cycle does indeed appear to be generated by its interaction with the snowshoe hare. As we shall see next, empirical, especially field-manipulation studies provide support for this same web of interactions but also modify and extend it (Krebs, 2011 ; Krebs *et al.*, 2017).

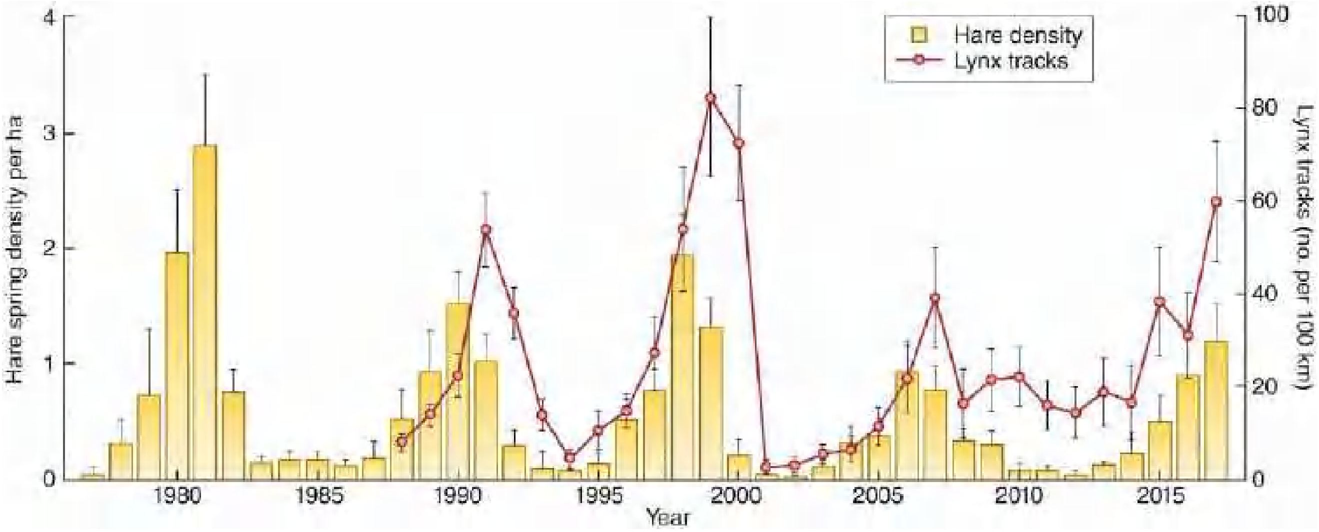


Figure 14.11 **Hare and lynx cycles.** Regular cycles in the abundance of the snowshoe hare, *Lepus americanus*, and the Canada lynx, *Lynx canadensis*, 1977–2017, at Kluane Lake, Yukon, Canada. Bars are 95% CIs.

Source: Charles J. Krebs, personal communication.

The demographic patterns underlying the hare cycle are well established. Both fecundity and survival begin to decline before peak densities are reached, but do not arrive at their minima until around two years after the peak. The immediate cause of hare death, in the vast majority of cases, appears to be predation rather than starvation, but fecundity remains low even after predator numbers have also dropped to low levels and when there is no obvious shortage of food for the hares. It is this extended delay that accounts for the relatively lengthy period of the hare cycle. Lynx abundance follows that of the hare with a delay of around one year.

Work on the system to try to understand these cycles has been in the vanguard in the application of field-scale manipulations (Krebs *et al.*, 2017). Experiments in which artificial food was added, or natural food was supplemented, or food quality was enhanced, all pointed in the same direction. Food supplementation improved individual condition and led to higher densities but did nothing to prevent the decline phase of the cycle, where, again, the overwhelming cause of death was predation. On the other hand, experiments in which either predators were excluded, or they were excluded and food was also supplemented, had much more dramatic effects, and survival was highest of all when food supplementation and predator exclusion were combined (Figure 14.12a). This is consistent with Stenseth *et al.*'s (1997) modelling in suggesting a role for both food and predators in the cycle, but suggests a dominant role for predation.

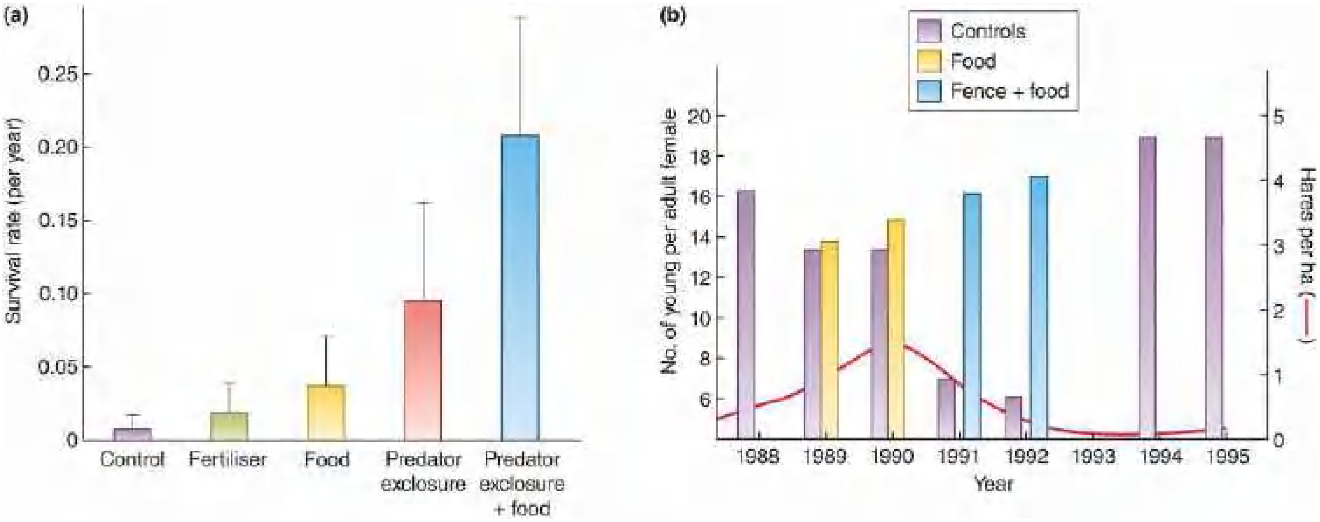


Figure 14.12 Survival and reproduction of the cyclic snowshoe hare. (a) Annual survival rates of snowshoe hares at Kluane Lake, Yukon, Canada in treatment categories as indicated, during the decline phase of their abundance cycle from 1990 to 1992. The fertiliser treatment improved food quality. Bars are 95% CIs. (b) Reproductive output over a hare cycle from 1988 to 1995 at Kluane Lake (red line). It was possible to compare control values with those from treatments of food supplementation in 1989 and 1990, and with those where food was supplemented and mammalian predators excluded in 1991 and 1992.

Source: (a) After Krebs *et al.* (2017). (b) After Krebs *et al.* (2001).

Furthermore, food supplementation slightly reduced the initial decline in fecundity prior to peak densities (Figure 14.12b), but the combination of food supplementation and predator exclusion brought fecundity up to almost maximum levels at what would otherwise be the phase of lowest fecundity following the density peak. Unfortunately, it was not possible to measure fecundity in a treatment where only food was supplemented – an example of the disappointments that almost inevitably accompany large field experiments – so the effects of food and predators could not be disentangled, leaving unexplained the extended period of low fecundity in the hares, continuing beyond the decline in hare (and predator) abundance.

Further experiments, however, shed light on this. It is well known that fecundity is reduced by high levels of corticosteroid hormones, which are associated with stress in mammals. This has been confirmed in the hares: at higher concentrations of corticosteroids, litters are smaller and individual offspring lighter (Figure 14.13a, b). The concentrations themselves vary over the course of the cycle – indeed, peaking after the peak in density (Figure 14.13c) so that, as with aggressiveness in grouse, stress in hares is highest after the peak density has passed. And crucially, there is a clear ‘maternal effect’: mothers with high concentrations of the hormones give birth to offspring with high concentrations (Figure 14.13d), giving rise to females at relatively low density and low predation risk having reduced fecundity as a result of the stress experienced by their mothers.

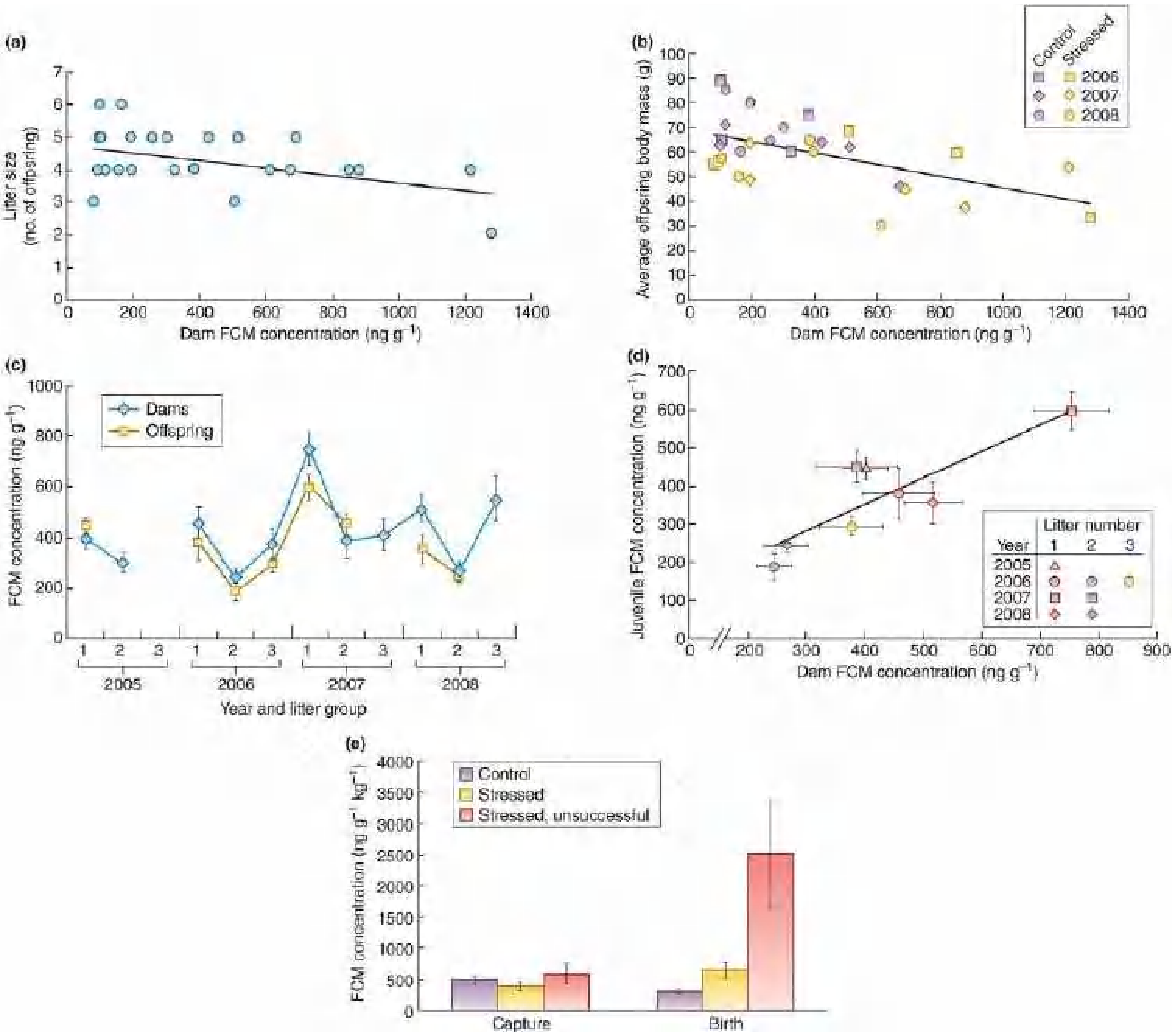


Figure 14.13 **The effects of stress hormones in the cyclic snowshoe hare.** (a) and (b) show, respectively, for free-ranging snowshoe hares, that litter sizes were smaller ($P < 0.05$) and individual offspring lighter ($P < 0.05$) when concentrations of corticosteroid hormones in the mother (dam) were greater, measured as the concentration of corticosteroid metabolites in the faeces (FCM). (c) The variation for these free-ranging hares in adult and juvenile FCM concentration from 2005 to 2008 (from the increase to the decline phase in their cycle). Bars are SEs. (d) The association between FCM concentrations in free-ranging dams and their juvenile offspring in successive litters from these years. Bars are SEs. The general least squares regression line is shown ($r^2 = 0.73$, $P = 0.007$). (e) FCM concentrations, both when they were captured and when they subsequently gave birth, in control pregnant females and pregnant females that were subjected to the stress of simulation predator exposure, or were subjected to that exposure and failed to give birth. Bars are SEs.

Source: (a, b, e) After Sherriff *et al.* (2009). (c, d) After Sherriff *et al.* (2010).

The question remains, though, as to what gives rise to this stress. One possibility is exposure to, and experience of, the predators (see Section 9.6.1), and the data in Figure 14.13e lend support to this. Pregnant hares were captured and either subjected to a daily episode of exposure to a predator for the last 15 days of their gestation (a trained dog was taken on a leash into the hares' enclosure) or left as unstressed controls. The different groups of hares had similar levels of stress hormones at the outset, but following treatment, stressed hares had significantly higher hormone levels than the controls, and of the stressed hares, those that failed to give birth had the highest levels of all. We therefore have a series of links, all supported by experimental evidence – from high ratios of predators to hares around the peak of hare abundance, to high exposure rates of hares to predators, to high levels of stress hormones in these hares, and thence to reduced fecundity and also to high hormone levels in their offspring, giving rise in turn to lowered fecundity some time after the peak in density.

Taken together, this suggests roles for both extrinsic and intrinsic factors in the hare cycles, but it reminds us that intrinsic factors (in this case high levels of stress) are unlikely to be fundamentally intrinsic in the sense of being self-generated, but rather, are likely to be generated themselves by extrinsic factors. In this case, there is evidence for exposure to predators generating these intrinsic changes, but this does not rule out the possibility that shortage of high-quality food may also contribute to high levels of stress and ultimately to reduced fecundity. Either way, the maternal effect, through which the offspring of stressed mothers themselves suffer reduced fecundity, seems crucial in driving the delayed response of hares to reductions in predation pressure and food shortage, and thus crucial, too, in generating the 10-year cycle. Finally, while the effects of predators are clearly profound, not only in driving these intrinsic changes but also in accounting, ultimately, for most hare deaths, it remains an open question whether the susceptibility of hares to predation is itself increased by infection (the immune response is compromised under stress) or food shortage, as we have seen in previous chapters (Krebs *et al.*, 2017).

14.6.3 Microtine rodents: lemmings and voles

many microtines cycle – but not all

More effort has been expended overall in studying population cycles in microtine rodents (voles and lemmings) than in any other group of species. These cyclic dynamics, with periods typically three or four years in length, or much more rarely two or five years or even longer, have been convincingly identified in voles (*Microtus* spp. and *Myodes* spp.) in Fennoscandia (Finland, Norway and Sweden); lemmings (*Lemmus lemmus*) elsewhere in montane habitats in Fennoscandia; lemmings (*Lemmus* spp. and *Dicrostonyx* spp.) in the tundra of North America, Greenland and Siberia; voles (*Myodes rufocanus*) in Hokkaido, northern Japan; common voles (*Microtus arvalis*) in central Europe; and field voles (*Microtus agrestis*) in northern England. There are other repeated features in these cycles: synchrony between different species in the same area (suggesting a common driver), synchrony across large geographic areas, and a tendency for individuals to be large in the increase and peak phases and small in the decline and low phases (sometimes called the 'Chitty effect') (Hanski & Henttonen, 2002 ; Krebs, 2011). A quite different pattern, of irregular and spectacular irruptions in abundance and mass movement, is shown by just a few lemming populations, notably in Finnish Lapland. It is these whose apparently suicidal behaviour has been so misrepresented in the name of film-makers' poetic licence, unfairly condemning all lemmings to popular misconception (Henttonen & Kaikusalo, 1993). On the other hand, there are other microtine populations that show no evidence of multiannual cycles, including voles in southern Fennoscandia, elsewhere in Europe, and many locations in North America (Turchin & Hanski, 2001).

Given this diversity of species, habitats and dynamics, the first question is whether to expect a single explanation to apply to them all, and if so, whether any such generality should apply to the details (is it food? is it predation?) or only to the underlying structure of that explanation. Clearly, it is more likely to be the latter. We can acknowledge that the cycles are the result of a 'second-order' process (Bjørnstad *et al.*, 1995 ; Turchin & Hanski, 2001) (see Section 14.5); that is, that they reflect the combined effects of direct and delayed density-dependent processes. But it is less likely that the direct and delayed processes will be the same in every cyclic population. What is important is that two such processes act in conjunction.

the specialist predation model

The search for an explanation has focused especially on the vole (and to a lesser extent lemming) populations of Fennoscandia, simply because these have been the most intensively studied. The most influential explanation for the cycles has been provided by the '*specialist predation model*', which has itself been subject to successive refinements (Hanski *et al.*, 2001). The model has already been touched on in Sections 10.3.2 and 14.5. It has the following key features: (i) logistic population growth in the microtine prey, to reflect the directly density-dependent effects of food shortage on the microtines, preventing their populations from growing too large before specialist predators 'catch up'; (ii) specialist predators (especially weasels, but also other mustelids) with a population growth rate that declines as the ratio of specialist predators to prey increases, acting in a delayed density-dependent manner; (iii) generalist predators (mammals that switch between prey types, or wide-ranging avian predators) that act in a directly density-dependent manner by responding immediately to changes in microtine density; and (iv) seasonal differences in the breeding of voles and weasels in the summer and winter, incorporated by replacing constant vital rates with ones that vary over the course of the year according to a sine wave function. Despite its name, therefore, the model includes both predators *and* food. Food provides the baseline direct density dependence (though, of course, other directly density-dependent processes would justify the same model formulation), while specialist predators provide the delayed density dependence. Generalist predators then provide a further source of direct density dependence that can be varied to mimic both their known decline in abundance moving north, and the parallel decline in their contact with their microtine prey, as the length of the snow-cover period increases: voles and weasels live below the snow, but the generalist predators do not.

experimental support?

In terms of experimental support for the model, field manipulations in which predators were removed have typically generated a common pattern of results (Klemola *et al.*, 2000 ; Hanski *et al.*, 2001 ; Korpimaki *et al.*, 2002). Vole densities have been increased several-fold, up to 20-fold, but the populations have still crashed, and the pattern of cycles has therefore been maintained. The various experiments have also been subject to the same type of criticism: that they were too short term, or small scale, or they affected too many, or too few, of the predator species (failing often to distinguish specialist and generalist predators), and they often involved the erection of protective fences that are likely to have affected movements of the prey (voles) as well. Conclusive experiments may be crucial, but this does not make them any easier! Thus, while the results of these experiments have indicated an important role for predators in vole survival and abundance, they have failed to prove a role for the predators in causing (as opposed, say, to amplifying) the vole cycles. In particular, there is little or no experimental support for an implicit but crucial prediction of the model – that the specialist predators' abundance exhibits a delayed numerical response to microtine abundance, reflecting their delayed density-dependent effect on their prey (Lambin, 2018).

supportive prediction?

On the other hand, when the model is parameterised with field data on the species' vital rates, predation rates and so on, it can recreate an impressive number of the features of the observed dynamics. Cycles are of broadly the correct amplitude and period, and both the period and indeed the amplitude of the cycles increase with latitude as the density of generalist predators decreases, as observed in nature (Figure 14.14). A related model for the collared lemming, *Dicrostonyx groenlandicus*, preyed upon by one specialist predator (the stoat, *Mustela erminea*) and three generalists (Gilg *et al*., 2003) was also able to recreate observed cycles in Greenland when parameterised with field data. Note, however, that consistencies between the model's output and field data do not prove that the model is correct – only that it is not *in*correct.

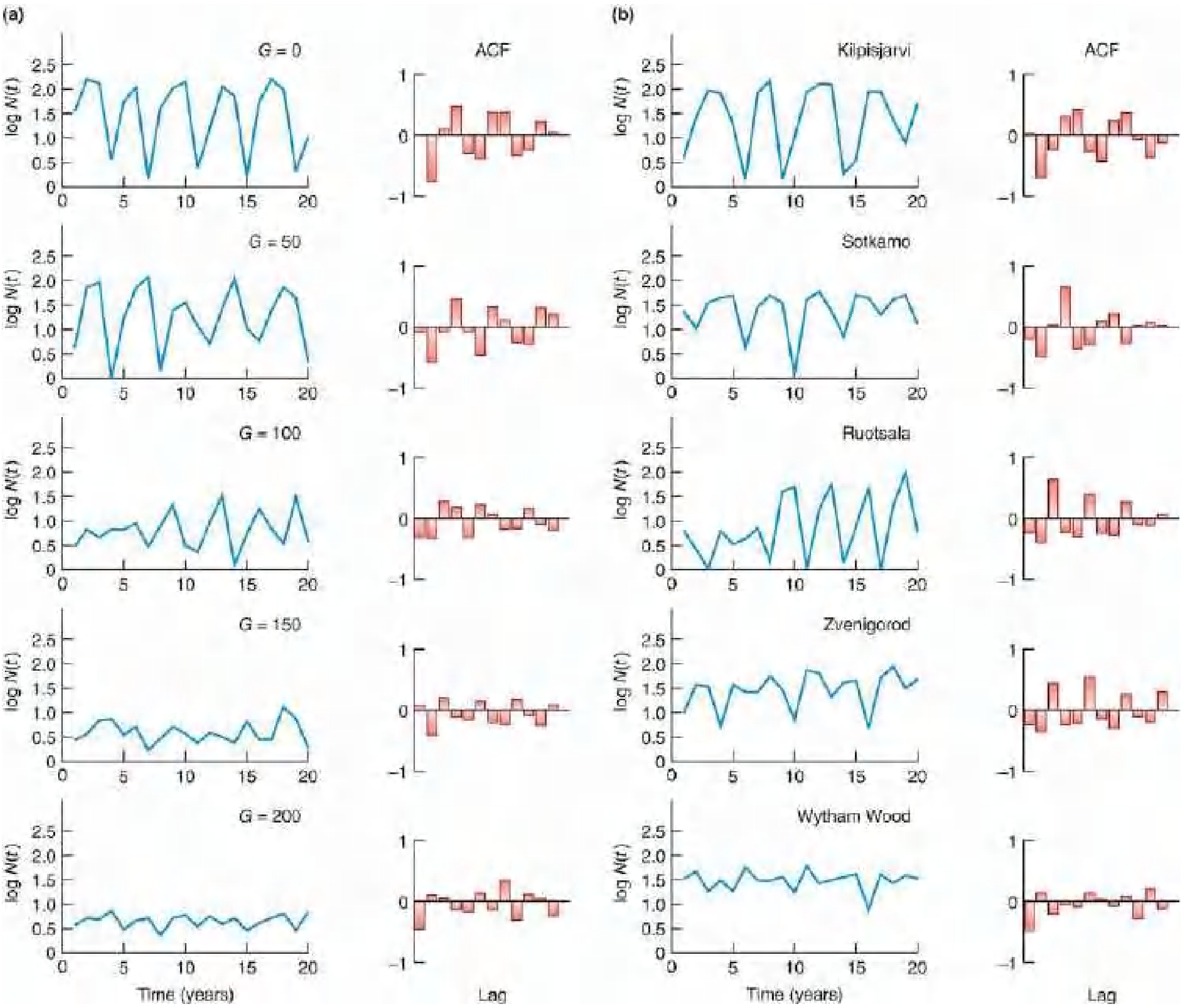


Figure 14.14 Analyses of model and microtine time series are consistent with the specialist predation hypothesis. (a) Sample data generated by the specialist predation model, and the associated autocorrelation functions (ACFs), for various values of generalist predator abundance, *G*. As *G* increases, cycle period increases and cycle amplitude decreases, and at sufficiently high values the dynamics are sufficiently stabilised for the cycles to disappear altogether. (b) Comparable time series from five field sites: Kilpisjarvi (69°N; period = 5), Sotkamo (64°N; period = 4), Ruotsala (63°N; period = 3), all in Finland, Zvenigorod, Russia (57°N; period = 3) and Wytham Wood, UK (51°N; no significant periodicity).

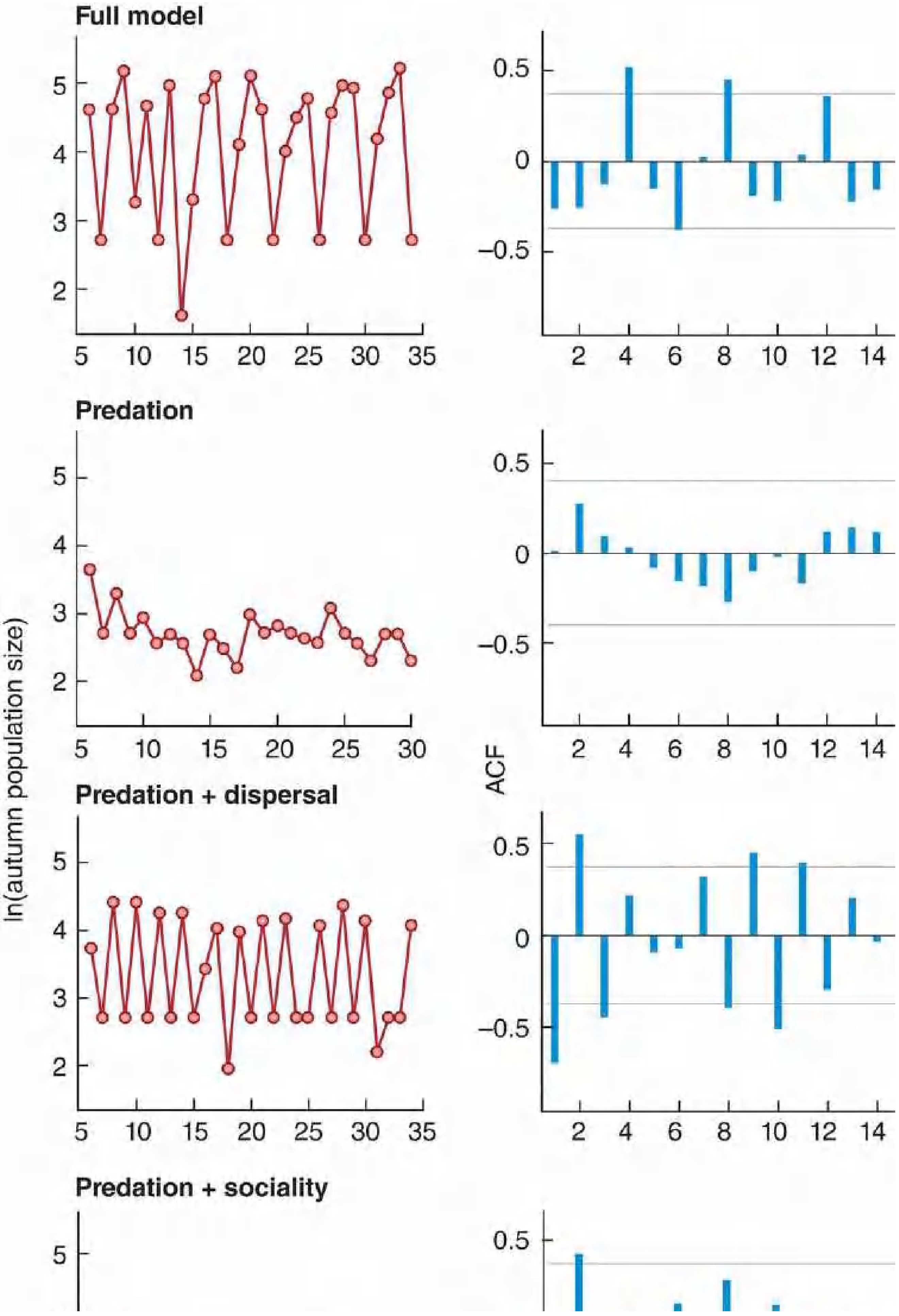
Source: After Turchin & Hanski (1997).

Indeed, not all studies have conformed to the predictions of the specialist predation model. Lambin *et al*. (2000) described regular cycles of field voles in Kielder forest, northern England (55°N), with a period of three to four years and an approximately 10-fold difference between peak and trough densities (a difference of 1 on a log scale, such as in Figure 14.14). Yet, parameterising the specialist predation model with the estimated intensity of generalist predation at this site would have predicted no cycles whatsoever. Moreover, a rigorous programme of reducing the numbers of the specialist predators, weasels, by about 60% in comparison with control sites, increased adult vole survival by about 25% but had no appreciable impact on the cyclic dynamics (Graham & Lambin, 2002). Again, too, the weasels showed no lagged numerical response to field vole abundance, which is perhaps not surprising given the relative brevity of their maturation period and inter-litter interval (Lambin, 2018). Likewise, Lambin *et al.* (2006) described regular cycles of the common vole, *Microtus arvalis*, in south-west France with many of the characteristics of the Fennoscandian cycles, but found negligible evidence of specialist predation. They interpret this as grounds for doubting the specialist predation model as a general explanation for microtine cycles, which in turn invites the question as to what, precisely, would be the ecological basis for the specialist predation model explaining cycles in Fennoscandia but not elsewhere. An alternative, as we have noted, would be to expect a structural similarity in the underlying model, but accept differences in detail between systems – specialist predators contributing the delayed density dependence in Fennoscandia, for example, but something else (as yet unknown) performing the same role in France and probably other systems. The attractions of a common explanation are easy to see, in terms of parsimony (not making something more complicated than it needs to be), and there are dangers in too readily compiling a series of plausible special cases, but these attractions cannot override an acceptance that not all microtines are the same, if that turns out to be the case.

intrinsic factors

Even for those populations where the specialist predation model has found support, interest in intrinsic factors has nonetheless been maintained. Voles and lemmings can achieve extremely high potential rates of population growth. Periods of overcrowding are therefore likely, and it would not be surprising if these generated changes in physiology, especially hormonal balance, and behaviour. Individuals may grow larger under different circumstances, or there might be increased pressure on some individuals to defend territories and on others to escape. And so on. All of these effects have been found or claimed by rodent ecologists (e.g. Lidicker, 1975; Krebs, 1978 ; Gaines *et al.*, 1979 ; Christian, 1980). However, what role they play in explaining the behaviour of rodent populations in nature remains an open question. Variations between individuals may be associated with different phases of the cycle, but this does not mean that they are *driving* the cycles. If individuals disperse more at particular cycle phases, say, or are larger, then this is likely to be a *response* to a present or past level of food or space availability, or to predation pressure. That is, intrinsic variations are more likely to explain the detailed nature of responses, whereas extrinsic factors are more likely to explain the causes of the responses.

Andreassen *et al.* (2013) brought together a range of results on intrinsic factors to see how these might combine with extrinsic drivers in generating microtine cycles in Fennoscandia. At the low- and early-increase phase, many females aggregate into social groups rather than defending separate territories, and their reproductive success is higher as a consequence. Dispersal rates are also high at this time, leading to the establishment of new social groups. Around the peak and crash phases, however, the predation of dominant males (which are especially vulnerable) from social groups reduces reproductive success in itself, but also induces an influx of immigrant males that reduce recruitment further through the infanticide they practice on the groups they join, promoting the break-up of these groups and an increased vulnerability of these disaggregated individuals to predation. Radchuk *et al.* (2016) then combined these intrinsic features of sociality and dispersal with specialist predation in an ‘individual-based model’ – outcomes for each individual in a model population were derived, and the population consequences determined by aggregating these outcomes, rather than modelling the population numbers themselves. Patterns similar to those observed are only generated by models that incorporate predation, sociality *and* dispersal (Figure 14.15). However, the models did not include generalist predators. Hence, we cannot say that these intrinsic factors improved the specialist predation model, nor that one or other of these two models is better. But the results do emphasise that while mortality and reduced reproduction may be externally imposed, the detailed patterns of their action, and hence the detailed dynamics of populations, may require the incorporation of intrinsic effects.


Full model
Predation
Predation + dispersal
Predation + sociality
ln(autumn population size)
ACF
5
4
3
2
5
10
15
20
25
30
35
0.5
0
−0.5
2
4
6
8
10
12
14

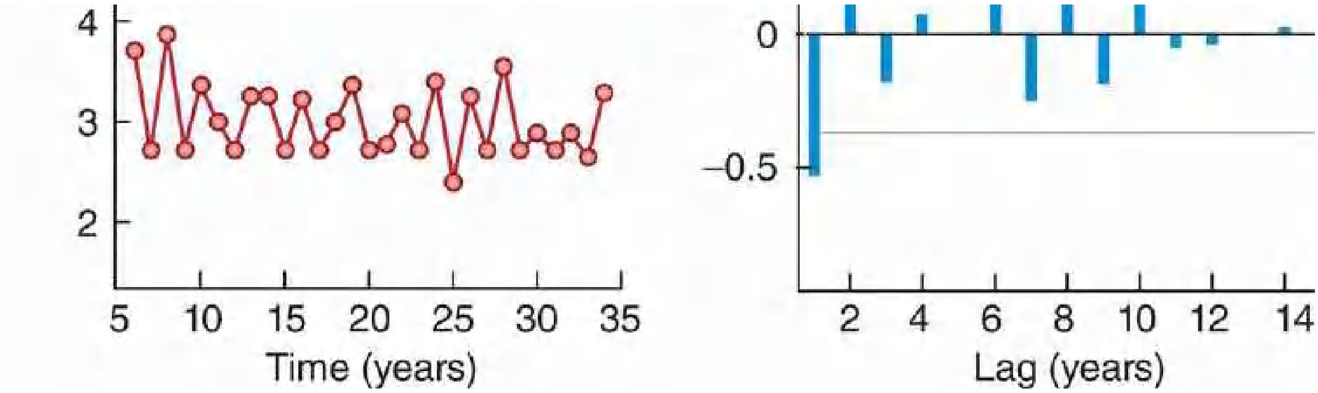


Figure 14.15 Effects of predation, dispersal and sociality on the dynamics of model microtine populations. Examples of simulation runs of various models, as indicated, of microtine rodent dynamics: a 'full' model incorporating predation, dispersal and sociality, and ones with predation only, predation and dispersal, and predation and sociality. Time series are to the left; autocorrelation functions (ACF) of those time series to the right. Only the full model recreated the type of dynamics observed in cyclic microtines.

Source: After Radchuk *et al*. (2016).

Clearly, we cannot claim to fully understand what drives the cycles of microtine rodents. Food shortage (or shortage of high-quality food) has received more support in the case of lemmings, whether from Canada or Fennoscandia, than for most other microtines. Some have emphasised, in particular, that lemmings may have a more devastating effect on their food than voles do (often mosses, rather than grasses and sedges), and that the subsequent regrowth of mosses is less immediate, introducing the required delayed density dependence (Turchin & Batzli, 2001 ; Oksanen *et al*., 2008). Along similar lines, the build-up of abrasive silica granules in heavily grazed grasses (see Section 9.3.5), and a one-year delay to return to normal levels once grazing is relaxed, has been proposed as a driver of the field vole cycles in northern England (Reynolds *et al*., 2012). However, the details of both the biology and the modelling as it applies to lemmings have been questioned (Gauthier *et al*. 2009), and we are left with the same combination of possible components – food, predation, social interactions – but insufficient evidence, especially from field experiments, to decide how best to combine them (Krebs, 2011).

The cycles of microtine rodents have been studied for longer and with greater intensity than those of any other species, and have generated more theories to explain them, and more disagreements amongst disputing advocates. There is a broad consensus that a conjunction of direct and delayed density dependence is required to account for observed patterns, but much less agreement about what the key processes are, and whether they are – or whether we should expect them to be – the same in all microtines. It is notable that even theories apparently focused on one component – for example, the specialist predation model – implicitly include, or have room within them for, food and intrinsic factors. The microtine work therefore reinforces some of the lessons learned from the red grouse and snowshoe hare studies: many factors are likely to have interacted before we reach a final outcome, and the process that causes a population's dynamics to be cyclic need not be the same one that amplifies (or dampens) that cycle. Ginzburg and Krebs (2015) argue, in particular, that there may be a widespread pattern in which the periods of cycles are defined by intrinsic factors while their amplitudes are driven by species interactions. Certainly, some at least of the examples we've discussed lend support to that view. Similar comments could be applied to the determination of abundance generally. In each case, resisting calls to declare a winning factor may seem like sitting on the fence, but these investigations are clearly marathons not sprints, and it may be, anyway, that cooperative teamwork between the factors is the best way to get us over the line.

14.7 Multiple equilibria: alternative stable states

In contrast to the regularity of population cycles, there are also populations, or sets of populations, that switch, often very rapidly, from one state to another – for example from an apparently persistent high abundance to an equally persistent low abundance. Such systems are said to exhibit 'multiple equilibria' or 'alternative stable states'. In principle, this could come about as a result of an equally profound change in the external environment, but typically there is no such association, suggesting that the change is a property of the system itself.

a predator–prey model with multiple equilibria

We can understand the general principle behind the generation of multiple equilibria within a system by returning to the graphical models of predator–prey dynamics examined in Chapter 10. In Figure 14.16, the prey zero isocline has both a vertical section at low densities and a hump. Referring back to Chapter 10, we can see that this could reflect a type 3 functional response of a predator that also has a long handling time, or perhaps the combination of an aggregative response and an Allee effect in the prey. As a consequence, the predator zero isocline crosses the prey zero isocline three times. The strengths and directions of the arrows in Figure 14.16a indicate that two of these points (X and Z) are fairly stable equilibria (although there are oscillations around each) each with their own 'basins of attraction', within which joint populations are drawn towards the equilibria. The third point (Y), however, is unstable, being on a ridge between the two basins of attraction. Populations near here will move towards either point X or point Z. Moreover, there are joint populations not far from point X where the arrows lead to the zone around point Z, and joint populations not far from point Z where the arrows lead back to the zone around point X. Even small environmental perturbations could put a population near point X on a path towards point Z, and vice versa. The behaviour of two hypothetical populations, consistent with the arrows in Figure 14.16a, are plotted in Figure 14.16b on joint abundance diagrams, and in Figure 14.16c as graphs of numbers against time. The prey population, in particular, may display an eruption in abundance from a low-density equilibrium to a high-density equilibrium and back again, or may move from one equilibrium to the other and simply stay there. In neither case would this reflect equally profound changes in the environment. Rather, this may be the result simply of environmental 'noise' close to the tipping point – one unusually warm year, for example – taking the system from one side of the point to the other, or there may be a gradual change in the environment that for a time holds the system within the same basin of attraction, with no noticeable change in behaviour, but which at some critical transition carries the system over the ridge and into the other basin of attraction.

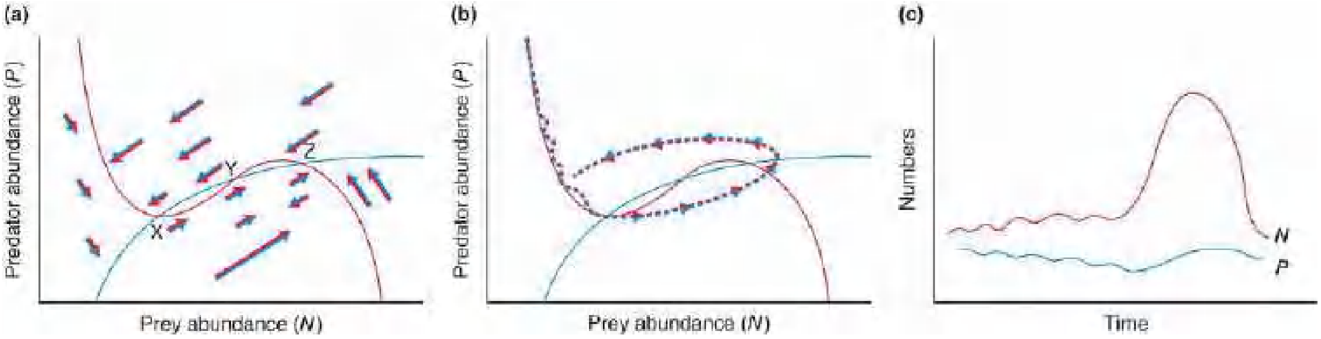


Figure 14.16 A predator–prey zero isocline model with multiple equilibria. (a) The prey zero isocline has a vertical section at low densities and a hump; the predator zero isocline can therefore cross it three times. Intersections X and Z are stable equilibria, but intersection Y is an unstable 'tipping point' from which the joint abundances move towards either intersection X or intersection Z. (b) A feasible path that the joint abundances might take when subject to the forces shown in (a). (c) The same joint abundances plotted as numbers against time, showing that an interaction with characteristics that do not change can lead to apparent 'outbreaks' in abundance.

Similar explanations have been invoked to explain comparable patterns of abundance in other systems. In each case, the underlying principles are the same. There is an unstable tipping point in the system, either side of which the system is drawn to quite different states (abundances). Only a small external perturbation, or a small additional change in the environment, is required to push the system from one side to the other, following which the system switches from one state to the other. But having thus been drawn to the new state, a much more profound change in the external environment, or a much longer time, is likely to be required in order to reverse the process. This is therefore an example of the more general phenomenon of ‘hysteresis’ – the dependence of the state of a system on its history. We will discuss an example in Section 15.3, in which harvested populations (for example fisheries) are exploited in what appears to be a sustainable fashion, but where just a small increase in exploitation rate, or unusually unfavourable weather, may be enough to push the system to the other side of a tipping point and *en route* to extinction. Two further examples are shown in Figure 14.17.

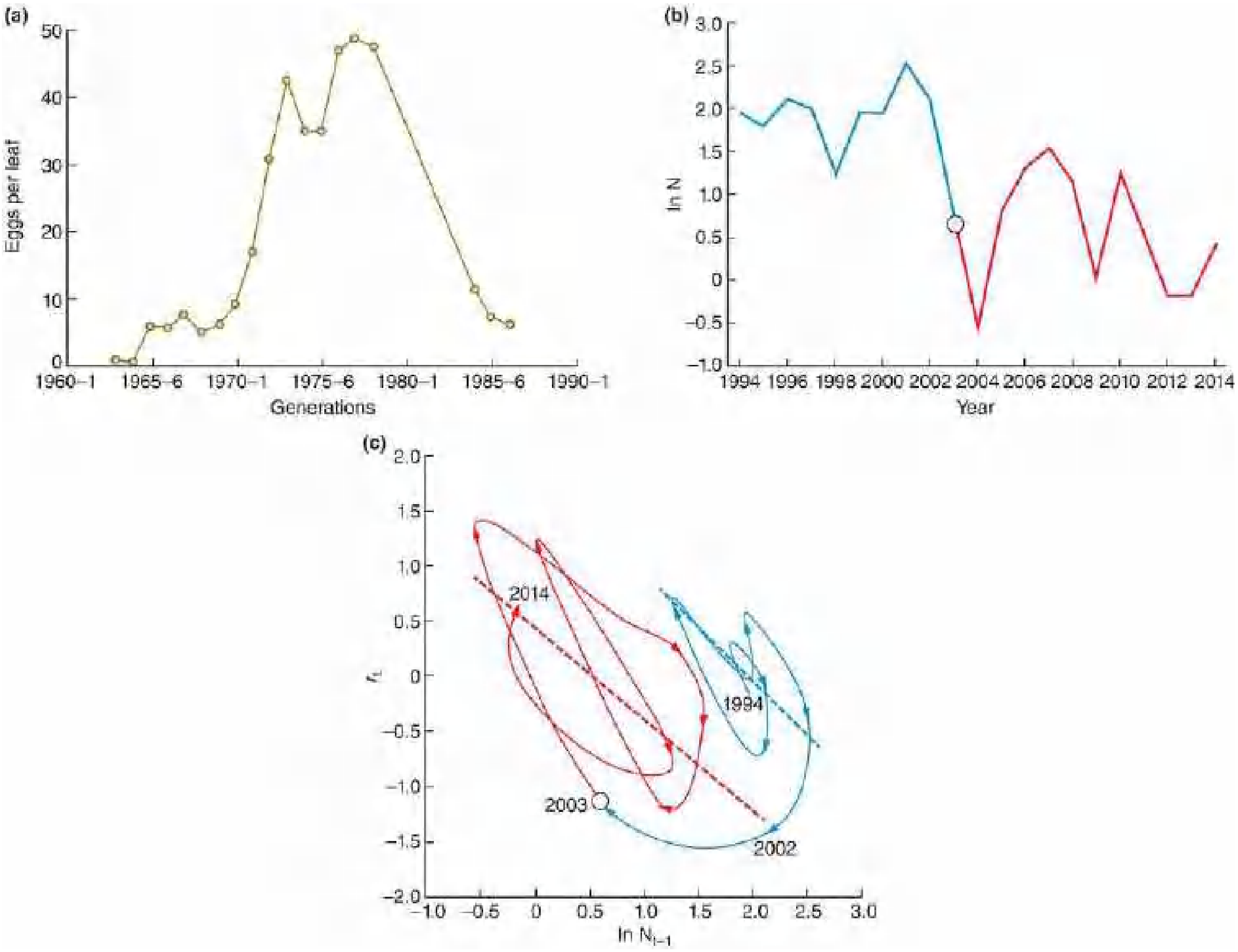


Figure 14.17 **Possible examples of outbreaks and multiple equilibria.** (a) The mean number of eggs per leaf of the viburnum whitefly, *Aleurotrachelus jelinekii*, on a viburnum bush in Silwood Park, Berkshire, UK, was initially low, then was high for a period, and then returned to initial levels. No samples were taken between 1978 and 1979, or 1984 and 1985. Parts (b) and (c) both follow the population dynamics of the midday gerbil, *Meriones meridianus*, in southern Russia as log-abundance (numbers capture per 100 trap-nights). Part (b) shows an apparent switch from high to low stable levels around 2003, and (c) supports this by plotting the population growth rate against log-abundance and showing that there were separate density-dependent relationships (the dashed straight lines) between the two before and after the switch ($r^2 = 0.48$, $P = 0.056$ and $r^2 = 0.50$, $P < 0.01$, respectively) but around different mean abundances.

Source: (a) After Southwood *et al.* (1989). (b, c) After Tchabovsky *et al.* (2016).

Alternative stable states have also been proposed for a number of plant–herbivore interactions, often where increased grazing pressure leads to the 'collapse' of the vegetation from a high biomass to a much lower one, which then fails to return to the high biomass state even when grazing pressure is severely reduced (van de Koppel *et al.*, 1997). The grasslands of the Sahel region of Africa, grazed by livestock, and the arctic plants along the coast of Hudson Bay in Canada, grazed by geese, are both examples. The conventional explanation (Noy-Meir, 1975) is essentially that depicted in Figure 14.16: an Allee effect in which plants driven to a low biomass have very little material above ground and hence very limited powers of immediate regrowth. It may also be, however, that the plants' problems at low biomass are compounded by soil deterioration – erosion, for example – introducing further positive feedback into the system: high grazing leading to low plant biomass, leading to poorer growing conditions, leading to lower plant biomass, leading to even poorer growing conditions, and so on (van de Koppel *et al.*, 1997). Such positive feedbacks are likely to be important more generally in holding systems at their new state once they have arrived there.

Indeed, similar patterns of sudden transition have been observed across whole communities, and for that matter in human societies, financial systems and so on. In these more complex systems, there are likely to be additional features, notably the heterogeneity amongst the components of the system, and the connectedness between them, that influence the propensity of a community or ecosystem to undergo these sudden transitions (Scheffer *et al.*, 2012). We therefore pick up this topic again, at the community level, in Chapter 17.

Chapter 15
Pest Control, Harvesting and Conservation

15.1 Managing abundance

Humans are very much a part of all ecosystems. We manipulate the abundances of other species in many ways. Sometimes our activities motivate us to suppress species we identify as pests. At other times, we aim to prevent the extinction of species we believe to be endangered, or we kill individuals for food. The desired outcomes are very different for pest controllers, harvest managers and conservation ecologists. But all need management strategies based on an understanding of what determines and regulates abundance. All, too, aim to develop strategies that are sustainable.

sustainability – an aim shared by pest controllers, harvest managers and conservation biologists

To call an activity 'sustainable' means that it can be continued or repeated for the foreseeable future. Concern has arisen, therefore, precisely because so much human activity is clearly unsustainable. We cannot continue to use the same pesticides if increasing numbers of pests become resistant to them. We cannot (if we wish to have fish to eat in future) continue to remove fish from the sea faster than the remaining fish can replace their lost companions. And we cannot claim to have conserved a species unless its future, at least for a time, looks assured – the very essence of sustainability

Sustainability has thus become one of the core concepts – perhaps the core concept – in an ever-broadening concern for the fate of the earth and the ecological communities that occupy it. In defining sustainability we used the words 'foreseeable future'. This is because, when an activity is described as sustainable, it is on the basis of what is known at the time. But many factors remain unknown or unpredictable. Things may take a turn for the worse (as when adverse oceanographic conditions damage a fishery already threatened by overexploitation) or some unforeseen additional problem may be discovered (resistance may appear to some previously potent pesticide). On the other hand, technological advances may allow an activity to be sustained that previously seemed unsustainable (new types of pesticide may be discovered that are more finely targeted on the pest itself rather than innocent bystander species). However, there is a real danger that we observe the many scientific advances that have been made in the past and assume that there will always be a technological 'fix' to solve our present problems, too. Unsustainable practices cannot be accepted simply on the basis of faith that future advances will make them sustainable after all.

The recognition of the importance of sustainability as a unifying idea in applied ecology has grown gradually, but there is something to be said for the claim that sustainability really came of age in 1991. This was when the Ecological Society of America published ‘The sustainable biosphere initiative: an ecological research agenda’, a ‘call-to-arms for all ecologists’ with a list of 16 co-authors (Lubchenco *et al.*, 1991). In the same year, the World Conservation Union (IUCN), the United Nations Environment Program and the World Wide Fund for Nature jointly published Caring for the Earth: A Strategy for Sustainable Living (IUCN/UNEP/WWF, 1991). The detailed contents of these documents are less important than their existence. They indicate a growing preoccupation with sustainability, shared by scientists, pressure groups and governments, and recognition that much of what we do is not sustainable. More recently, the emphasis has shifted from a purely ecological perspective to one that adds in the social and economic conditions influencing sustainability. This is sometimes referred to as the ‘triple bottom line’ of sustainability – a term originally coined by Elkington (see Elkington, 2002).

Note again that this chapter contains no ‘Application’ sections. This is because the entire focus of the chapter is on strategies and protocols to solve environmental problems.

15.2 The management of pests

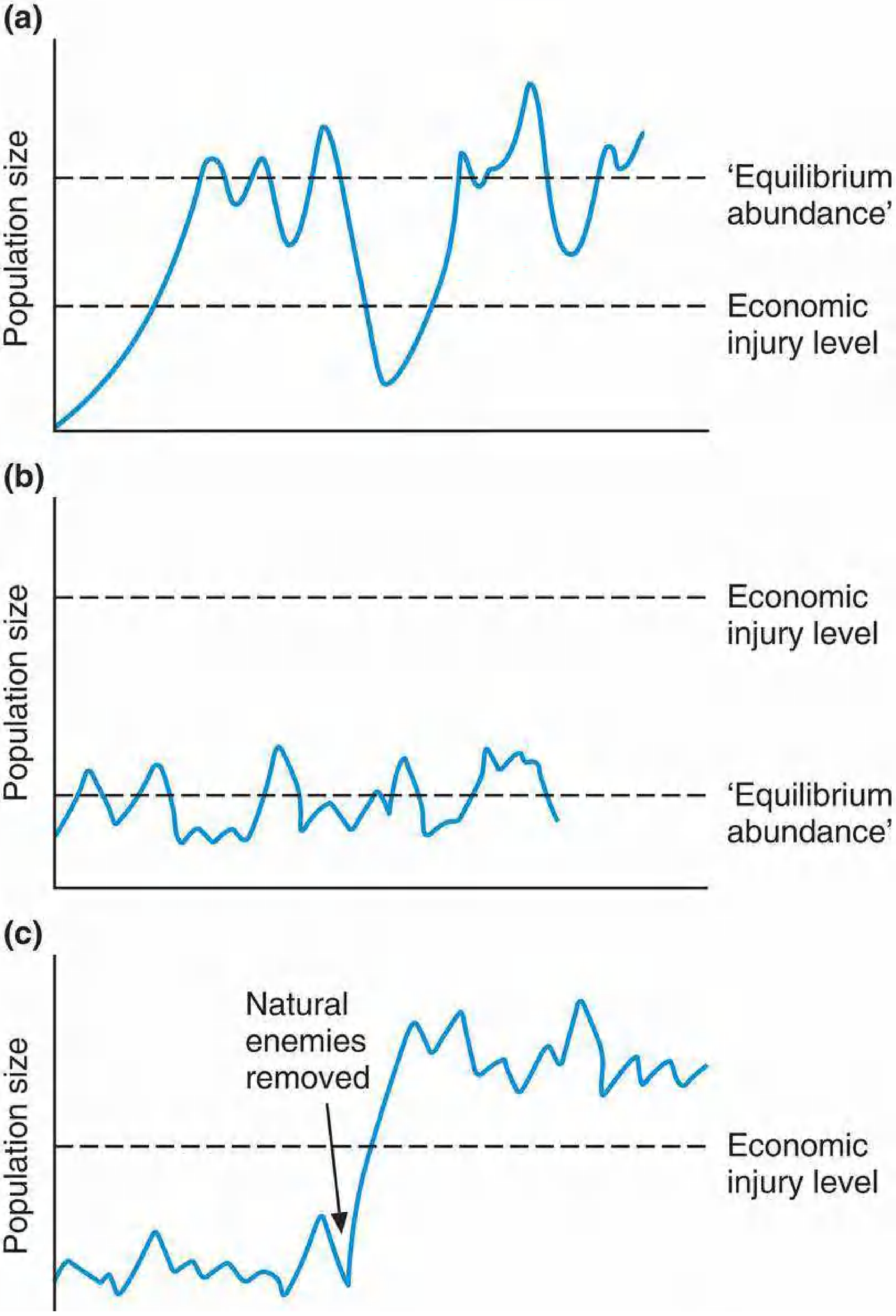
what is a pest?

A pest species is one that humans consider undesirable. This definition covers a multitude of sinners: mosquitoes are pests because they carry diseases or because their bites itch; *Allium* spp. are pests because when harvested with wheat these weeds make bread taste of onions; rats and mice are pests because they feast on stored food; mustellids are pests in New Zealand because they are unwanted invaders that prey upon native birds and insects; garden weeds are pests for aesthetic reasons. People want to be rid of them all.

15.2.1 Economic injury levels and economic thresholds

economic injury levels define actual and potential pests

One might imagine that the aim of pest control is always total eradication of the pest, but this is not the general rule. Rather, the aim is to reduce the pest population to a level at which it does not pay to achieve yet more control. This is referred to as the *economic injury level* or EIL. The EIL for a hypothetical pest is illustrated in Figure 15.1a: it is greater than zero (eradication is not profitable) but it is also below the typical, average abundance of the species. If the species was naturally self-limited to a density below the EIL, then it would never make economic sense to apply ‘control’ measures, and the species could not, by definition, be considered a ‘pest’ (Figure 15.1b). There are other species that have a carrying capacity in excess of their EIL, but have a typical abundance that is kept below the EIL by natural enemies (Figure 15.1c). These are potential pests. They can become actual pests if their enemies are removed. In practice, since the damage that pests do is contingent on the state of the crop itself and many other factors, EILs typically include details not only of the number of pests, but also of the crop itself, perhaps the pest’s natural enemies, and so on.


(a)
Population size
'Equilibrium abundance'
Economic injury level
(b)
Population size
Economic injury level
'Equilibrium abundance'
(c)
Population size
Natural enemies removed
Economic injury level

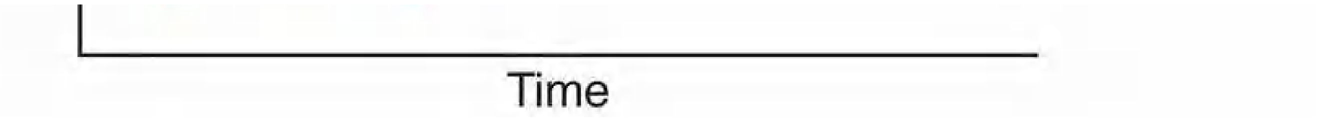


Figure 15.1 Uncontrolled pests typically exceed their economic injury level: non-pests do not. (a) The population fluctuations of a hypothetical pest. Abundance fluctuates around an ‘equilibrium abundance’ set by the pest’s interactions with its food, predators, etc. It makes economic sense to control the pest when its abundance exceeds the economic injury level (EIL). Being a pest, its abundance exceeds the EIL most of the time (assuming it is not being controlled). (b) By contrast, a species that cannot be considered a pest fluctuates always below its EIL. (c) ‘Potential’ pests fluctuate normally below their EIL but rise above it in the absence of one or more of their natural enemies.

the economic threshold – getting ahead of the pests

However, when a pest population has reached a density at which it is causing economic injury, it is generally too late to start controlling it. Hence it can be more important to know the *economic threshold* (ET): the density of the pest at which action should be taken to prevent it reaching the EIL. An example of how cost of treatment, value of crop, and damage by pests come together to determine EILs and ETs, is provided by a study on alfalfa (*Medicago sativa*), an important forage crop fed to dairy cattle in the USA, of which the potato leafhopper, *Empoasca fabae*, is a major pest (Chasen *et al*., 2015). The EIL brings these various factors together as follows:

$$\mathrm{EIL} = C/VDP, \tag{15.1}$$

where C is the cost of control per unit area (the EIL is greater when control is costly), V is the value of the crop per unit of yield (the EIL is lower when the crop is valuable), D is the yield loss caused by each pest (the EIL is lower when individual pests are particularly damaging) and P is the effectiveness of the control measures (the EIL is lower when control is effective). In this case, the ET for potato leafhopper infestations on alfalfa is taken to be 75% of the EIL (control should be exerted when pest levels reach 75% of the EIL). This is a pragmatic rather than a precisely estimated figure, as is frequently the case. It encourages pre-emptive action by crop managers, but has not been refined by detailed studies of the abundance of planthoppers as they progress from threatening to damaging levels. The alfalfa EIL had not been altered for 30 years. Thus, acknowledgement of the increasing value of alfalfa (V in Equation 15.1) and a reassessment of rates of yield loss per insect (D) led to a revision downwards in the ET from 0.2–2.0 planthoppers per sampling-net sweep (depending on the height of the alfalfa crop at the time) to 0.1–0.8 planthoppers, though these revisions would themselves be sensitive to variations in the cost of insecticides (Chasen *et al*., 2015). Alfalfa is now more valuable: pest controllers therefore need to step in earlier to protect it.

15.2.2 Chemical pesticides and their unintended consequences

Chemical pesticides are a key part of the armoury of pest managers, but as we shall see, there are several undesirable consequences of their use. Even in organic agriculture, where synthetic chemicals are banned, naturally occurring chemicals (plant products, for example) are allowed.

The history of the use of chemical pesticides is a journey that began with products, often naturally occurring, that are toxic to virtually all organisms, and that can be used as pesticides only by directing them at the pests and adjusting dosages so as to (hopefully) harm the pests but not other species, including the pest controllers. These dominated pest control from the time of the ancient civilisations to the era of increasingly scientific pest management in the 19th and early 20th century. However, in the middle of the 20th century, these pesticides were largely superseded by synthetic organic chemicals – chlorinated hydrocarbons (such as DDT), organophosphates and carbamates, for example, in the case of insecticides – that were more effective than their simpler antecedents but often not much more selective in their action. More recent decades, then, have seen a drive towards increased selectivity (targeting of pests) but inevitably increased costs, too. Among insecticides, for example, less selective organics have increasingly been replaced by synthetic pyrethroids. Their chemical structure is based on that of pyrethrum, a chemical produced naturally, as an insecticide, by chrysanthemums – one of a number of naturally occurring botanical insecticides. The synthetic pyrethroids are far more effective, but their increased refinement has meant increased complexity and increased costs. A fuller account of the use of chemical insecticides is provided by Yu (2014), of herbicides by Cobb and Reade (2010) and of fungicides by Oliver and Hewitt (2014).

target pest resurgence

A pesticide gets a bad name if, as is often the case, it kills more species than the one it was aimed at. However, in the context of sustainability, the bad name is especially justified if it kills the pests' natural enemies and so contributes to undoing what it was employed to do. '*Target pest resurgence*' occurs when the application of a pesticide kills large numbers of the pest but also large numbers of its natural enemies, such that the numbers of the pest increase rapidly some time after the initial application, because pest individuals that survive the pesticide, or that migrate into the area later, find themselves with a plentiful food resource but few, if any, natural enemies. An example is shown in Figure 15.2a, where numbers of the plant bug, *Lygus hesperus*, a pest of cotton and other crops, were higher in plots treated with the broad-spectrum insecticide, acephate, than on untreated control plots (Asiimwe *et al.*, 2014). There were also significant negative effects of acephate on the abundances of a wide range of natural enemies of the bug. Neither of these effects were detected in plots where the selective insecticide flonicamid was used, which does negligible harm to the natural enemies. All of this supports the pivotal role of effects on natural enemies in target pest resurgence, as do many other examples. On the other hand, there is mounting evidence that pesticides may exhibit *hormesis*, a reversal in the dose-response relationship at low concentrations such that low concentrations, perhaps left as a residuum, have a positive not a negative or zero effect on the target species (Guedes & Cutler, 2014). This, too, can undermine or even reverse the detrimental effects that pesticides have on pests.

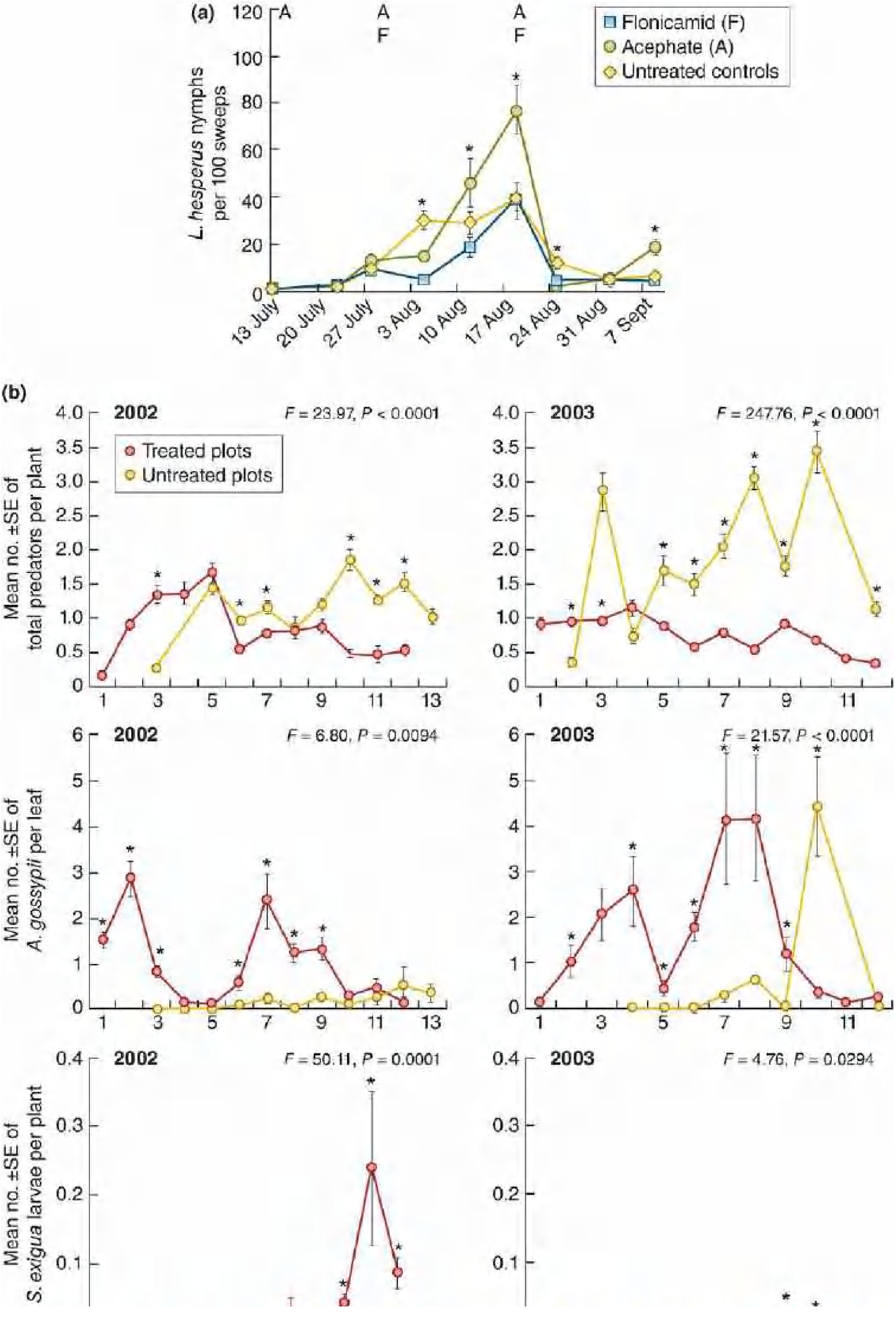

(a)
A
A F
A F
Flonicamid (F)
Acephate (A)
Untreated controls
L. hesperus nymphs per 100 sweeps
0
20
40
60
80
100
120
13 July
20 July
27 July
3 Aug
10 Aug
17 Aug
24 Aug
31 Aug
7 Sept
(b)
2002
F = 23.97, P < 0.0001
Treated plots
Untreated plots
Mean no. ±SE of total predators per plant
2003
F = 247.76, P < 0.0001
2002
F = 6.80, P = 0.0094
Mean no. ±SE of A. gossypii per leaf
2003
F = 21.57, P < 0.0001
2002
F = 50.11, P = 0.0001
Mean no. ±SE of S. exigua larvae per plant
2003
F = 4.76, P = 0.0294
1
3
5
7
9
11
13

Figure 15.2 Pesticides can lead to target pest resurgence and secondary pest outbreaks. (a) The effect, compared with untreated controls, of the selective insecticide flonicamid and the broad-spectrum insecticide acephate on the abundance of nymphs of the plant bug, *Lygus hesperus*, on cotton in Arizona, USA. A and F indicate when treatments were made. Bars are SEs; * indicates a significant difference ($P < 0.05$) at the date concerned. (b) The numbers of predators (top) and of the secondary pests cotton aphid, *Aphis gossypii* (middle), and beet armyworm, *Spodoptera exigua* (bottom), in two years (left and right) when cotton plants were treated to control the boll weevil, *Anthonomus grandis grandis*, in Texas, USA. Statistics to the upper right relate to the whole period; * indicates a significant difference ($P < 0.05$) at the date concerned.

Source: (a) After Asiimwe *et al*. (2014). (b) After Knutson *et al*. (2011).

secondary pest outbreaks

Indeed, when a pesticide is applied, it may not only be the target pest that resurges. Alongside the target are likely to be a number of potential pest species that had been kept in check by their natural enemies (see Figure 15.1c). If the pesticide destroys these (or there is hormesis), the potential pests become real ones – and are called *secondary pests*. We see an example in Figure 15.2b, where application of the broad-spectrum insecticide malathion to control the boll weevil, *Anthonomus grandis grandis*, on cotton in Texas, USA gave rise to subsequent outbreaks of two other pests, the beet armyworm, *Spodoptera exigua*, and the cotton aphid, *Aphis gossypii*.

mortality of non-target species in general

Sometimes the unintended effects of pesticide application have been much less subtle than target pest resurgence or secondary pest outbreaks. Problems have been especially apparent for the many classes of pesticides that tend to be retained by organisms by the process of *biomagnification*: an increase in the concentration of the pesticide in body tissues as we move up a food chain as a result of successive predators consuming and then concentrating the pesticides carried by their prey. The potential for disaster is illustrated, for example, by estimates that around 75% of surface waters contain levels of neonicotinoids that are harmful to the invertebrates in those waters (Morrissey *et al*., 2015). Neonicotinoids are the fastest growing class of insecticides applied to crop production. It is sobering, too, that landscape-scale usage of fungicides applied for crop protection in the USA is the best predictor of range contraction of declining bumblebee species (key pollinators), and of the increasing prevalence of the bee pathogen, *Nosema bombi*, in declining species (McArt *et al*., 2017). These unintended, harmful effects to bystander and even positively beneficial species argue in favour of a precautionary approach in any pest management exercise. They argue, too, for improved understanding of the toxicity and persistence of pesticides, and the development of more specific and less persistent alternatives.

unintended effects of the genetic modification of crops

There can be unintended consequences, too, when crops are genetically modified to allow increased usage of pesticides. For example, a number of crops have been produced that are tolerant to the non-selective herbicide glyphosate. This allows the herbicide to be used to effectively control weeds without adverse effects on the crop itself, which, of course, further encourages its use. At commercial rates, herbicides appear to have few significant direct effects on animals. However, conservationists increasingly worry about the indirect effects when the weed plants are the food hosts for insects or their seeds are a key food item for birds.

Fat hen (*Chenopodium album*) is a plant that occurs worldwide and is one weed that we can expect to be affected adversely by the farming of genetically modified (GM) crops. But the seeds of fat hen are an important winter food source for farmland birds, including the skylark (*Alauda arvensis*). Watkinson *et al.* (2000) took advantage of the fact that the population dynamics of both fat hen and skylarks have been intensively studied. They incorporated both into a model of the impacts of GM sugar beet on skylark populations that, despite its simplicity, conveys some key messages. Skylarks forage preferentially in weedy fields and aggregate locally in response to weed seed abundance. Watkinson *et al.*'s model, therefore, included two key features: (i) before the introduction of GM technology the seeds were aggregated – most farms had a relatively low density of weed seeds, with a few farms having very high densities; and (ii) the probability of a farmer adopting GM sugar beet is related to weed seed density through a parameter ρ. Positive values of ρ mean that farmers are more likely to adopt the technology where seed densities are currently high (and the perceived need is greatest). This leads to an increase in the frequency of low-density fields (Figure 15.3a). Negative values of ρ, however, indicate that farmers are more likely to adopt GM sugar beet where seed densities are currently low, perhaps because a history of effective weed control is correlated with a willingness to adopt new technology. This leads to a decreased frequency of low-density fields (Figure 15.3a). ρ, therefore, is not an ecological parameter but reflects a socioeconomic response to the introduction of new technology, and it turns out that this relationship is more important to bird population density in the model than the direct impact on weed abundance (Figure 15.3b). Specifically, the threat to skylark populations from GM sugar beet will be greatest if it is planted, preferentially, by those with the greatest weed problems. This reminds us of the need for resource managers to think in terms of the triple bottom line of sustainability, with its ecological, social and economic dimensions.

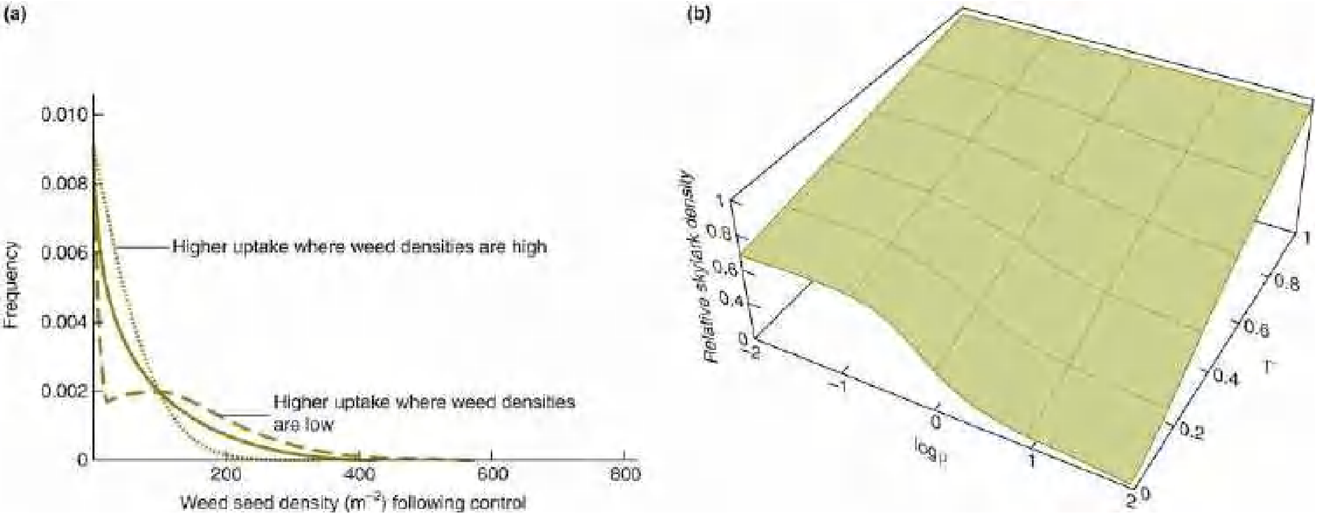


Figure 15.3 The threat to skylark populations from GM sugar beet is greatest if it is planted, preferentially, by farmers with the greatest weed problems. (a) Frequency distributions of mean seed densities across farms before the introduction of GM sugar beet (solid line), and in two situations where the technology has been adopted: where the technology is preferentially adopted on farms where weed density is currently high (ρ positive, dotted line) and where it is currently low (ρ negative, dashed line). (b) The relative density of skylarks in fields in winter (vertical axis; unity indicates field use before the introduction of GM crops) in relation to ρ and to the approximate reduction in weed seed bank density due to the introduction of GM crops (Γ). Realistic values of Γ are those less than 0.1. Hence, small positive or negative values of ρ can be expected to give quite different skylark densities.

Source: After Watkinson *et al.* (2000).

15.2.3 Evolution of resistance to pesticides

evolved resistance: a widespread problem

Chemical pesticides lose their role in sustainable agriculture if the pests evolve resistance. Such evolution is simply natural selection in action and is an almost inevitable consequence when vast numbers of individuals in a genetically variable population are killed in a systematic way by a pesticide. Initially, one or a few individuals may be unusually resistant (perhaps because they possess an enzyme that can detoxify the pesticide), but if the pesticide is applied repeatedly, each successive generation will contain a larger proportion of resistant individuals. Pests typically have a high intrinsic rate of reproduction. A few individuals in one generation may give rise to hundreds or thousands in the next.

This problem was often ignored in the past, even though the first case of DDT resistance was reported shortly after its introduction in 1946 (in house-flies, *Musca domestica*, in Sweden). The scale of the problem is illustrated in Figure 15.4 for insecticides and herbicides.

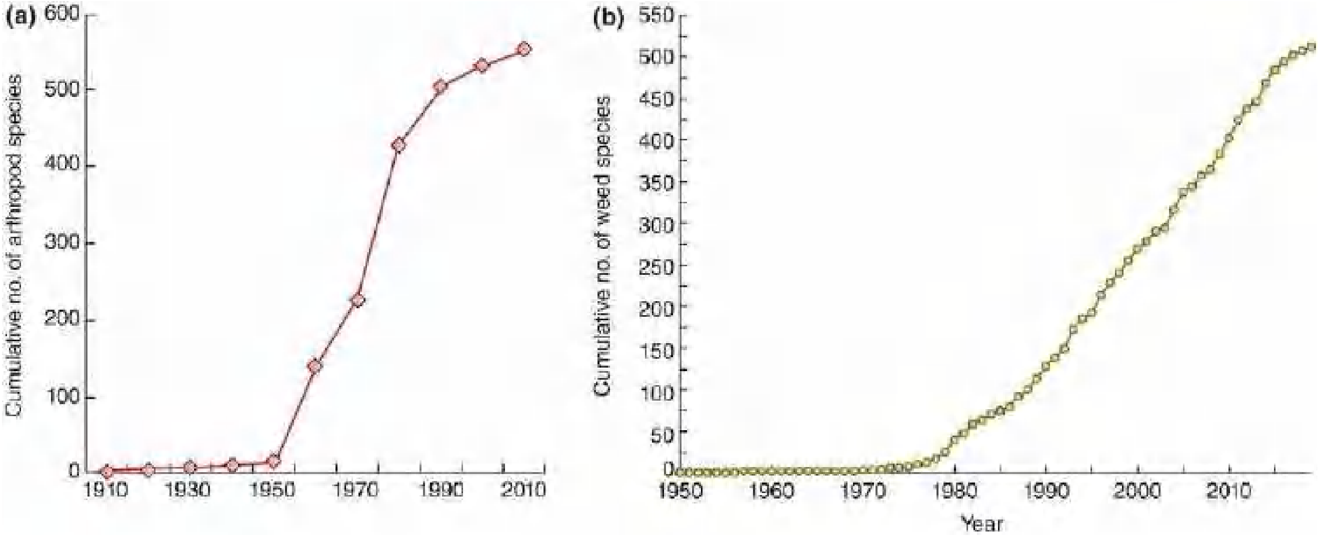


Figure 15.4 There has been a steady rise, for more than half a century, in the number of pests resistant to chemical pesticides. The increase in the number of (a) arthropod (insect and mite) species resistant to insecticides and (b) weed species resistant to herbicides.

Source: (a) After Sparks (2013).

managing resistance

The evolution of pesticide resistance can be slowed, though, by various resistance management strategies. The strategy developed by the World Health Organization for the insect vectors of malaria is a good example (WHO, 2012). Apart from only using them when necessary (see 15.2.5 Integrated Pest Management), it proposes using insecticides with different modes of action in rotation, or applying them in a spatial mosaic, or combining them in mixtures. In each case, the aim is to deny the insects repeated exposure to the same insecticide, which would be the ideal conditions for resistance to evolve.

However, if chemical pesticides brought nothing but problems – if their use was intrinsically and acutely unsustainable – then they would already have fallen out of widespread use. This has not happened. Instead, their rate of production has increased rapidly. The ratio of benefit to cost for the individual producer has often remained in favour of pesticide use, at least in the short term. A meta-analysis of 115 studies comparing conventional and organic farming (systems with and without artificial pesticides) found that yields were on average 19% higher in conventional systems (Ponisio *et al*., 2014). Moreover, in many poorer countries, the prospects of imminent mass starvation, or of an epidemic disease, are so frightening that the social and health costs of using pesticides have to be ignored. In cases such as these, the use of pesticides is justified by objective measures such as 'lives saved' or 'total food produced', and in this sense at least, their use may be described as sustainable. In practice, even if pesticide use is reduced, sustainability will also depend on continually developing new pesticides that keep at least one step ahead of the pests: pesticides that are less persistent and more accurately targeted at the pests themselves.

15.2.4 Biological control

Outbreaks of pests occur repeatedly and so does the need to apply pesticides. But if we can apply *biological control* (the manipulation of the natural enemies of pests), then it may be possible to eliminate or reduce the use of chemicals but control pests just as effectively, often at a much reduced cost, both economically and environmentally (Table 15.1). We have seen examples of biological control in previous chapters (for example, Applications 4.4, 9.1, 10.1, 10.3 and 12.3) but here we take a closer look.

Table 15.1 Comparison, based on various literature sources, between aspects of the development of products for chemical and biological control of pests. *Source:* After van Lenteren (2012).

	Chemical control	**Biological control**
Number of 'ingredients' tested	>3.5 million	3500
Success ratio	1:140 000	1:10
Development costs per product	US$256 million	US$2 million
Development time	10 years	10 years
Benefit/cost ratio	2:1	2.5–20:1
Risks of resistance	High	Nil/low
Specificity	Low	High
Harmful side effects	Many	Nil/few

conservation

There are a variety of categories of biological control. The most straightforward, perhaps, is *conservation* biological control, which simply involves sustaining the density or persistence of populations of generalist natural enemies of a pest to support some baseline level of control. For example, aphid pests of wheat are attacked by coccinellid and other beetles, heteropteran bugs, lacewings (Chrysopidae), syrphid fly larvae and spiders (Brewer & Elliott, 2004). Many of these natural enemies overwinter in the grassy boundaries at the edge of wheat fields, from where they disperse and reduce aphid populations around the field edges. The planting of grassy strips within the fields can enhance these natural populations and the scale of their impact on aphid pests.

cottony cushion scale insect: a classic case of importation ...

The most common type of biological control is the *importation* of a natural enemy from another geographic area – very often the area in which the pest originated prior to achieving pest status – in order that the control agent should persist and thus maintain the pest, long term, below its economic threshold. This is often also called *classical* biological control and sometimes *inoculation*.

One of the best examples of 'classical' biological control is itself a classic. Its success marked the start of biological control in a modern sense. The cottony cushion scale insect, *Icerya purchasi*, was first discovered as a pest of Californian citrus orchards in 1868. By 1886 it had brought the citrus industry close to the point of destruction. Ecologists initiated a worldwide correspondence to try to discover the natural home and natural enemies of the pest, eventually leading to the importation to California of about 12 000 individuals of a dipteran parasitoid, *Cryptochaetum iceryae*, from Australia, and 500 predatory ladybird beetles (*Rodolia cardinalis*) from Australia and New Zealand. Initially, the parasitoids seemed simply to have disappeared, but the predatory beetles underwent such a population explosion that all infestations of the scale insects in California were controlled by the end of 1890. Although these beetles have usually taken most or all of the credit, the long-term outcome has been that the beetles are instrumental in keeping the scale in check inland, but *Cryptochaetum* is the main agent of control on the coast (Flint & van den Bosch, 1981).

... illustrating several general points

This example illustrates a number of important general points. Firstly, it emphasises that species may become pests simply because, by colonisation of a new area, they escape the control of their natural enemies (the enemy release hypothesis). Biological control by importation is thus, in an important sense, restoration of the status quo for the specific predator–prey interaction (although the overall ecological context is certain to differ from what would have been the case where the pest and control agent originated). Then, we see that biological control requires the classical skills of the taxonomist to find the pest in its native habitat, and particularly to identify and isolate its natural enemies. This may often be a difficult task – especially if the natural enemy has the desired effect of keeping the target species at a low carrying capacity, since both the target and the agent will then be rare in their natural habitat. Nevertheless, the rate of return on investment can be highly favourable. In the case of the cottony cushion scale, biological control has subsequently been transferred to 50 other countries and savings have been immense. In addition, this example illustrates the importance of establishing several, hopefully complementary, enemies to control a pest. Finally, we should note that classical biological control, like natural control, can be destabilised by chemicals. The first use of DDT in Californian citrus orchards in 1946–1947 against the citricola scale insect *Coccus pseudomagnoliarum* led to an outbreak of the (by then) rarely seen cottony cushion scale when the DDT almost eliminated the ladybirds. The use of DDT was terminated.

augmentation

Augmentation differs from importation in requiring the mass production of control agents prior to their release. It is therefore often used where the agent is unable to persist long term. Glasshouses are probably the best example, where crops are removed, along with the pests and their natural enemies, at the end of the growing season. Augmentation therefore typically makes use of local, not exotic, imported agents. As of 2012, 230, mostly arthropod natural enemy species were available commercially for augmentation control, but for its adherents at least, there remains a frustratingly low rate of uptake, despite the favourable combination of economic success and environmental harmlessness that it provides (van Lenteren, 2012).

microbial control of insects via inundation – and genetic manipulation

Finally, *inundation* is the release of large numbers of a natural enemy, with the aim of killing those pests present at the time, but with no expectation of providing long-term control as a result of the control agent's population increasing or maintaining itself. By analogy with the use of chemicals, agents used in this way are referred to as *biological pesticides*. One common form of inundation involves the use of insect pathogens to control insect pests (Bailey *et al*., 2010). By far the most widespread and important agent is the bacterium *Bacillus thuringiensis* ('Bt'), which can easily be produced on artificial media. Unlike most other microbial agents, the pesticide in this case comprises one or more of the toxins produced by the bacterium rather than the living microbe itself. Its advantages include its powerful toxicity against specific target insects – after being ingested by the larvae, death typically occurs 30 minutes to three days later – and its lack of toxicity against organisms outside this narrow group (including ourselves and most of the pest's natural enemies). It has been used against insect vectors of disease as well as pests of crops and stored products. Perhaps most significantly, though, several major crop species have been genetically modified to express Bt toxins, and the use of these crops has grown massively (Figure 15.5). During 2012, Bt corn (*Zea mays*) accounted for 67% of the corn planted in the USA. And between 2010 and 2012, Bt cotton (*Gossypium* spp.) accounted for 75–97% of cotton planted in Australia, India, China and the USA (Tabashnik *et al*., 2013). Unsurprisingly, resistance to the Bt toxins in major pest species has followed closely in the wake of this increased usage (Figure 15.5). This in turn has focused attention on strategies of resistance management, including the use from the outset of strains expressing more than one of the Bt toxins, and especially the planting of sufficient 'refuge' plots containing plants without Bt toxins, promoting the survival of Bt-susceptible pests with which rare Bt-resistant individuals are likely to mate, delaying the evolution of resistance overall (Tabashnik *et al*., 2013).

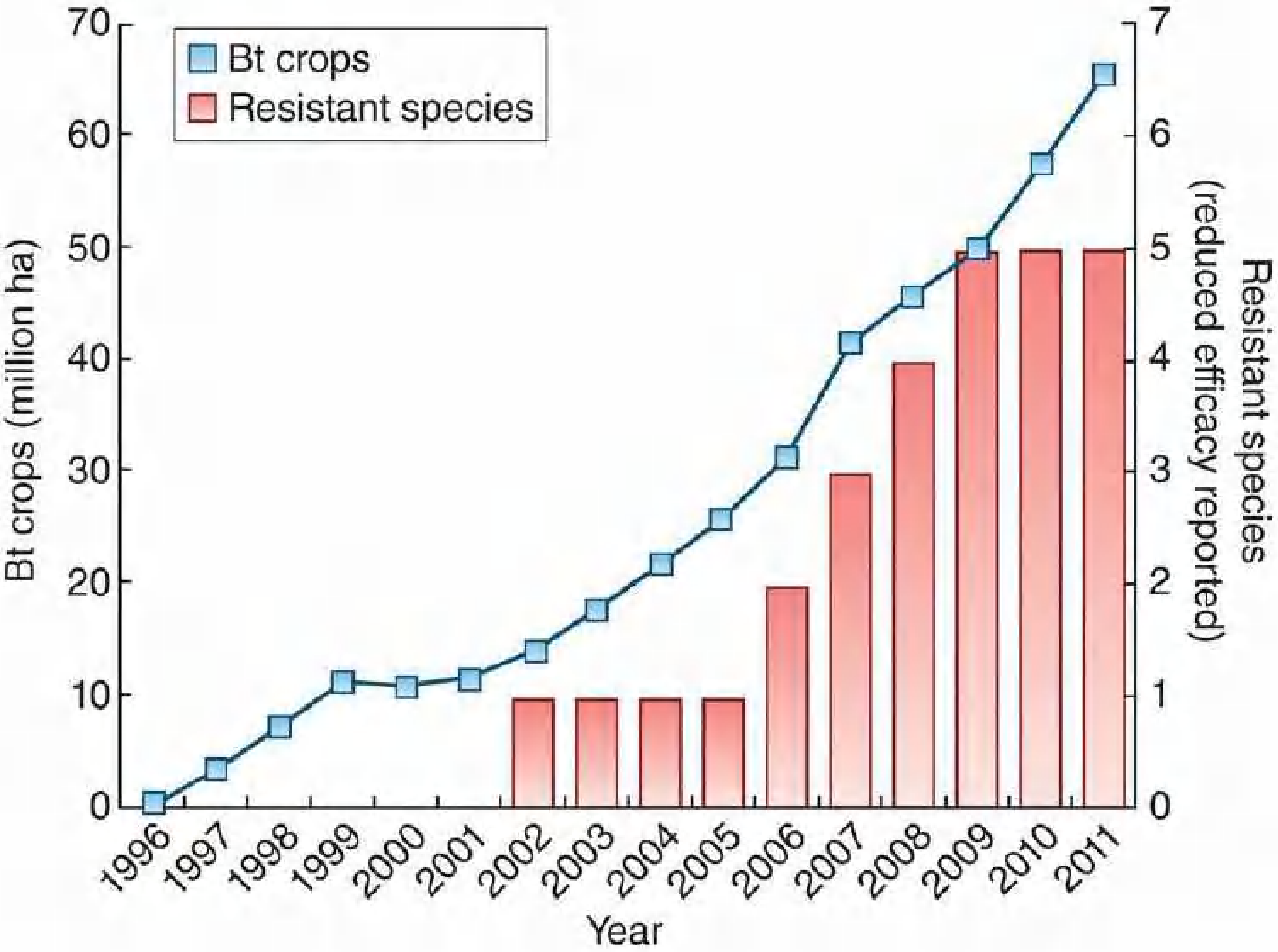


Figure 15.5 ***Bacillus thuringiensis*** **(Bt) use and resistance to it have both risen sharply.** The increases in the use of crops engineered to produce toxins of Bt for the control of insect pests, and of pest species reported to be resistant to the toxins.

Source: After Tabashnik *et al*. (2013).

a need for caution

There are many examples where biological control, once developed, has been successful (Gurr & Wratten, 2000). But we should not pretend that things never go wrong. For example, a seed-feeding weevil (*Rhinocyllus conicus*), introduced to North America to control exotic *Carduus* thistles, attacks more than 30% of native thistles (of which there are more than 90 species), reducing thistle densities (by 90% in the case of the Platte thistle *Cirsium canescens*) with consequent adverse impacts on the populations of a native picture-winged fly (*Paracantha culta*) that feeds on thistle seeds (Louda *et al*., 2003a). Louda *et al*. (2003b) reviewed 10 biological control projects that included the unusual but worthwhile step of monitoring non-target effects and concluded that relatives of the target species were most likely to be attacked whilst rare native species were particularly susceptible. Their recommendations for management included the avoidance of generalist control agents, an expansion of host-specificity testing and the need to incorporate more ecological information when evaluating potential biological control agents.

15.2.5 Integrated pest management

IPM: an ecologically rather than chemically based philosophy

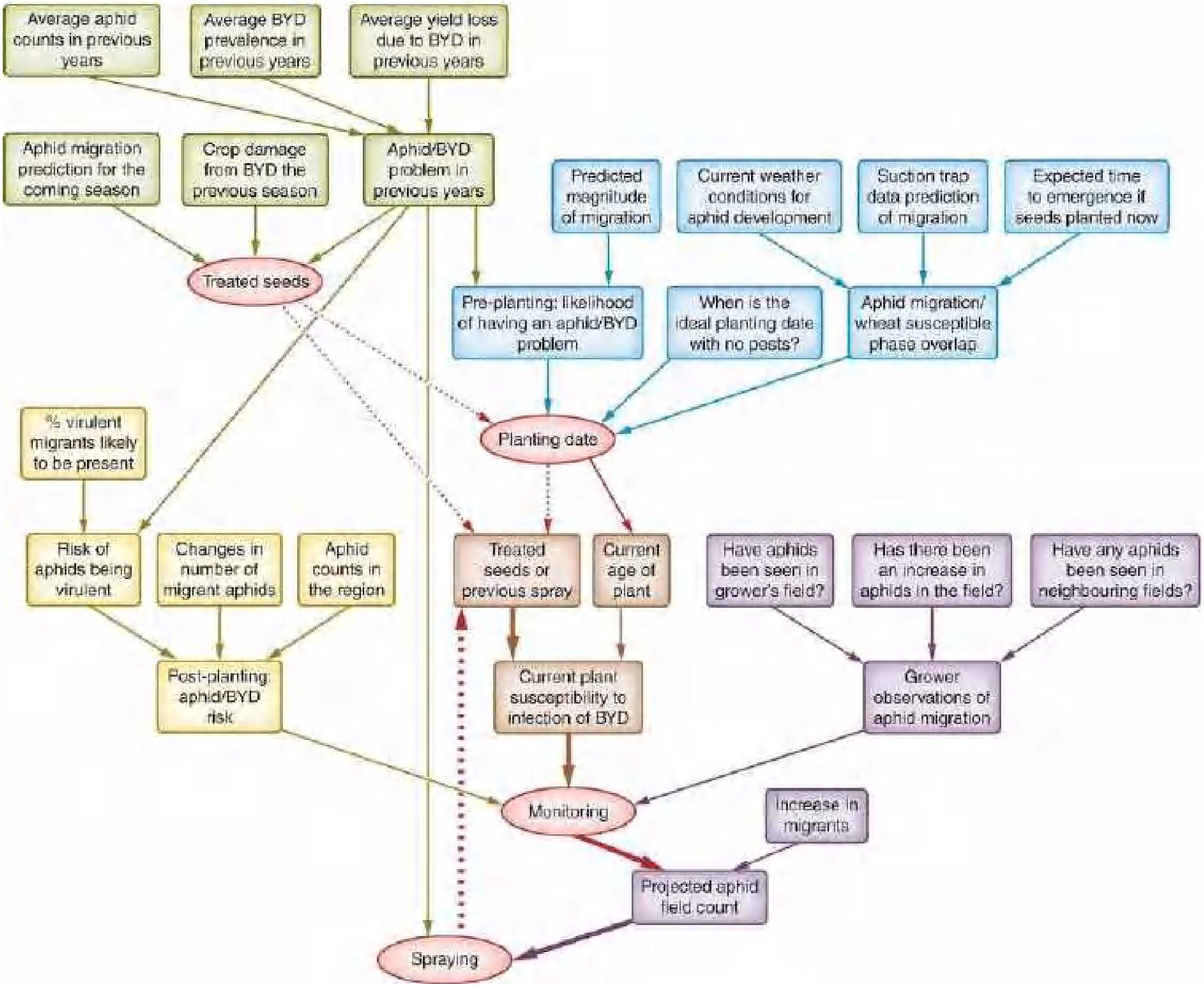
Starting about 50 years ago, recognition of the need to control pests on the one hand, and of the difficulties of doing so on the other, especially through the use of chemicals, led to the development of an approach termed integrated pest management (IPM). A recent survey suggests there have been more than 60 definitions of IPM (Young, 2017). This tells us that IPM is more an approach to pest control, or a philosophy, than a prescribed set of procedures or instructions. Its key features are, first, that action against pests of any sort should only be taken when it is necessary to do so – that is, when economic injury levels have been breached, or preferably, economic thresholds suggest they are liable to be breached. This, in turn, requires both that research effort is invested in determining what the EIL or ET is, and that monitoring effort is invested in assessing the status of the pest population, the crop and so on, relative to those thresholds.

Secondly, IPM aims to utilise the full range of control methods: physical control (for example, simply keeping pests away from crops by erecting barriers), cultural control (for example, rotating the crops planted in a field so pests cannot build up their numbers over several years), biological control, chemical control, and the use of resistant varieties of crop. But in choosing amongst these options, there is always a preference for the interventions that cause least damage to other components of the community. Hence, there is a disposition in favour, especially, of biological control and against chemical control. Chemical control is not ruled out, however. If the EIL is being breached, and no alternatives exist, then IPM schemes would generally prescribe the use of chemicals, though advice is likely to include precautionary measures aimed at targeting the pest and avoiding more widespread contamination.

The essence of the IPM approach, therefore, is to make the control measures fit the pest problem, and no two problems are the same – even in adjacent fields. Thus, IPM typically requires a farmer or specialist pest manager to first diagnose and assess pest problems before suggesting appropriate responses, and then monitoring the effects of these. Often this involves the development of computer-based decision support systems (DSSs) (Damos, 2015).

a decision support system for an IPM programme to control barley yellow dwarf

An example is outlined in Figure 15.6 for barley yellow dwarf (BYD), a virus disease of cereals causing massive grain losses worldwide and affecting more than 150 species. It is caused by viruses (five main ones) spread by aphids (four main species), which are therefore the primary target for the control of the disease. The details in the figure are much less important than the general principles. The most fundamental of these is that the various aspects of control – whether to pre-treat seeds, when to plant, whether and when to monitor for aphids, and whether and when to spray with insecticide – are not applied according to some predetermined schedule but on the basis of combining past and present information for the local area and acting accordingly. What ‘acting accordingly’ is, in this case, can be determined via a website or a smart phone app, opening up the prospect, in turn, of local farmers or pest managers not only accessing recommendations on site, but also feeding in local information to help others in the area (Walls *et al.*, 2016).



[Figure 15.6](#) Decision support systems allow flexible implementation of integrated pest management programmes. Outline of a decision support system for the control of barley yellow dwarf (BYD), a virus disease of cereals carried by aphids, which therefore need to be controlled. Data inputs over which the pest manager has no control are in rectangles; actions that the manager controls are in ovals. Simple arrows indicate the connectedness on which decision making is based. Decisions are made chronologically from top to bottom: whether to treat seeds, when to plant, whether and when to monitor the aphids, whether and when to spray with insecticide. Dotted arrows indicate the necessity for the local user to input additional data. The bold arrows, including the dotted one, represent a feedback loop through which monitoring and spraying may take place several times in a season.

Source: After Walls *et al.* (2016).

integration of IPM in sustainable farming systems

Implicit in the philosophy of IPM is the idea that pest control cannot be isolated from other aspects of food production, and it is especially bound up with the means by which soil fertility is maintained and improved. These broader sustainable agricultural systems are often referred to as integrated farming systems (IFS). Just as IPM strategies seek to replace synthetic pesticides with less environmentally invasive alternatives whenever possible, so IFS strategies also do the same with synthetic fertilisers. The adoption of both is certainly increasing. For example, in 2009, the European Union issued a Sustainable Use of Pesticides Directive insisting that from 1 January

2014 all farming within the Union should operate according to IPM principles, though the actual level of adoption is unknown and certainly less than 100% (Lefebvre *et al*., 2015). Ultimately, of course, adoption requires a perception that any shortfalls in yield are outweighed by wider environmental benefits – in terms of both short-term reductions in toxicity and longer-term gains in sustainability – and equally important, that those shortfalls in yield are accompanied by reductions in the costs of inputs (pesticides and fertilisers) such that profits are little if at all affected. Increasingly, studies are showing IFS systems delivering in these terms. Figure 15.7a, for example, shows a comparison between three farming strategies applied between 2003 and 2011 in Iowa, USA (Davis *et al*., 2012). The first was a conventional two-year maize-soybean rotation with 'normal' levels of herbicide application. The other two were IFS strategies, both with much reduced levels of herbicides and use of composted cattle manure, one a three-year and the other a four-year rotation, with both incorporating crops produced for livestock feed in the rotation. As Figure 15.7a, shows, the IFS strategies were far better in terms of environmental toxicity and so on, but also as good or even better than conventional farming in terms of yields and profit. A case can even be made for modern organic strategies, in which synthetic pesticides and fertilisers are banned altogether, despite their reduced yields, performing better than conventional farming not only environmentally but also in terms of profit (Figure 15.7b). The case for sustainable agriculture only being achievable at the expense of profits looks increasingly to be a weak one.

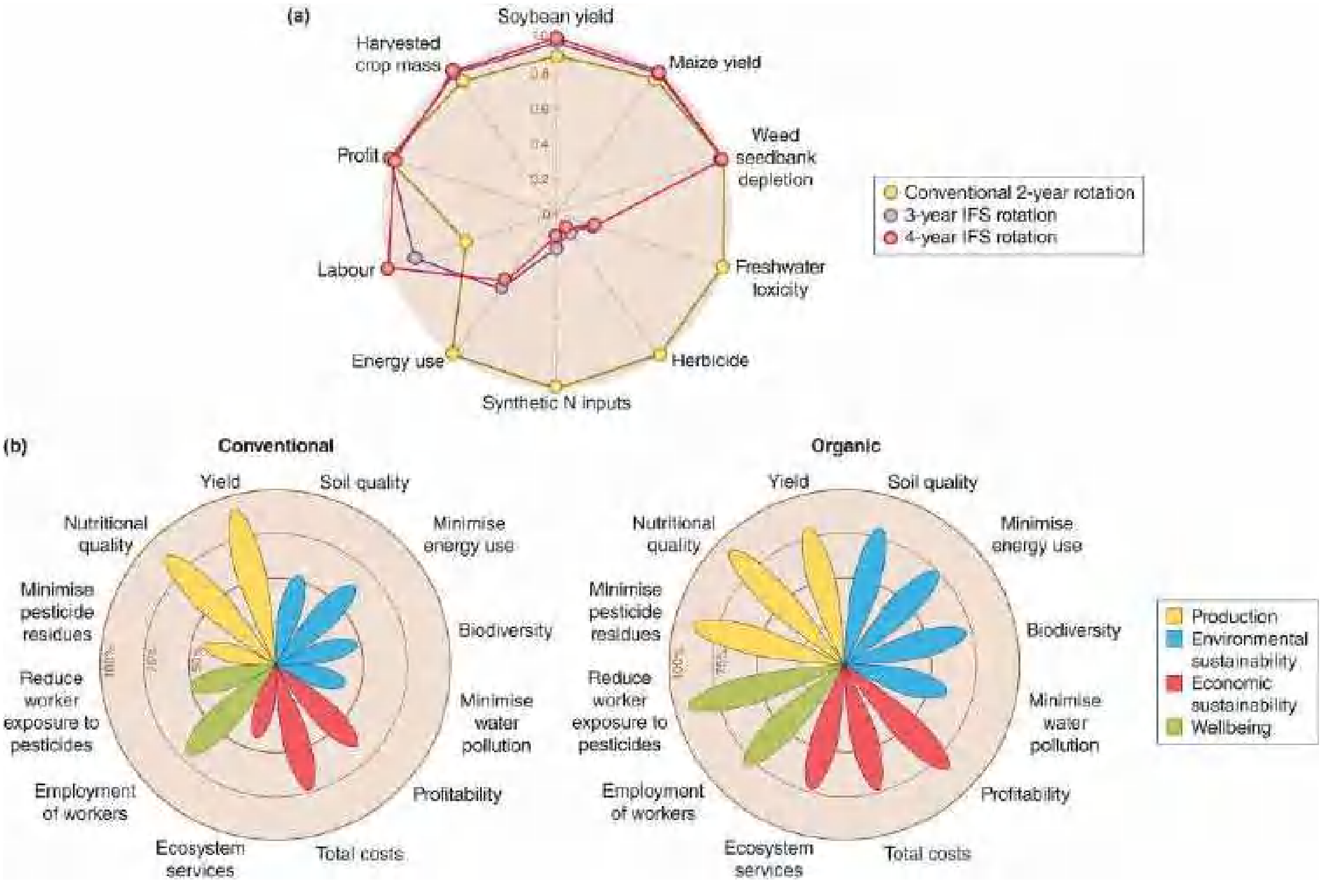


Figure 15.7 Integrated farming systems and even fully organic farming can be good for profits as well as for the environment. (a) The performance of three cropping systems for the production of maize and soybean in Iowa, USA, against multiple aspects of performance, as indicated. Means for each aspect are normalised on a scale from 0 to 1, with 1 representing the value for the best system for the variable concerned. The two integrated farming systems (IFS) performed as well as or better than the conventional system in each case. (b) A similar assessment of organic farming relative to conventional farming (made by advocates of organic farming from a wide review of the literature) in four colour-coded areas of sustainability, using ‘flower petals’: production (yellow), environmental sustainability (blue), economic sustainability (red) and wellbeing (green). Circles represent levels of performance: 25%, 50%, 75% and 100%. Ecosystem services are functions provided by ecosystems that support human interests, generally saving a cost that would otherwise need to be paid (see Section 15.4.1).

Source: (a) After Davis *et al*. (2012). (b) From Reganold & Wachter (2015).

15.3 Harvest management

harvesting aims to avoid over- and underexploitation

When a natural population is exploited by removing a portion of the population for our benefit (for food, timber and so on), while leaving a portion behind to continue to grow or to reproduce, we refer to this as harvesting or culling. In doing this, we want, on the one hand, to avoid overexploitation, where too many individuals are removed and the population is driven into biological jeopardy or economic insignificance – or perhaps even to extinction. But harvest managers also want to avoid underexploitation, where fewer individuals are removed than the population can bear, and a crop of food, for example, is taken which is smaller than necessary, threatening both the health of potential consumers and the livelihood of the harvesters. As we shall see, the best position to occupy between these two extremes is not easy to determine, since it needs to combine considerations that are not only biological (the well-being of the exploited population) and economic (the profits being made from the operation), but also social (local levels of employment and the maintenance of traditional lifestyles and human communities) (Clark, 2010). We begin, though, with the biology.

15.3.1 Maximum sustainable yield

MSY: the peak of the net recruitment curve

High yields are obtained from populations held below, often well below, their carrying capacity. This fundamental pattern is captured by the model population in Figure 15.8, where the natural net recruitment (or net productivity) of the population is described by a dome-shaped curve like those described in Section 5.4.2. Recruitment rate is low when there are few individuals and low when there is intense intraspecific competition. It is zero at the carrying capacity (K). The density giving the highest net recruitment rate depends on the exact form of intraspecific competition. This density is $K/2$ in the logistic equation (see Section 5.7) but, for example, is only slightly less than K for many large mammals (see Figure 5.14b). Always, though, the rate of net recruitment is highest at an 'intermediate' density, less than K.

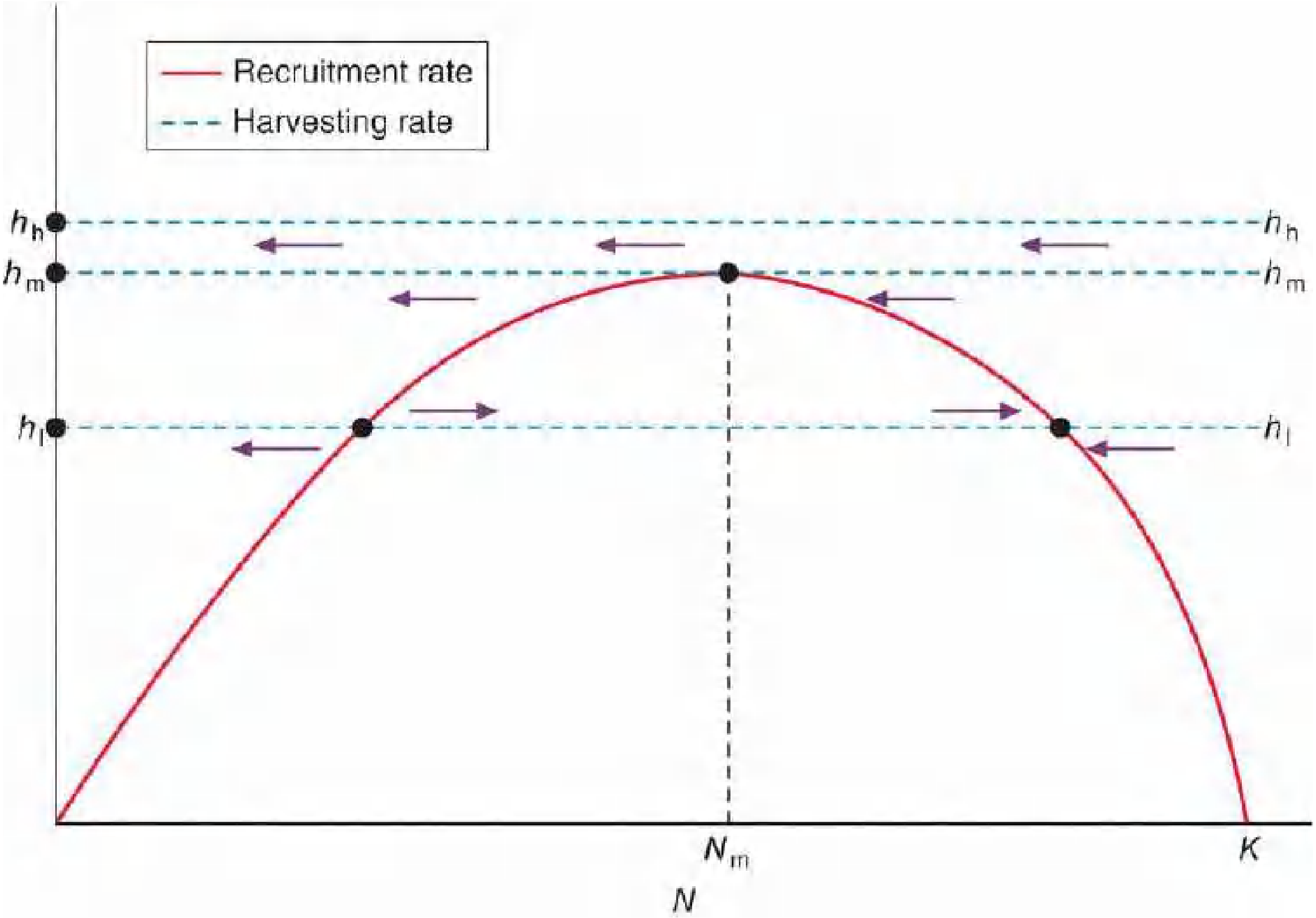


Figure 15.8 Fixed quota harvesting can achieve a maximum sustainable yield (MSY) by catching the peak of the net recruitment curve, but is a fragile strategy for doing so. The figure shows a single recruitment curve and three fixed quota harvesting curves: high quota (h_h), medium quota (h_m) and low quota (h_l). Arrows in the figure refer to changes to be expected in abundance under the influence of the harvesting rate to which the arrows are closest. •, equilibria. At h_h the only 'equilibrium' is when the population is driven to extinction. At h_l there is a stable equilibrium at a relatively high density, and also an unstable breakpoint at a relatively low density. The MSY is obtained at h_m because it just touches the peak of the recruitment curve (at a density N_m): populations greater than N_m are reduced to N_m, but populations smaller than N_m are driven to extinction.

Figure 15.8 also illustrates three possible harvesting regimes, although in each case there is a *fixed harvesting rate*, that is, a fixed number of individuals removed during a given period of time, or '*fixed quota*'. When the harvesting and recruitment lines cross, the harvesting and recruitment rates are equal and opposite; the number removed per unit time by the harvester equals the number recruited per unit time by the population. Of particular interest is the harvesting rate, h_m, the line that crosses (or, in fact, just touches) the recruitment rate curve at its peak. This is the highest harvesting rate that the population can match with its own recruitment. It is known as the *maximum sustainable yield* (MSY), and as the name implies, it is the largest harvest that can be removed from the population on a regular and repeated (indeed indefinite) basis. It is equal to the maximum rate of recruitment, and it is obtained from the population by depressing it to the density at which the recruitment rate curve peaks. More generally, simple models like this can be referred to as *surplus yield* models, since they are based on the idea of the harvester removing a crop that the population can 'afford' to give up.

MSY has severe shortcomings but has been frequently used

The MSY concept is central to much of the theory and practice of harvesting. This makes the recognition of its shortcomings all the more essential. Firstly, by treating the population as a number of similar individuals, or as an undifferentiated biomass, it ignores all aspects of population structure such as size or age classes and their differential rates of growth, survival and reproduction. Alternatives that incorporate structure are considered below. Secondly, by being based on a single recruitment curve it treats the environment as unvarying. Thirdly, in practice, it may be impossible to obtain a reliable estimate of the MSY. And finally, as we also discuss below, achieving an MSY is by no means the only, nor necessarily the best, criterion by which success in the management of a harvesting operation should be judged.

Nonetheless, in many systems, the MSY concept remains a guiding principle – for example, in 2013 it was reconfirmed as the basis for one of the three pillars of the European Common Fisheries Policy of the European Union (EU, 2013) – and by pursuing the concept, a number of the basic tenets of harvesting can be explained. Therefore, we begin with the MSY concept, but then examine its various shortcomings in more detail.

15.3.2 Harvesting strategies based on MSY

fixed-quota harvesting ...

The MSY density (N_m) is an equilibrium (gains = losses), but when harvesting is based on the removal of a fixed quota, as it is in Figure 15.8, N_m is a very fragile equilibrium. If the density exceeds the MSY density, then h_m exceeds the recruitment rate and the population declines towards N_m, as it should. But if, by chance, the density drops even slightly below N_m, then h_m will once again exceed the recruitment rate. Density will then decline even further, and if this fixed MSY quota is maintained, the population will decline until it is extinct. Similarly, if the MSY is even slightly overestimated, the harvesting rate will always exceed the recruitment rate (h_h in Figure 15.8), and again, repeatedly taking an 'MSY' quota will inevitably lead to extinction. Thus, a fixed quota at the MSY level might be desirable and reasonable in a wholly predictable world about which we have perfect knowledge. But in the real world of fluctuating environments and imperfect datasets, these fixed quotas are open invitations to disaster.

... whose dangers are illustrated by the Peruvian anchovy fishery

Nevertheless, a fixed-quota strategy was frequently used in the past. An example is provided by the Peruvian anchovy (*Engraulis ringens*) fishery. From 1960 to 1972 this was the world's largest single fishery, and it constituted a major sector of the Peruvian economy. Fisheries experts advised that the MSY was around 10 million tonnes annually, and catches were limited accordingly. But the fishing capacity of the fleet expanded, and in 1972 the catch crashed. Overfishing seems at least to have been a major cause of the collapse, although its effects were compounded with the influences of profound climatic fluctuations (see later). A moratorium on fishing would have been an ecologically sensible step, but this was not politically feasible: 20 000 people were dependent on the anchovy industry for employment. The stock took more than 20 years to recover (Figure 15.9).

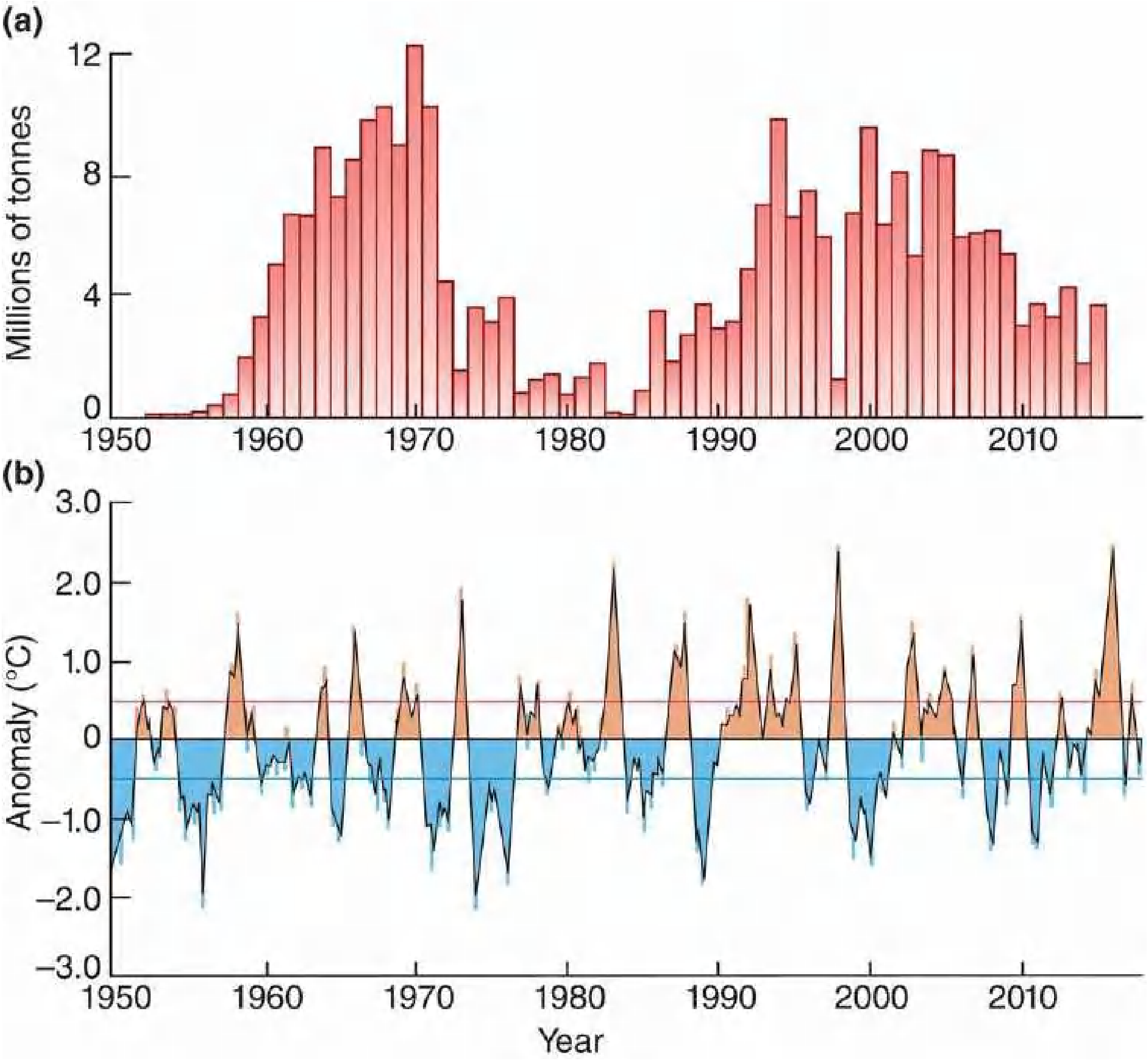


Figure 15.9 The Peruvian anchoveta fishery collapsed in 1972 as a result of overfishing, but has also been adversely affected by El Niño events. (a) Landings of the Peruvian anchovy, *Engraulis ringens*, since 1950. (b) The El Niño–Southern Oscillation (ENSO) over the same period as measured by sea surface temperature anomalies (differences from the mean) in the equatorial mid-Pacific. The El Niño events are significant negative anomalies.

Source: (a) After Salvatteci *et al.* (2018).

regulating harvesting effort is less risky – but leads to a more variable catch

The risk associated with fixed quotas can be reduced if instead there is regulation of the harvesting *effort*. The yield from a harvest (H) can be thought of, simply, as being dependent on three things. It increases with the size of the harvested population, N, with the level of harvesting effort, E (e.g. the number of 'trawler-days' in a fishery or the number of 'gun-days' with a hunted population), and with harvesting efficiency, q. Hence,

$$H = qEN. \tag{15.2}$$

On the assumption that this efficiency remains constant, Figure 15.10a depicts an exploited population subjected to three potential harvesting strategies differing in harvesting effort. Figure 15.10b then illustrates what is generally referred to as a *yield-effort curve*: the overall relationship to be expected, in a simple case like this, between effort and average yield. There is an apparently 'optimum' effort, giving rise to the MSY, E_m, with efforts both greater and less than this giving rise to smaller yields.

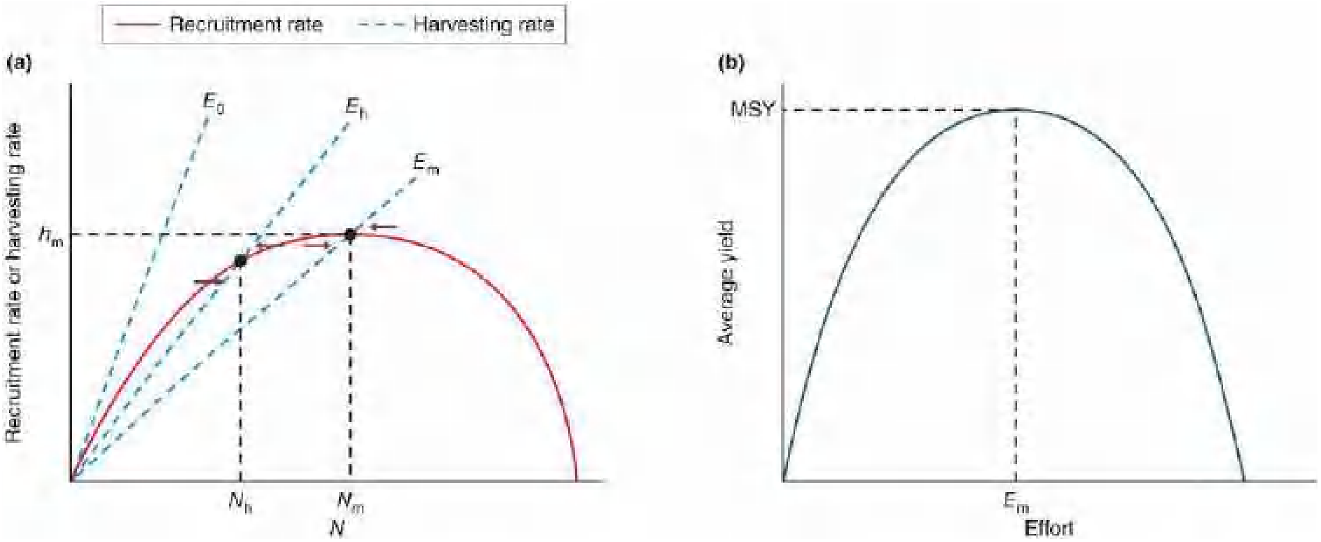


Figure 15.10 Fixed effort harvesting can deliver a stable maximum sustainable yield (MSY). (a) Curves, arrows and dots are as in Figure 15.8. The MSY is obtained with an effort of E_m, leading to a stable equilibrium at a density of N_m with a yield of h_m. At a somewhat higher effort (E_h), the equilibrium density and the yield are both lower than with E_m but the equilibrium is still stable. Only at a much higher effort (E_o) is the population driven to extinction. (b) The overall relationship between the level of the fixed effort and average yield.

Adopting an 'MSY effort' is a much safer strategy than fixing an MSY quota, since now, in contrast to Figure 15.8, if density drops below N_m (Figure 15.10a), recruitment exceeds the harvesting rate and the population recovers. In fact, there needs to be a considerable overestimate of E_m before the population is driven to extinction (E_o in Figure 15.10a). However, because there is a fixed effort, the yield varies with population size. Whenever the population size, as a result of natural fluctuations, drops below N_m, yield will be less than the MSY. The appropriate reaction is to reduce effort whilst the population recovers. But an understandable (albeit misguided) reaction might be to compensate by increasing the effort. This, however, is likely to depress population size further (E_h in Figure 15.10a), and it is therefore easy to imagine the population being driven to extinction as very gradual increases in effort chase an ever-diminishing yield.

To prevent this, it is necessary first to determine an appropriate intensity of harvesting – limiting the number of gun licences issued to hunters, or regulating the size and composition of a fishing fleet – then to monitor that effort, and ultimately to restrain or even punish those exceeding what's allowed. None of these are easy. For example, vessel monitoring systems of various sorts are now in place around the world, allowing managers to track remotely the activities of vessels that are obliged to carry appropriate technology on board. Yet even in relatively advanced economies such as the UK, the systems have limitations, being required only for vessels exceeding 12 metres in length, leaving 85% of the UK fleet, concentrated in inshore waters, unmonitored in this way. Survey data collected on a voluntary basis from these smaller vessels shows how foci of intense fishing activity may thereby be missed (Figure 15.11).

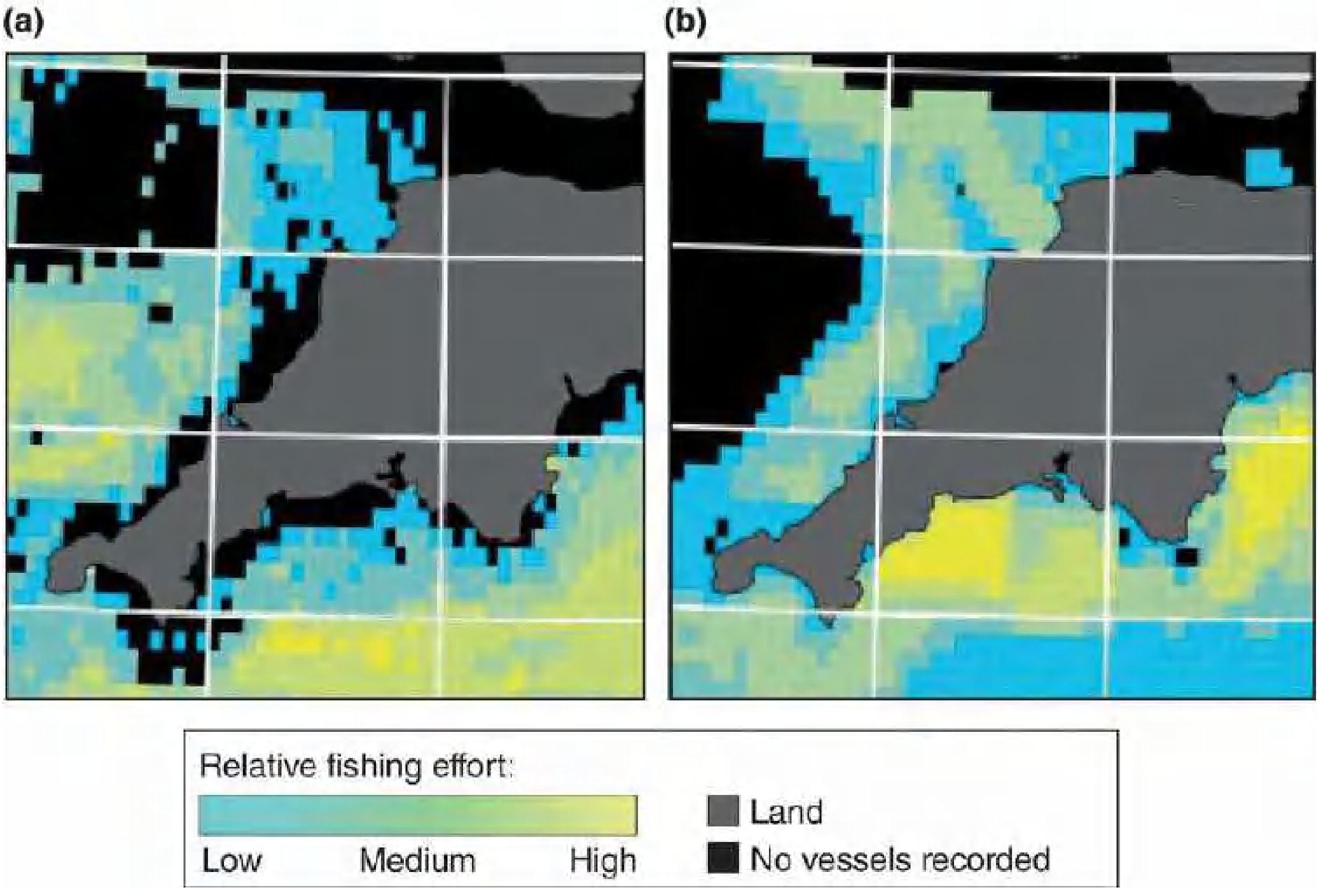


Figure 15.11 The discrepancy between official data from vessels with monitoring devices and less technical but more broadly based data shows how difficult it can be to monitor and regulate fishing effort. Fishing effort around the south-west of the UK, coded from blue (light) through green to yellow (high). Grey indicates land; black indicates that no vessels were recorded. (a) The effort for vessels over 15 m in length, 2007–2010, obtained from official Vessels Monitoring Systems (VMS) data using devices carried by the vessels, based on fishing hours and hence combining the number of vessels with the length of time each fished. (b) The effort for vessels less than 15 m, obtained from fishermen interviews, since smaller vessels do not carry the monitoring devices.

Source: After Enever *et al*. (2017).

fixed proportion and fixed ‘escapement’ harvesting

Two further management strategies are based on the simple idea of a surplus yield. Firstly, a *fixed proportion* of the population can be harvested (this is equivalent to fixing a hunting mortality rate and should have the same effect as harvesting at constant effort); and second, a fixed number of breeding individuals can be left at the end of each hunting or fishing season (*fixed escapement*) – an approach that involves major investment in monitoring but is particularly useful for annual species, because they lack the buffer provided by immature individuals in longer lived species (Milner-Gulland & Mace, 1998). For example, the Alaskan Bristol Bay fishery for sockeye salmon (*Oncorhynchus nerka*) and other salmonids is the largest high-value salmon fishery in North America. It is managed to accomplish location-specific escapement targets chosen, and repeatedly updated, to deliver an MSY. Fisheries are not opened until it is established that the ‘run strength’ (as the fish return from the sea to their rivers to breed) is sufficient to meet the escapement target, following which the surplus is allocated to harvest (Steiner *et al*., 2011).

regulating fixed proportion harvesting through monitoring mortality compensation

Another way of viewing the dynamics captured by Figure 15.10 is to relate the harvest yield (that is, the mortality attributable to harvesting) to the natural mortality in the population. An ideal, in terms of sustainability, would be if the survival in harvested populations increased enough to compensate for the losses due to harvesting mortality. In such a case, harvesting would simply be removing what natural mortality would have removed anyway, sometimes called the ‘doomed surplus’ (Errington, 1934). At the other extreme, there might be no compensation – natural and harvesting mortality might be purely additive – and between the two extremes, mortality might be partially compensatory. The pattern of compensation will vary from case to case (Sandercock *et al*., 2011). For example, mortality is more likely to be compensatory at lower harvesting rates and at higher abundances, close to the carrying capacity, where natural mortality rates are high. It is more likely to be additive where abundance is lower, and the population has less capacity to bounce back. Perhaps most important in practice, the level of compensation will also depend on the timing of harvesting, and of phases of high natural mortality, within any seasonal cycle exhibited by the population. If harvesting precedes a period of high natural mortality, then the response is more likely to be compensatory; whereas when harvesting occurs during or after such a period, the response is more likely to be additive.

The willow ptarmigan, *Lagopus lagopus* (closely related to the red grouse, *L. lagopus scoticus*, that we have met elsewhere in this text) is hunted, and managed, at many locations in northern latitudes. The ptarmigan, like most upland game birds, is harvested according to a fixed proportion strategy. The question therefore arises as to how overall mortality rates will respond to variations in the proportional harvest. In a study in Norway, the experimental hunting season ran from mid-September to the end of October (though normally it would extend to late February or March). At five sites, over three years, survival was monitored year-round, while the sites themselves were subjected, at random, to harvesting rates of 0% (control), 15% or 30%. At 15%, a fall in the rate of natural mortality partially compensated for the harvesting mortality; but at 30%, harvesting and natural mortality were close to being additive (Figure 15.12).

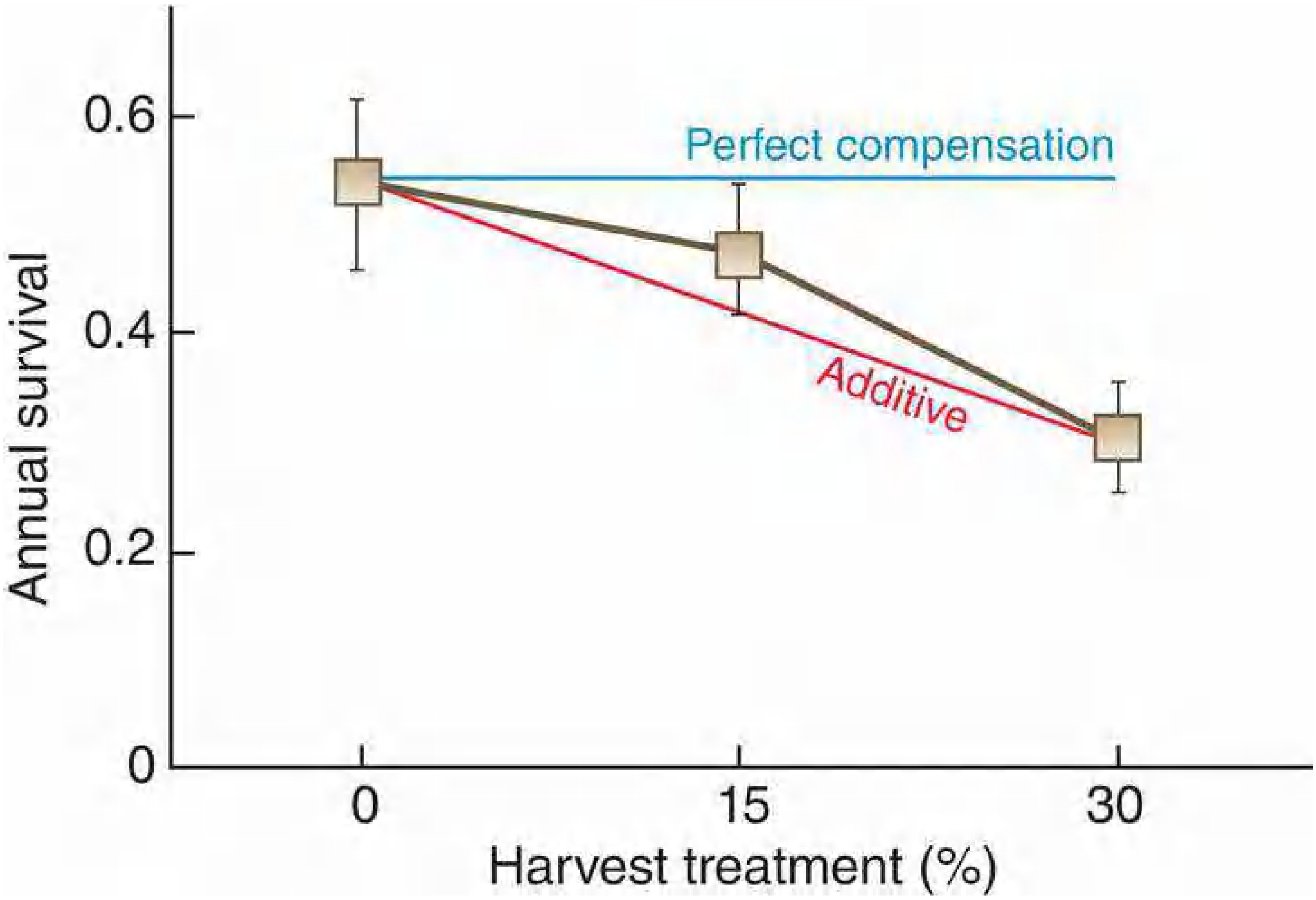


Figure 15.12 Density compensation following fixed proportional harvesting in the willow ptarmigan. Annual survival rates (with SEs) of willow ptarmigan, following harvesting, as indicated, at experimental sites in Norway. The lines for how the relationship would look for perfectly compensatory and additive mortality are also shown, though these are indicative rather than exact, since the exact relationship depends also on timing of mortality, actual harvesting mortality rates, etc.

Source: After Sandercock *et al*. (2011).

The key here, as in many cases, seems to be the seasonality and nature of potential compensatory mechanisms. One important period in the year of high natural mortality for the willow ptarmigan is in September and October. This coincides with harvesting, and is associated with juvenile dispersal and the early establishment of male territories. The coincidence itself limits the potential for compensatory increases in survival. Equally important, though, the main compensatory mechanism in this case, increasing at low densities, would be the immigration of juveniles. However, these juveniles are themselves subject to hunting, which does not discriminate between juveniles and adults, and this limits both the supply of juveniles (from areas that are also harvested) and their survival should they arrive. Together, these factors, and the apparent absence of other compensatory mechanisms at this and other times of the year, are likely to account for the only limited compensation at moderate harvesting levels and its absence when levels are higher. Whatever the underlying cause, results like these point to the level of proportional harvesting that may be sustainable – possible at rates of around 15%, but unlikely up at around 30%.

15.3.3 Economic and social factors

the economically optimum yield – typically less than the MSY

Simple MSY models can also be used to highlight some of the shortcomings of an overreliance on MSYs. One of the most obvious of these is the failure, through a purely ecological approach, to recognise that the exploitation of a natural resource is usually a business enterprise, in which the value of the harvest must be set against the costs of obtaining that harvest. Even if we distance ourselves from any preoccupation with ‘profit’, it makes no sense to struggle to obtain the last few tonnes of an MSY if the money spent in doing so could be much more effectively invested in some other means of food production. The basic idea is illustrated in Figure 15.13. We seek to maximise not total yield but net value – the difference between the gross value of the harvest and the sum of the fixed costs (interest payments on ships or factories, insurance, etc.) and variable costs, which increase with harvesting effort (fuel, crew’s expenses, etc.). Figure 15.13 shows that the *economically optimum yield* (EOY) is likely to be less than the MSY, and is obtained through a smaller effort or quota. However, the difference between the EOY and the MSY will vary, being least in enterprises where most costs are fixed (the ‘total cost’ line is virtually flat). This is especially the case in high investment, highly technological operations such as deep-sea fisheries, which are therefore most prone to overfishing even with management aimed at economic optima.

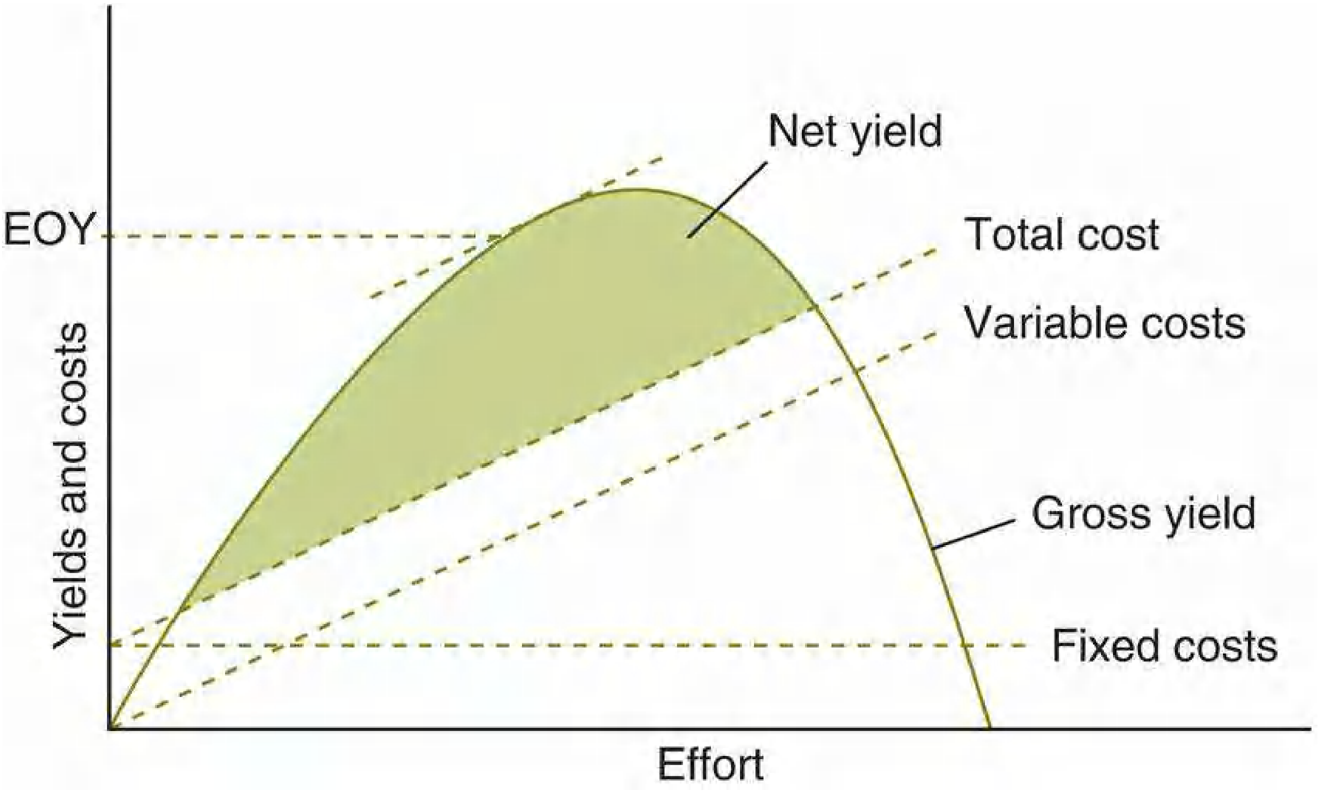


Figure 15.13 The economically optimum yield (EOY) is often lower than the maximum sustainable yield. The EOY, that which maximises ‘profit’, is obtained to the left of the peak of the yield-against-effort curve, where the difference between gross yield and total cost (fixed costs plus variable costs) is greatest. At this point, the gross yield and total cost lines have the same slope.

Source: After Hilborn & Walters (1992).

discounting: liquidating stocks, or leaving them to grow?

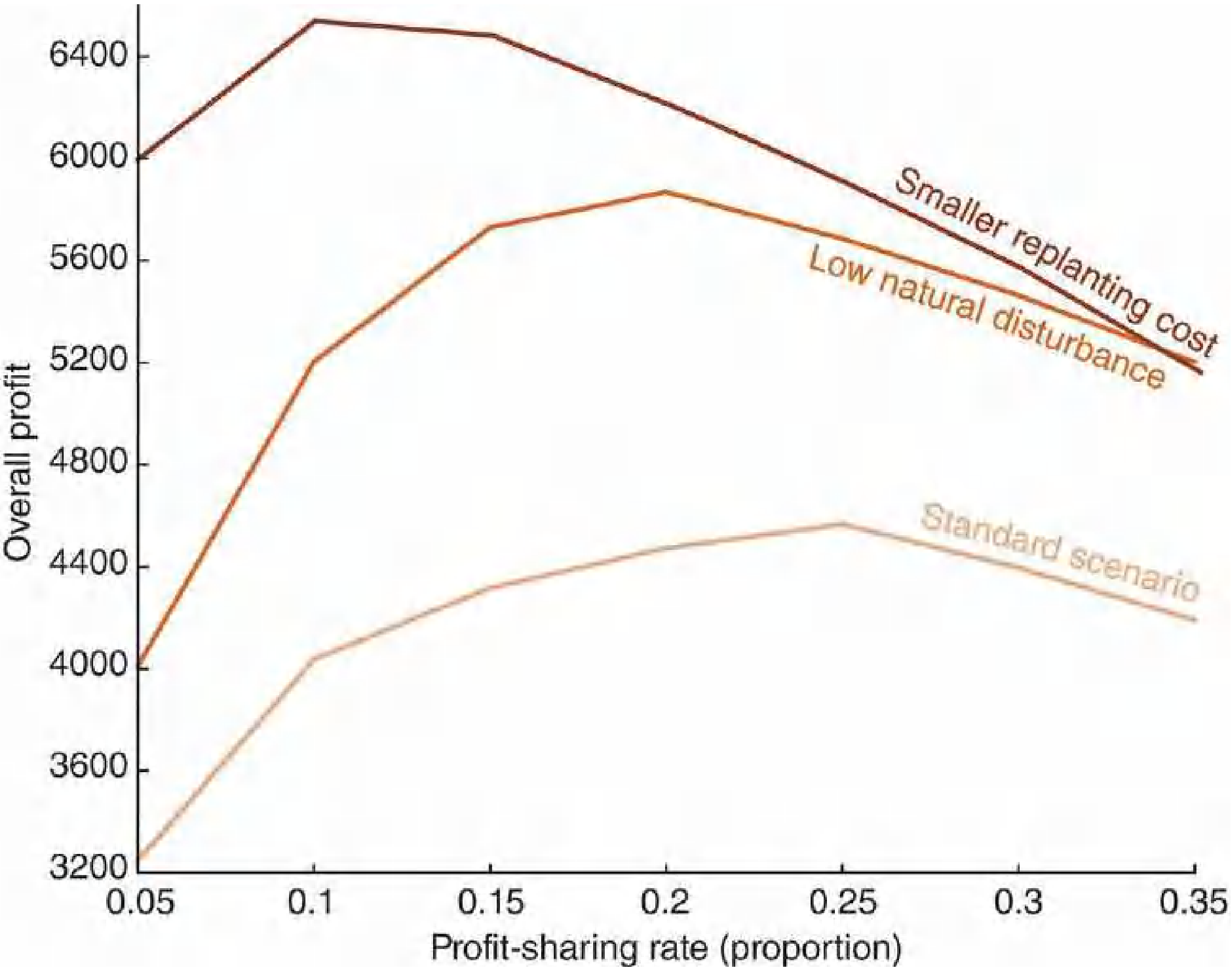
A second important economic consideration is the proposition that in economic terms, each fish caught now is worth more than an equivalent fish caught in the future – referred to technically as ‘discounting’. The reason, put simply, is that the value of the current catch can be placed in the bank to accrue interest, so that its total value increases. In fact, a commonly used discount rate for natural resources is 10% per annum (90 fish now are as valuable as 100 fish in one year’s time) despite the fact that the difference between the interest rates in the banks and the rate of inflation is usually much less than this. The justification is a desire to incorporate ‘risk’: a fish caught now has already been caught, whereas one still in the water might or might not be caught – a bird in the hand is worth two in the bush.

On the other hand, the caught fish is dead, whereas the fish still in the water can grow and breed (although it may also die). In a very real sense, therefore, each uncaught fish may be worth more than ‘one fish’ in the future. In particular, if the stock left in the water grows faster than the discount rate, as is commonly the case, then a fish put on deposit in the bank is not so sound an investment as a fish left on deposit in the sea. Nevertheless, even in cases like this, discounting provides an economic argument for taking larger harvests from a stock than would otherwise be desirable.

Moreover, in cases where the stock is less productive than the discount rate – for example, many whales and a number of long-lived fish – it seems to make sense, in purely economic terms, not only to overfish the stock, but actually to catch every fish (‘liquidate the stock’). The reasons for not doing so are partly ethical – it would be ecologically short sighted and a disdainful way of treating the hungry mouths to be fed in the future. But there are also practical reasons: jobs must be found for those previously employed in the fishery (or their families otherwise provided for), alternative sources of food must be found, and so on. This emphasises, first, that a ‘new economics’ must be forged in which value is assigned not only to things that can be bought and sold – like fish and boats – but also to more abstract entities, like the continued existence of whales or other ‘flagship species’ (TEEB, 2010). It also stresses the danger of an economic perspective that is too narrowly focused. The profitability of a harvesting operation cannot sensibly be isolated from the implications that the management of that operation has in a wider sphere.

profit sharing to prevent illegal logging of teak

An instructive example of where biological, economic and social considerations coincide is the production of teak timber, *Tectona grandis*, in Indonesia. The rate of production by the major producer, The Javanese State Forestry Company, had been seriously decreased by illegal logging, prompting the introduction of a management scheme in which local people were paid to participate in surveillance to suppress illegal logging by offering them a share in the profits from the timber harvest, and were also allowed to cultivate cash crops on the land while the planted trees were too small to shade them (Fujiwara *et al.*, 2012). A model was then developed to investigate what the optimal level of profit-sharing might be, that is, optimal for the management company (Lee *et al.*, 2018): too low a level and locals would feel insufficiently compensated and motivated, and surveillance would be ineffective in stemming the losses to illegal logging – too high a level and the company’s profits would be unnecessarily cut. In the model, locals were paid both for their surveillance and for replanting trees following harvest. Some of the results are shown in Figure 15.14.



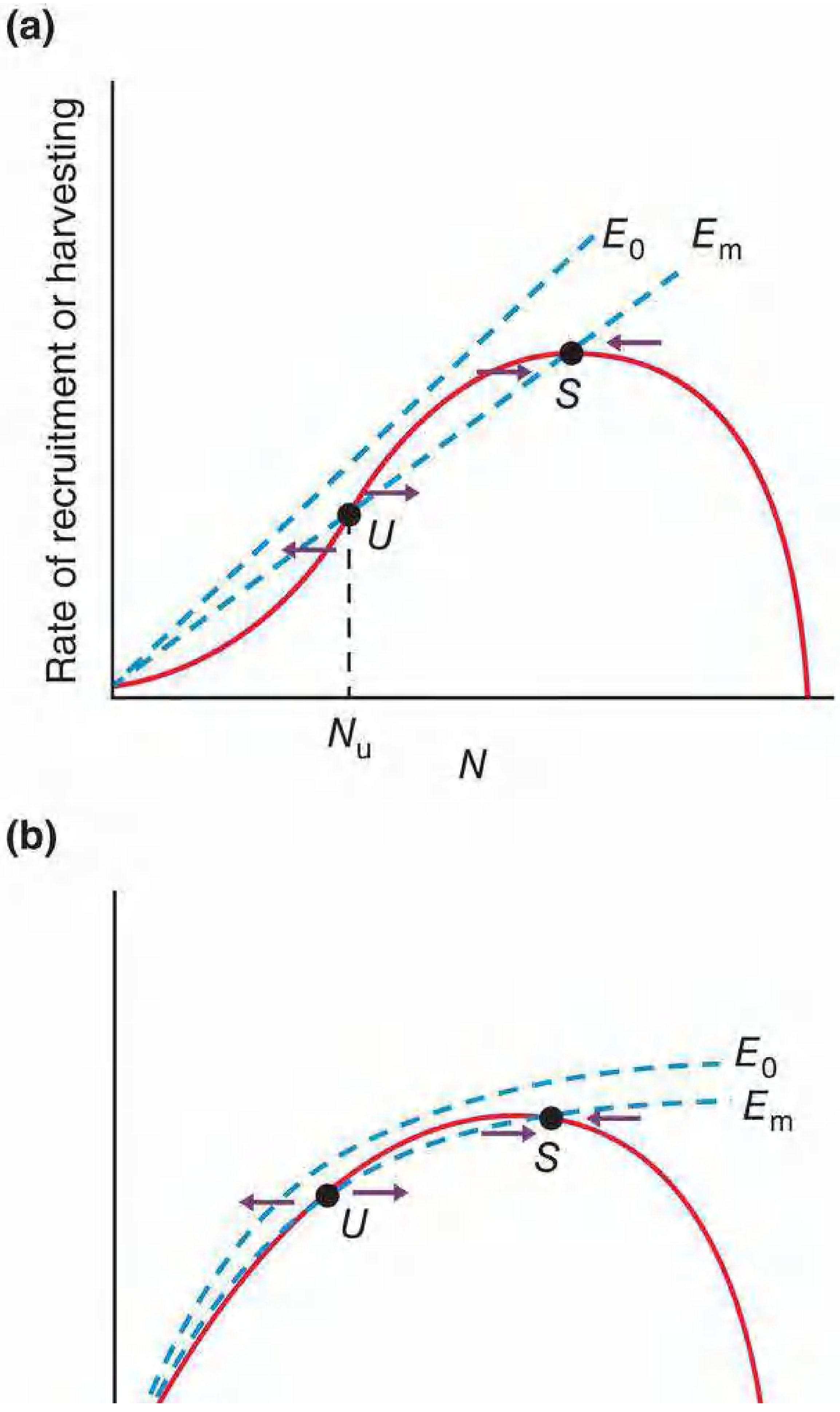
[Figure 15.14](#) Profit sharing in the harvesting of teak can be adjusted so as to maximise overall profit. Each curve describes the estimated overall profit (model output, arbitrary units) from harvesting teak, *Tectona grandis*, in Indonesia at various rates of profit sharing (the proportion of profit paid to local workers), either in a standard scenario, or when levels of natural disturbance are low, or where replanting costs are reduced.

Source: After Lee *et al*. (2018).

These confirm that an intermediate level is optimal, but they also indicate some of the factors that determine what that level should be. The underlying biology is important. When rates of natural disturbance in the forest are low, the optimal tree age for harvest is increased since there is less chance of tree loss prior to harvest. These older, larger trees are more valuable, decreasing the optimal level of profit sharing, because locals are more willing to participate in surveillance at lower disturbance rates given the greater value of each harvested tree. The detailed psychology of the interacting agents also plays a part. If payments to locals for replanting are reduced, locals are less tempted to overlook illegal logging that would bring replanting forward. Their increased motivation again reduces the optimal level of profit sharing (Figure 15.14). Clearly here, as so often, biology is important but not the only consideration.

15.3.4 Instability of harvested populations: depensation and multiple equilibria

Finally, simple MSY models highlight the fact that the dynamics of harvested populations may be highly fragile because the system within which they operate has 'multiple equilibria' (see Section 14.7). There are two scenarios in particular where this may be the case. The recruitment rate may be particularly low in the smallest populations (a pattern known as *depensation*) or more generally as an Allee effect (Section 10.3.2) (Figure 15.15a). Allee effects can arise for a variety of reasons. In marine fisheries, and elsewhere, these fall broadly into two categories. Firstly, there may be increases in natural mortality at low densities, or decreases in birth rate as a result of reduced mating frequencies. In fact, analyses of available datasets suggest such patterns may be uncommon, although apparent rarity may be more a reflection of shortcomings in the datasets themselves or the ways they have been analysed (Hutchings, 2014 ; Perälä & Kuperinen, 2017). Alternatively, multiple equilibria may arise because harvesting efficiency increases in small populations (Figure 15.15b). For instance, many clupeids (sardines, anchovies, herring) are especially prone to capture at low densities, because they form a small number of large schools that follow stereotyped migratory paths that the trawlers can intercept.


(a)
Rate of recruitment or harvesting
E_0
E_m
S
U
N_u
N
(b)
E_0
E_m
S
U

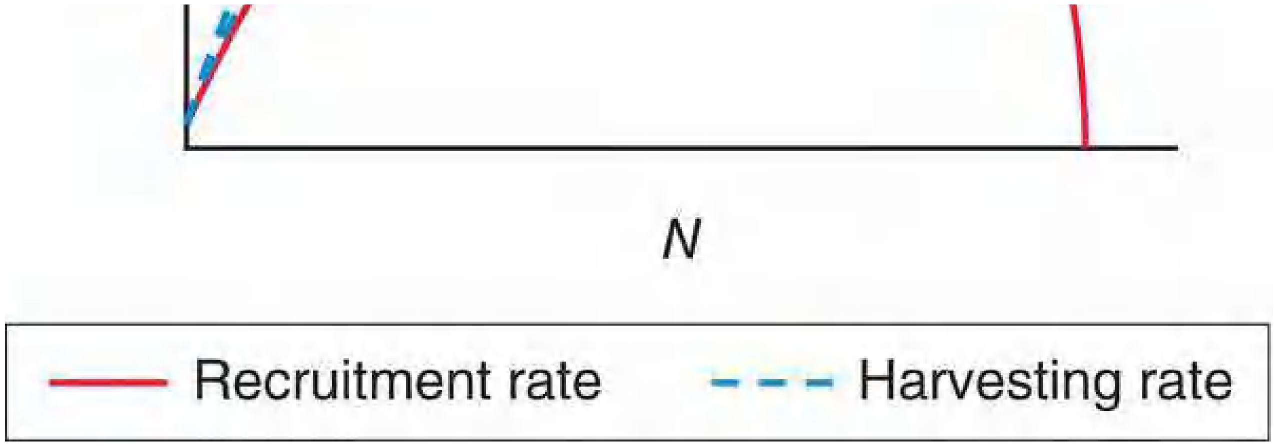
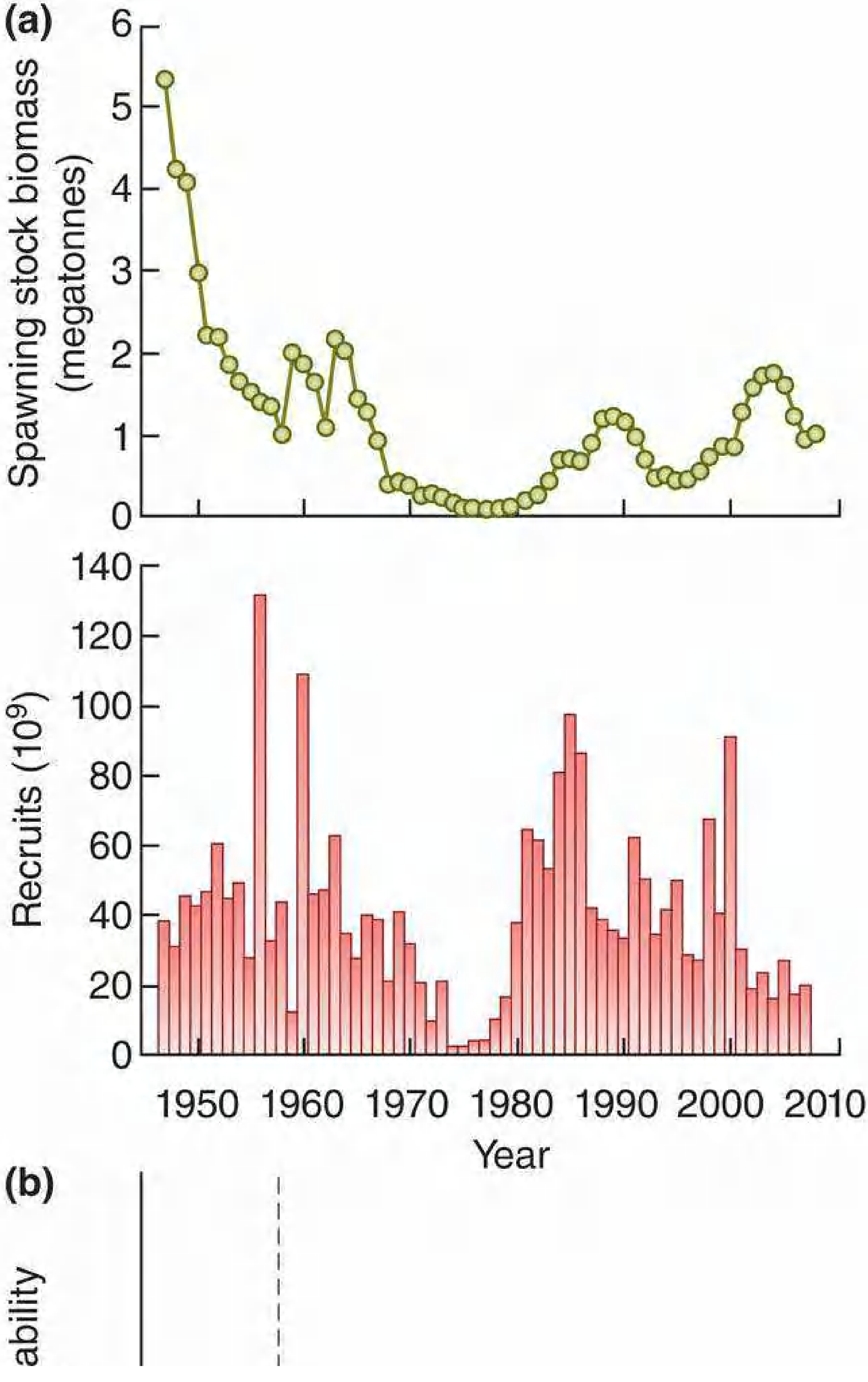


Figure 15.15 Harvesting operations may have multiple equilibria. (a) When recruitment rate is particularly low at low densities, the harvesting effort giving the MSY (E_{m}) has not only a stable equilibrium (S) but also an unstable breakpoint (U). Below this abundance, the population declines to extinction. The population can also be driven to extinction by harvesting efforts (E_{o}) not much greater than E_{m}. (b) Similar patterns are observed when harvesting efficiency declines at high densities.

In both cases, the harvesting line crosses the recruitment curve at two points (Figure 15.15). One, point S, is the usual stable equilibrium at the MSY abundance, but point U is an unstable 'breakpoint'. If the population drops below the MSY abundance but above the abundance at U (N_{u}), it returns to the MSY abundance (Figure 15.15a). However, if it drops even slightly below N_{u}, perhaps as a result of only a very small increase in effort, then the harvesting rate will be greater than the recruitment rate, and the population will be *en route* to extinction. And once the population is on this slippery slope, much more than a marginal reduction in effort is required to reverse the process. This is the crucial, practical point about multiple equilibria: a very slight change in behaviour can shift the system from one state to another – a shift that may be difficult to reverse.

Certainly, there are examples of drastic collapses in stock abundance followed by slow or only partial recovery. We saw this with the Peruvian anchovies, and another example, the North Sea herring (*Clupea harengus*), is shown in Figure 15.16a. To assess whether depensation may play a part in driving these dynamics, Perälä and Kuperinen (2017) examined the relationship between abundance and recruitment in the stock and fitted statistical models that included a parameter, c, the value of which indicated whether that relationship was depensatory (per capita recruitment decreasing at lower densities; $c > 1$) or compensatory (per capita recruitment increasing at lower densities; $c < 1$). The analysis suggested strongly that there was depensation (Figure 15.16b): having crashed, the stock found itself on the wrong side of an unstable breakpoint, such that only a profound and sustained reduction in harvesting effort could allow it to recover.


(a)
Spawning stock biomass (megatonnes)
6
5
4
3
2
1
0
Recruits (10^9)
140
120
100
80
60
40
20
0
1950
1960
1970
1980
1990
2000
2010
Year
(b)
ability

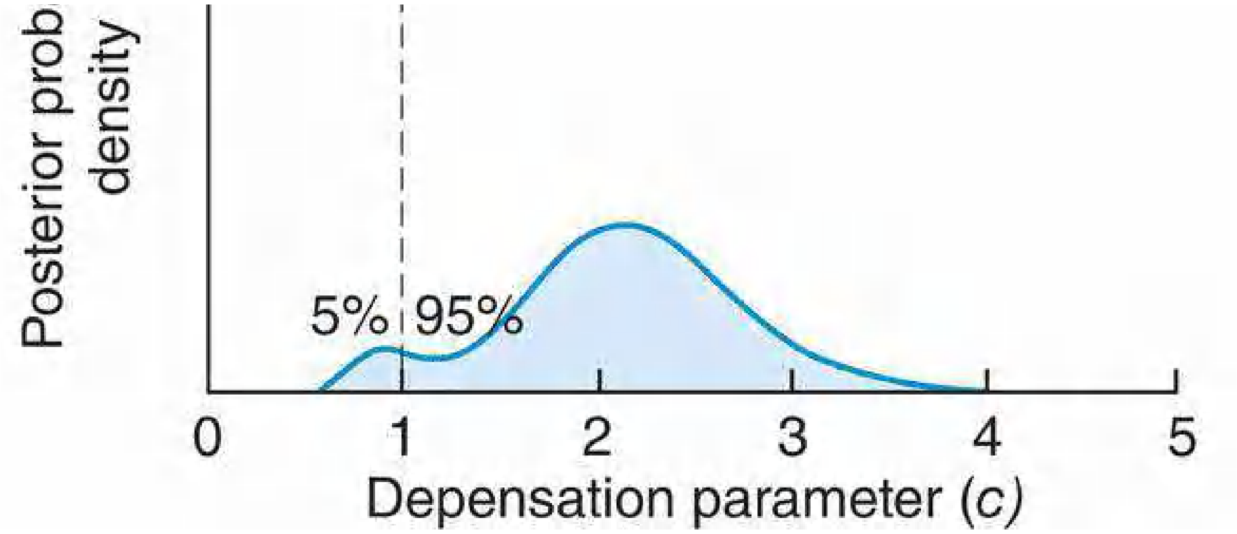


Figure 15.16 Depensation effects led to the sudden collapse and slow recovery of the North Sea herring fishery. (a) The spawning stock biomass (above) and the recruitment into the population (below) of North Sea herring, *Clupea harengus*, 1947–2009. (b) Results from a statistical model fitted to the relationship between stock and recruitment for North sea herring, which included a parameter, *c*, which signifies depensation when $c > 1$.The figure shows the 'posterior marginal probability' for *c*, indicating that depensation is highly probable.

Source: (a) After Dickey-Collas *et al.* (2010). (b) After Perälä & Kuperinen (2017).

15.3.5 Instability of harvested populations: environmental fluctuations

Instability in harvested populations may also stem from the effects of external forces. Not all fisheries' collapses are simply the result of overfishing and human greed. Fishing pressure does indeed often exert a great strain on the ability of natural populations to sustain levels of recruitment that counteract overall rates of loss. But the immediate cause of a collapse – in one year rather than any other – is often the occurrence of unusually unfavourable environmental conditions. Moreover, when this is the case, the population is more likely to recover (once conditions have returned to a more favourable state) than it would be if the crash was the result of overfishing alone.

the Peruvian anchovy and El Niño

Returning to the classic example of the Peruvian anchovy (Figure 15.9), prior to its major collapse from 1972 to 1973, the fishery had already suffered a dip in the upward rise in catches in the mid-1960s, as a result of an 'El Niño event': the incursion of warm tropical water from the north severely reducing ocean upwelling, and hence productivity, within the cold Peruvian current coming from the south (see Section 2.4.1). By 1973, however, because fishing intensity had so greatly increased, the effects of a subsequent El Niño event were much more severe. Moreover, whilst the fishery showed some signs of recovery from 1973 to 1982 in spite of largely unabated fishing pressure, a further collapse occurred in 1983 associated with yet another El Niño event, and there have been subsequent El Niño-associated setbacks. Clearly, it is unlikely that the consequences of these natural perturbations to the current flow would have been so severe if the anchovy had not been exploited or had been only lightly fished. It is equally clear, though, that the history of the Peruvian anchovy fishery cannot be understood properly in terms simply of fishing, as opposed to natural events, and indeed, the details of how the consequences of those natural events play out, and interact with fisheries' practices, remain a matter of debate (Salvatteci *et al.*, 2018).

15.3.6 Recognising structure in harvested populations: dynamic pool models

‘dynamic pool’ models recognise population structure

The simple ‘surplus yield’ models of harvesting that we have pursued so far are useful as a means of establishing basic principles (like MSY), and they are good for investigating the possible consequences of different types of harvesting strategy. But they ignore population structure, and this is a bad fault for two reasons. Firstly, ‘recruitment’, in practice, is a complex process incorporating adult survival, adult fecundity, juvenile survival, juvenile growth, and so on, each of which may respond in its own way to changes in density and harvesting strategy. Secondly, most harvesting practices are primarily interested in only a portion of the harvested population (e.g. mature trees, or fish that are large enough to be saleable). These complications are taken into account in what are usually called ‘dynamic pool’ models (for reviews, see Shepherd & Pope, 2002a , 2002b).

The general structure of a dynamic pool model is illustrated in Figure 15.17. There are four submodels (for recruitment rate, growth rate, natural mortality rate and fishing rate of the exploited stock) that combine to determine the exploitable biomass of the stock and the way this translates into a yield to the fishing community. In contrast to the surplus yield models, this biomass yield depends not only on the number of individuals caught but also on their size (past growth); whilst the quantity of exploitable (i.e. catchable) biomass depends not just on ‘net recruitment’ but on an explicit combination of natural mortality, harvesting mortality, individual growth and recruitment into catchable age classes.

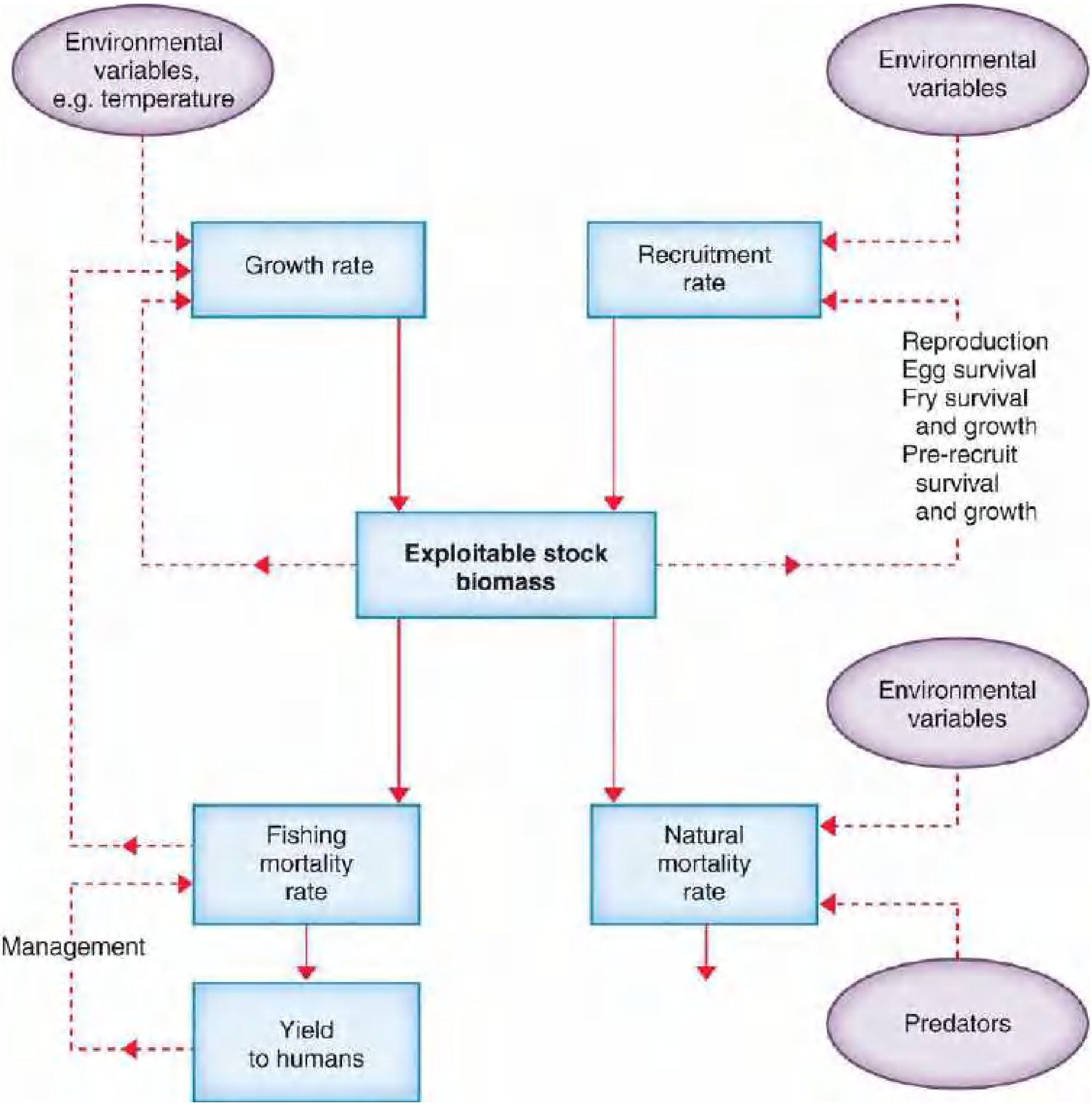


Figure 15.17 The dynamic pool approach to fishery harvesting and management, illustrated as a flow diagram. There are four main 'submodels': the growth rate of individuals and the recruitment rate into the population (which add to the exploitable biomass), and the natural mortality rate and the fishing mortality rate (which deplete the exploitable biomass). Solid arrows refer to changes in biomass under the influence of these submodels. Dashed arrows refer to influences either of one submodel on another, or of the level of biomass on a submodel or of environmental factors on a submodel. Each of the submodels can itself be broken down into more complex and realistic systems. Note that yield-per-recruit analyses (see Figure 15.18) ignore the important arrow from exploitable stock biomass to recruitment.

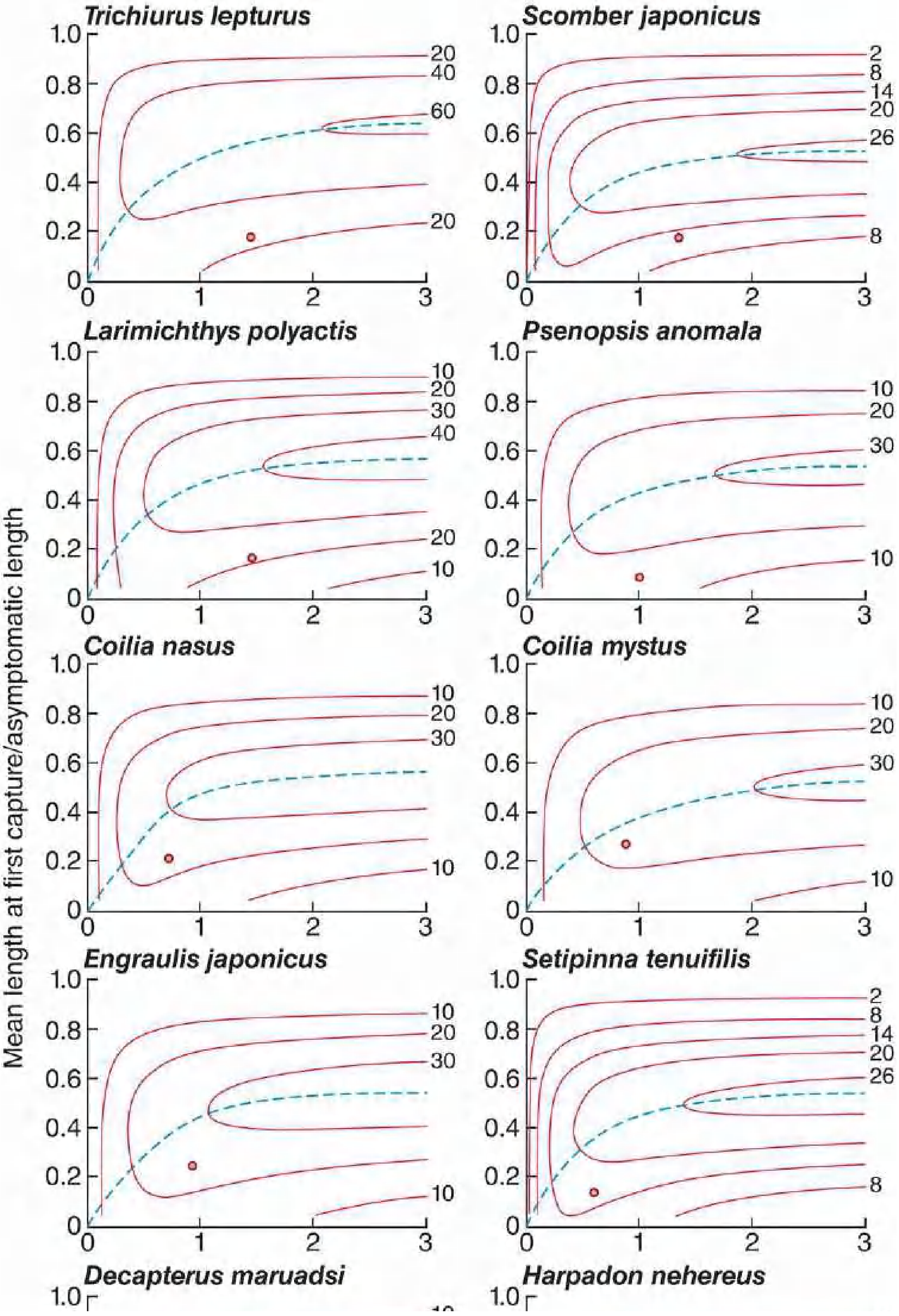
Source: After Pitcher & Hart (1982).

There are many variations on the general theme. For example. the submodels can be dealt with separately in each of the age classes and submodels can incorporate as much or as little information as is available or desirable (Shepherd & Pope, 2002a , 2002b). In all cases, though, the basic approach is the same. Available information (both theoretical and empirical) is incorporated into a form that reflects the dynamics of the structured population. This then allows the yield and the response of the population to different harvesting strategies to be estimated. This in turn should allow a recommendation to the stock manager to be formulated. The crucial point is that in the case of the dynamic pool approach, a harvesting strategy can include not only a harvesting intensity, but also a decision as to how effort should be partitioned amongst the various age classes, which for fisheries typically relates to the selectivity of the fishing gear, especially the mesh size of the nets that are used.

yield-per-recruit analyses of Chinese fish

Since dynamic pool models incorporate realities that surplus yield models ignore, it would be easy to say, simply, that they are to be preferred. On the other hand, the data required to use dynamic pool models are very often not available, or are available only in a partial or imperfect form. Surplus yield models, therefore, are still very much in use, and there is a wide variety of age-structured approaches that go beyond surplus yield models but fall short of a fully supported dynamic pool model – indeed, far too wide a range to be reviewed here (but see, for example, Haddon (2002)). We can, though, get an idea of the questions that can be addressed, by looking at a particular approach, a *yield-per-recruit analysis*. This seeks to determine how growth, natural mortality and fishing mortality interact to relate the numbers in the youngest age class (the recruits) to the exploited biomass from the fishery (the yield). Knowing this we can calculate not only what fishing effort, but also what minimum size of capture, would be most sustainably productive. The underlying principle is that short-term yield will increase with effort and with allowing few fish, even the smallest, to escape. Longer term, however, it may be better to increase mesh size, and perhaps to reduce effort, so as to allow fish to grow and reproduce more before they are caught.

Results from one example, for 10 marine fish species exploited in the seas around China, are shown in Figure 15.18 (Liang & Pauly, 2017). The study is instructive in illustrating how limited data can nonetheless be put to good use. No direct information was available for either the growth or mortality rates of the species. Nonetheless, ignoring details here, on the basis of reasonable assumptions it was possible to use length-frequency data from catches – that is, the proportions of the population in successive size classes – to estimate these growth and mortality rates.


Mean length at first capture/asymptomatic length
Trichiurus lepturus
20
40
60
20
Scomber japonicus
2
8
14
20
26
8
Larimichthys polyactis
10
20
30
40
20
10
Psenopsis anomala
10
20
30
10
Coilia nasus
10
20
30
10
Coilia mystus
10
20
30
10
Engraulis japonicus
10
20
30
10
Setipinna tenuifilis
2
8
14
20
26
8
Decapterus maruadsi
Harpadon nehereus
1.0
0.8
0.6
0.4
0.2
0
0
1
2
3

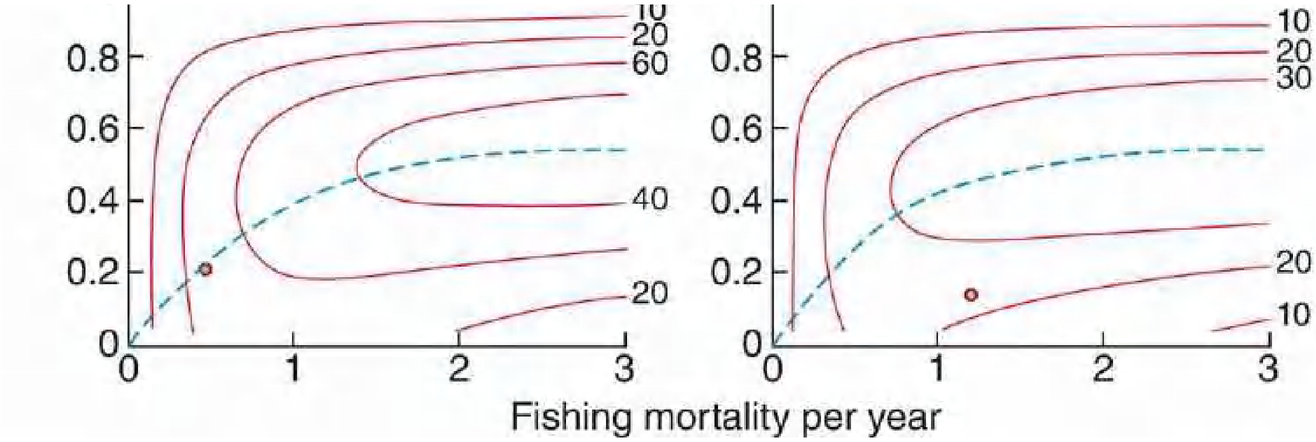


Figure 15.18 Peak (optimal) fishery yields around China are obtained at intermediate mesh sizes, but the mesh sizes used are typically smaller than this. Yield per recruit contours (grams) for 10 species of fish from the seas around China, as indicated (red lines), for different values of fishing mortality or effort (tonnes per year) and different values of mean length at first capture relative to asymptotic (fully grown) length, reflecting the mesh size. The blue dashed lines link optimal values for length at first capture. The red dots show the state of the fishery – the effort and mesh size – at the time of publication.

Source: After Liang & Pauly (2017).

Looking at Figure 15.18, it is apparent, first, that the mean length at first capture (effectively, mesh size) that maximises yield tends to increase with fishing effort, though with a clear levelling off. As fishing mortality is increased, it is necessary to allow fish to grow more before they are caught. Hence, for a given effort, yield peaks at an intermediate mesh size, allowing small fish to grow but not for so long that they are likely to die before they get caught. It is also clear that in almost every case the mesh sizes currently used are too small, catching too many small fish and generating smaller yields than should be possible. It is not surprising, therefore, that Liang and Pauly (2017) were able to point to a history of declines in the catches from many of these stocks, and in the average size of the fish being caught. Yield-per-recruit analyses like these are limited in the sense that fishermen are interested in *overall* yield, which obviously combines yield per recruit with the number of fish recruited – which itself depends on the stock-recruitment relationship (see Figure 15.16). In the present case, however, small mesh sizes are likely to compound the problems apparent in Figure 15.18, not only reducing the yield per recruit, but also removing fish from the population before they are best able to produce recruits of their own. Applied cautiously, these analyses can be an important diagnostic in assessing the state of a fishery.

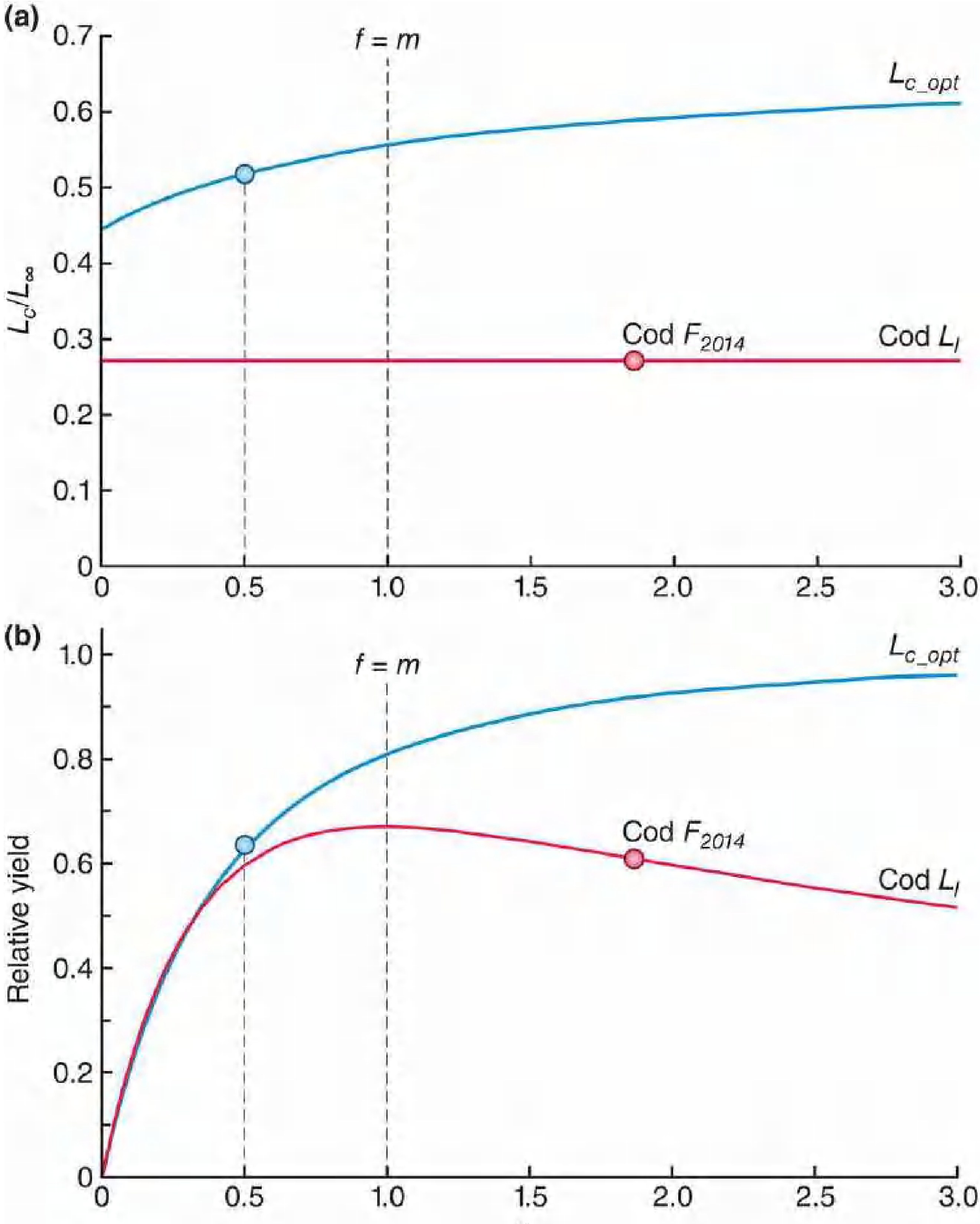
15.3.7 Rules of thumb for sustainable harvesting

More generally, the tension that inevitably exists between, on the one hand, the need for populations to be harvested sustainably, and on the other, the frequent lack of sufficient data on which to base a detailed strategy, makes it attractive to have relatively simple (but enlightened) rules of thumb that can be applied broadly, and in the absence of system-specific calculations. Froese *et al*. (2016) have proposed three such rules that each relate to principles we have already discussed here. There will always be those that object to what might be seen as the *over*simplification of rules of thumb, or to any particular set of rules, but an examination of the rules is nonetheless informative in suggesting a practical way of behaving responsibly in the absence of certainty. Their first rule is: *take less than nature* – that is, ensure that mortality as a result of fishing is less than the natural mortality rate. The basis for this comes from a simplified view of the yield-recruitment curves we saw, for example, in Figure 15.10 (fixed effort harvesting). There, the harvesting (or fishing mortality) rate at the MSY, H_{MSY}, will be equal to the natural mortality rate, m, if we assume that the net recruitment curve in the figure is symmetrical, as a result of a linear increase with abundance in the numbers dying set against an unchanging number being born. The rule, therefore, proposes, as we have seen previously, that the harvesting rate should never exceed H_{MSY}, and that in the absence of good data with which to estimate H_{MSY}, we can assume this to be equal to m. In fact, as Froese *et al.* emphasise, various meta-analyses have suggested that fisheries often suffer serious reductions in yield or changes in age structure at harvesting rates much less than H_{MSY}. Setting the harvesting rate to m should therefore be seen as an upper limit, best avoided, and setting it to $0.5m$ can be viewed as a more appropriate, precautionary target.

Their second rule is to *maintain populations at more than half their natural abundance*. This, too, relates to the net recruitment curves we discussed earlier and a simplifying assumption of symmetry. In fact, when curves are not symmetrical, the MSY often occurs at less than half the natural level, rather than more than half, so this rule, too, has a precautionary element.

Thirdly, they advocate that we should *let fish grow and reproduce*, specifically by aiming for a mean length in the catch that would maximise yields and profits. Without going into the calculations underlying the specifics of this recommendation, we can see that the rule reflects results like those in Figure 15.18, where yield was maximised by only catching fish once they had reached an intermediate, optimum size, and where this optimum could be estimated without the need for extensive catch data. This time, though, the emphasis is on designing fishing gear so as to generate not a minimum fish size, but a mean size in the catch, since this is easier to estimate and to enforce.

Some results of applying these rules of thumb to the fishery of the North Sea cod, *Gadus morhua*, are shown in Figure 15.19a. As in Figure 15.18, we see that as fishing effort increases, the mean catch size required to maximise yield also increases. Specifically, if fishing mortality, f, is equal to natural mortality, m (the ‘risky’ rule of thumb), then the mean size of fish in the catch needs to be 56% of their maximum size (72 cm); whereas if f is kept to $0.5m$ (the precautionary rule of thumb) then the mean size would need to be set to just 52% of the maximum (67 cm). Crucially, however, the 50% saving in effort gained from this halving of the fishing mortality, and the increased resilience of the fishery that would result, would give rise to only a 23% fall in yield (Figure 15.19b).



(a)
L_c/L_∞
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
$f = m$
L_{c_opt}
Cod F_{2014}
Cod L_l
0
0.5
1.0
1.5
2.0
2.5
3.0
(b)
Relative yield
1.0
0.8
0.6
0.4
0.2
0
$f = m$
L_{c_opt}
Cod F_{2014}
Cod L_l
0
0.5
1.0
1.5
2.0
2.5
3.0
f/m

Figure 15.19 Reducing mesh size to allow smaller North Sea cod to escape leads to a small drop in yield but to a larger drop in the effort expended and to greater fishery resilience. (a) The solid blue curve shows, for North Sea cod, the combination of length of first capture (L_c) relative to full-grown length (L_∞) and fishing mortality relative to natural mortality (f/m) that would give rise to an optimum mean length of the catch. Lengths when $f = m$ and $f = 0.5m$ are indicated. The horizontal red line is set for the 2014 length of first capture (L_l) and shows the state of the fishery at that time (cod F_{2014}). (b) Yield per recruit relative to the theoretical maximum (but unsustainable) yield for increasing values of f/m, for both the blue curve in (a), $L_{c\text{-}opt}$, and the 2014 fishery, L_l.

Source: After Froese *et al.* (2016)

.

By contrast, some sense of the actual health of the fishery can be gained by noting its position in Figure 15.19a in 2014 – with a fishing mortality rate almost twice the natural rate, and a minimum legal landing size (35 cm) way below that required to maximise yields even at the much lower recommended efforts. This suboptimal combination of fishing effort and mesh size, apart from the cost of the high fishing effort and the threat this poses to long-term sustainability, gives rise to a yield lower than that predicted when f is $0.5m$ (Figure 15.19b). Even these simple rules, therefore, can guide a fishery away from current practices towards an attractive, balanced combination of safety and acceptable short-term yields and profits.

15.3.8 Ecosystem-based fisheries management?

fishing down marine food chains

As we close this section on the abundance of single species, it is appropriate that we look briefly at the management of fisheries from a broader, more multispecies perspective (in later chapters the focus will turn more to whole communities and ecosystems). In fisheries management, as elsewhere in ecology, nobody doubts that focusing on single species, or even pairs of species, is a matter of scientific convenience rather than a reflection of an underlying reality, but the importance of a broader perspective was brought sharply into focus by Pauly *et al.*'s (1998) analysis of the effect of commercial fishing on the mean trophic level (MTL) of fisheries landings. We discuss trophic levels, food chains and food webs in more detail in Chapter 17, but the underlying ideas are simple and widely known. In marine systems in particular, algae occupy trophic level 1 (TL1), on which all levels ultimately depend for food. They are preyed upon by herbivorous zooplankton in TL2, which are preyed upon by larger zooplankton and small fishes in TL3, and these in turn are preyed upon by larger fish (and some invertebrates), which, depending on the precise mix within their diet, may be classified as anything from TL3.5 (feeding on a mixture of planktivorous and piscivorous fish) to TL4.5 and beyond. Unlike terrestrial systems, there is almost always an increase in size as we move up these marine food chains. An example of Pauly *et al.*'s (1998) analysis is shown in Figure 15.20a.

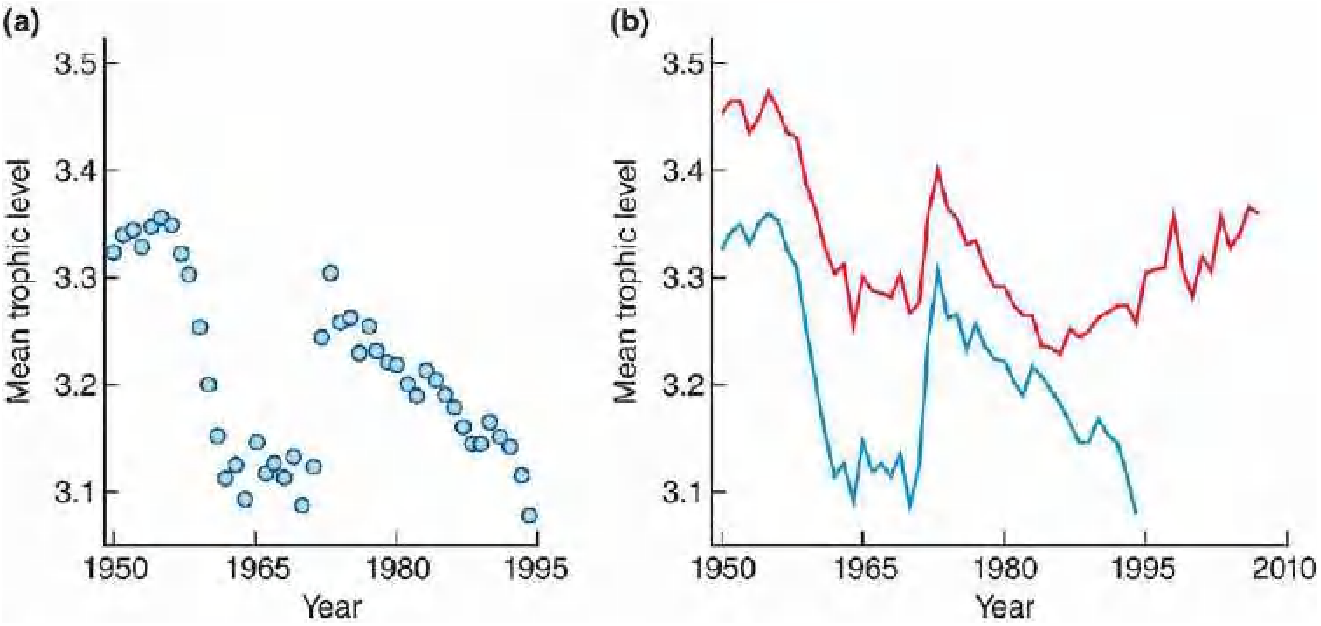


Figure 15.20 **The mean trophic level (MTL) of fisheries catches have appeared to decline over time but this seems not to be a general rule.** (a) Changes in the estimated MTL of marine fisheries landings globally, 1950–94. (b) The blue line shows the same data as in (a). The red line shows a reanalysis of those data with MTL re-estimated, and with the dataset extended to 2009.

Source: (a) After Pauly *et al.* (1998). (b) After Branch *et al.* (2010).

The figure shows an apparent decline in the MTL of catches between 1950 and the mid-1990s, especially obvious from the early 1970s. Pauly *et al.* attributed this trend to fishermen turning to smaller species, lower down the food chain, as stocks of larger top predators were increasingly depleted by overfishing – referred to as ‘fishing down the food web’ (or down the food chain). The reality and meaning of this trend was questioned almost immediately (Caddy *et al.*, 1998), but it quickly attracted widespread credence, to the extent that in 2000, the United Nations-conceived Convention on Biological Diversity adopted the MTL (sometimes called the Marine Trophic Index) as one of eight indicators of ecosystem health, applied widely to assess the state of the marine environment (Branch *et al.*, 2010).

Certainly the initial analysis was valuable, at least, as an acknowledgement that fishing, and especially overfishing, is likely to have consequences extending far beyond the target species itself. It is valuable, too, where possible, to have simple indices of ecosystem health, allowing different sites and systems to be compared and trends to be followed. However, it has become apparent, first, that ‘fishing down’ is not the only force acting to depress the catch MTL. For example, ‘fishing through’ the food web, where fisheries at lower trophic levels are sequentially added to the portfolio without the collapse of those at higher levels, has very similar effects on MTL, with one major survey concluding that fishing down had occurred in nine ecosystems, but fishing through in 21 (Essington *et al.*, 2006). It has also become apparent that a decline over time in catch MTL is neither inevitable nor necessarily an indication of declining biodiversity. Indeed, Essington *et al.* (2006) found that ‘fishing up’, moving on to higher trophic levels, had occurred in 18 of the ecosystems they analysed; while Figure 15.20b shows a reanalysis of the data in Figure 15.20a, with 15 years of additional data and with all data updated to include reassessments of the trophic levels occupied by the different species. There have, indeed, been periods of catch MTL decline, but since the mid 1980s the overall trend appears to have been upwards.

Moreover, Figure 15.21 shows a range of examples in which time series have been traced for the MTL in catches, but also, where available, for MTL in the marine community as a whole, as assessed by surveys (trawls carried out to provide consistent representative samples) and by compilations from all available data sources. It is, after all, the health of the wider community, not the contents of the catch, that should be our most fundamental concern. We can see, first, that there is no general rule of declining catch MTL over time. Indeed, in several, increases are apparent, especially since the mid-1980s. But in addition, of those fisheries where correlations were possible, catch MTL was negatively correlated with survey MTL in 13 of 29 cases, and negatively correlated with compilation MTL in four of nine cases. Not surprisingly, therefore, these negative correlations were also found when data were combined from across the different fisheries. This casts doubt, at the very least, on the uncritical use of catch MTL as a measure of ecosystem health, and allied to the range of fisheries' strategies that might translate into changes in catch MTL, argues for an approach that focuses on the underlying biological community (not assuming catch composition is a reliable proxy) and that monitors carefully the responses of fisheries' behaviour to diminishing stocks and financial pressures (Branch *et al.*, 2010). Given the difficulties of obtaining enough reliable data even for single species, adopting and applying this whole-community approach represents a formidable challenge, but one that we should be prepared to take up.

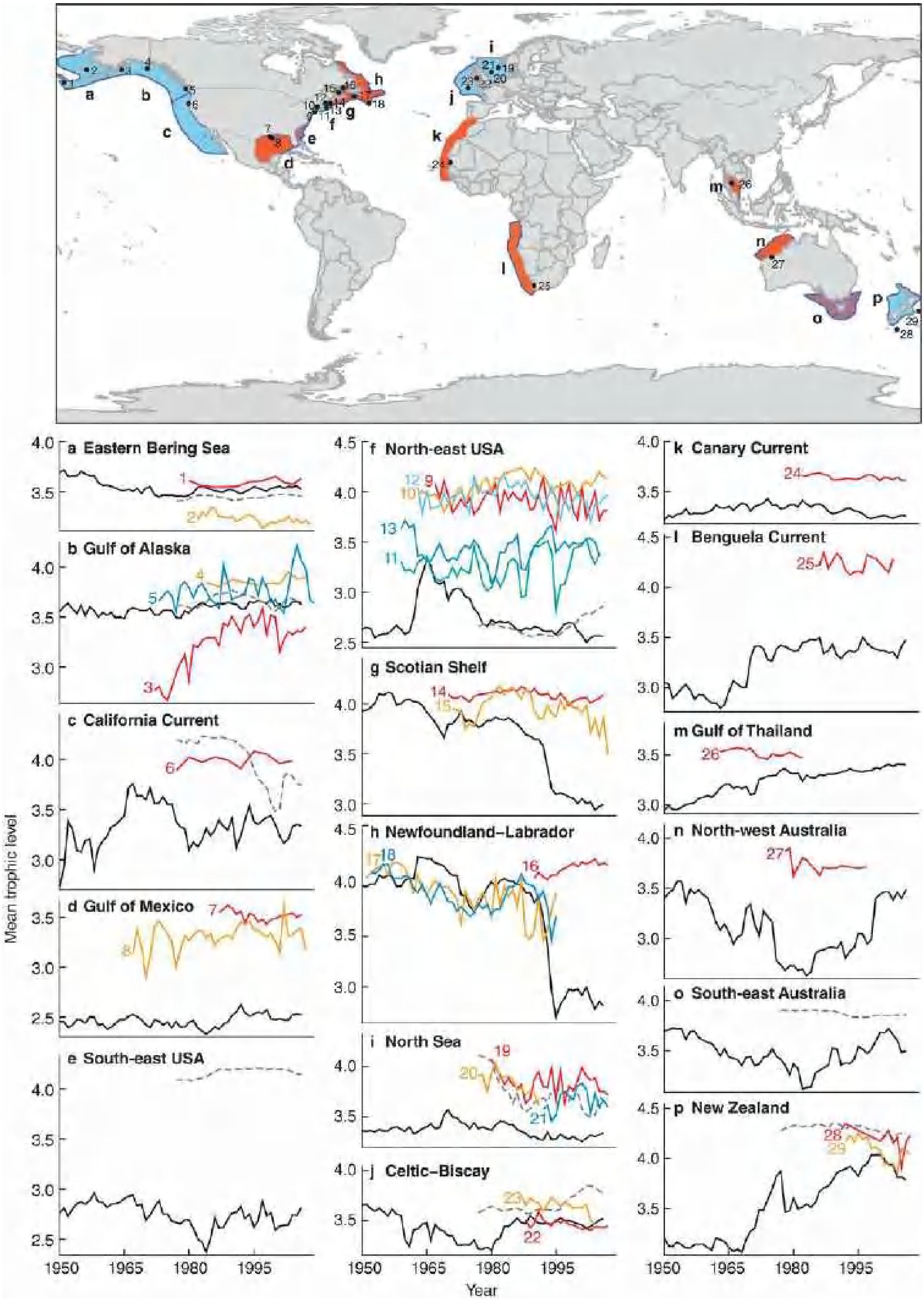

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
a Eastern Bering Sea
b Gulf of Alaska
c California Current
d Gulf of Mexico
e South-east USA
f North-east USA
g Scotian Shelf
h Newfoundland–Labrador
i North Sea
j Celtic–Biscay
k Canary Current
l Benguela Current
m Gulf of Thailand
n North-west Australia
o South-east Australia
p New Zealand
Mean trophic level
Year
1950
1965
1980
1995
2.5
3.0
3.5
4.0
4.5

Figure 15.21 Mean trophic level (MTL) in catches shows a variety of trends in different fisheries around the world, both over time and in relation to estimates for the fishery community as a whole. MTLs estimated for the ecosystems of 16 fisheries, as indicated. MTLs from catches are shown in black, from surveys in various colours, and from compilations from various sources in grey. The map shows locations with data from all three sources (blue), from catches and surveys (red) and from catches and assessments (purple). Numbers refer to the location of each survey.

Source: After Branch *et al.* (2010).

15.4 Conservation ecology

15.4.1 Introduction

We have discussed the processes by which new species are generated (Section 1.3), by which they may migrate to new areas (see Chapter 6), and by which their abundance may be determined (see especially Chapter 14). We also know that species may become extinct, either locally or globally. Regrettably, current rates of extinction greatly exceed those of species creation or immigration (Figure 15.22). *Conservation* describes the various actions we can take to slow down or even reverse these losses of species and, in the case of collections of species, biodiversity. Many of the Applications in previous chapters have touched on aspects of species conservation, but here we take a more concerted view. Although the topics are closely connected, we focus more on biodiversity conservation in Chapter 19.

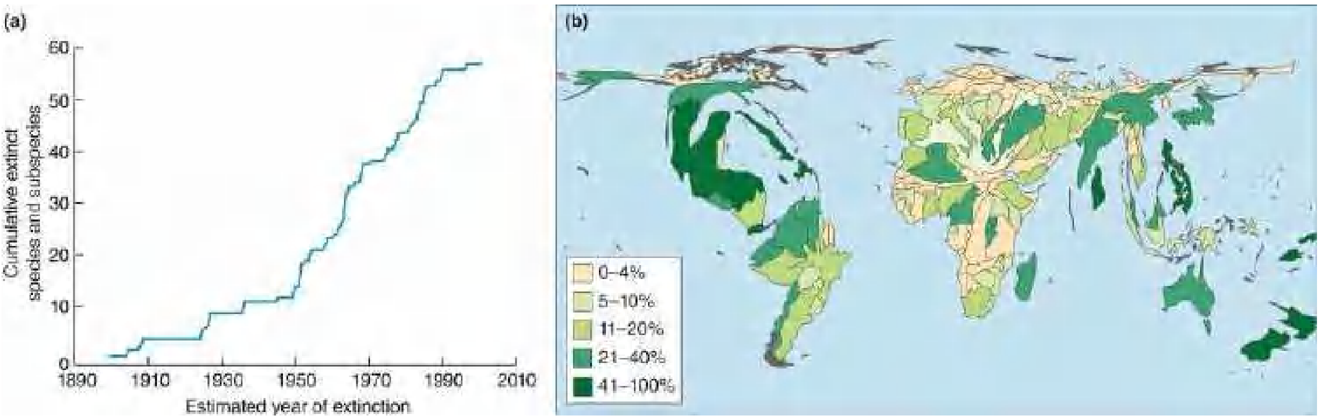


Figure 15.22 Freshwater fish in North America and amphibians worldwide illustrate the rapid loss of species currently. (a) The cumulative loss of species (39) and subspecies (18) of freshwater fish in North America since 1900. (b). A map of the world in which countries are distorted in proportion to the number of amphibian species living there, relative to their size. Then, countries are coded according to the percentage of amphibian species threatened with extinction (either 'vulnerable', endangered' or 'critically endangered' – see Figure 15.24).

Source: (a) After Burkhead (2012). (b) After Wake & Vradenburg (2008).

what is conservation?

In some cases, conservation efforts are focused on individual species and hence on biodiversity only indirectly. In others there is a more explicit focus on whole habitats and so on biodiversity itself. In either case, what are the options available to us? Broadly, we can either protect or restore. Among protective measures, we can simply fence off areas to keep out whatever the threat is, or we can manage an area (protect a food plant, eliminate hunting or disturbance) to reduce or even eliminate that threat. Alternatively, we may relocate individuals from an area where the species is thriving (relatively speaking) to one where it is now extinct, or breed individuals in an artificial setting before reintroducing them, or even recreate whole ecosystems, especially in areas where all or most of the habitat has been destroyed (for example by logging or mining). We can think of these cases of restoration as conservation in a broader sense, since the aim is to reverse a past failure to conserve.

how many species on Earth?

Ideally, to judge the scale of the problems we face, we need to know the total number of species currently, the rate at which these are going extinct, and how this rate compares with that of prehuman times. There are considerable uncertainties in our estimates of all three. About 1.8 million species have so far been named (Figure 15.23), but the real number must be much larger. Estimates have been derived in a variety of ways. One approach, for example, uses information about the rate of discovery of new species to project forward, group by taxonomic group, to a total estimate of up to 6–7 million species in the world. However, the uncertainties in estimating global species richness are profound, and our best guesses range from 5–10 million eukaryotes, plus the bacteria and viruses, with defensible numbers overall ranging from 3 to 100 million (May, 2010).

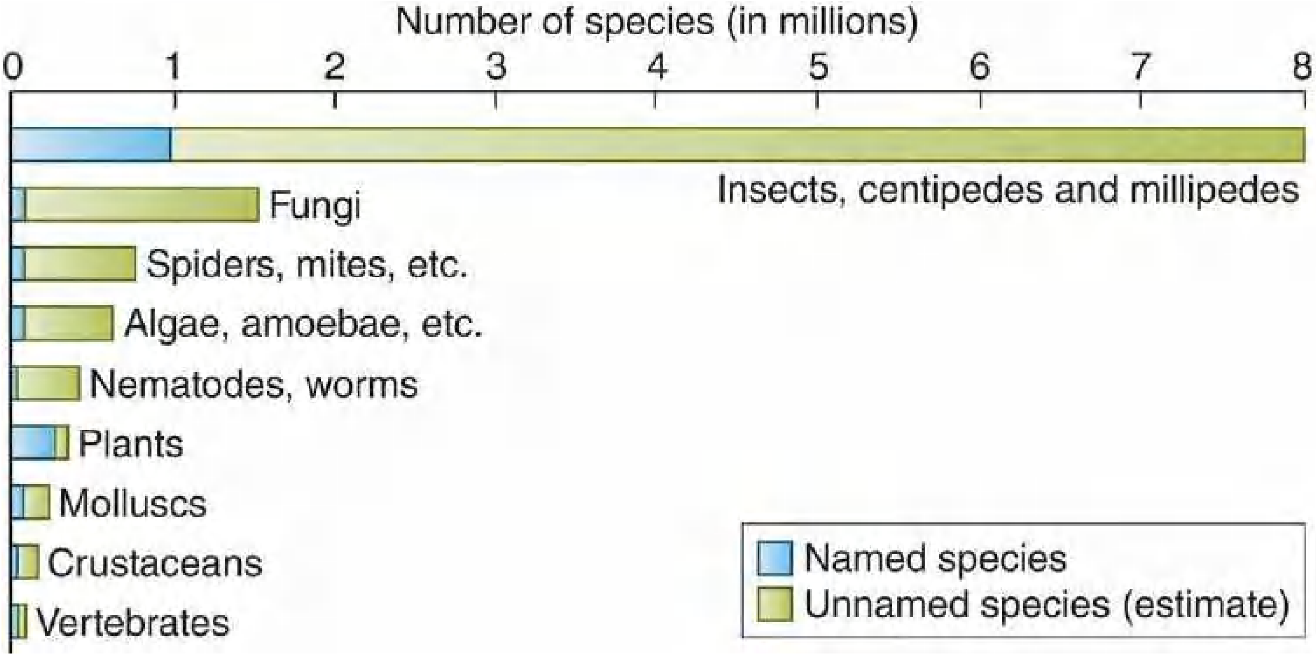


Figure 15.23 Most of the estimated 10 million-plus species of eukaryotes remain unnamed. Numbers of species identified and named are shown in blue and estimates of unnamed species that exist are shown in green.

Source: After MEA (2005).

modern and historical extinction rates

An important lesson from the fossil record is that the vast majority of species eventually become extinct – more than 99% of species that ever existed are now no longer with us. However, given that individual species are believed, on average, to have lasted about 1–10 million years, and if we estimate conservatively that the total number of species on Earth is 10 million, then we would predict that on average, only between 100 and 1000 species (0.001–0.01%) would go extinct each century. The current observed rate of extinction of birds and mammals is about 1% per century, 100–1000 times this ‘natural’ background rate (see also Figure 15.22). The evidence, then, while based only on estimates, suggests that our children and grandchildren may live through a period of species extinction comparable with the ‘natural’ mass extinctions evident in the geological record, when, for example, many whole families of shallow-water invertebrates were lost some 250 million years ago towards the end of the Permian period, and around 96% of the species then living may have gone extinct.

what is the value of conservation? – ecosystem services

But should we care? To most, the answer is an unhesitating 'yes'. It is nonetheless important to consider why we should care, that is, why the continued existence of individual species, and of biodiversity as a whole, is valuable. To do so we must elaborate the idea of *ecosystem services* (MEA, 2005a), to which individual species contribute. To reiterate, ecosystem services are functions or attributes provided by ecosystems (natural or managed) in support of human interests, generally saving a cost that would otherwise need to be paid. The concept provides a counterpoint to the purely market-driven economic reasons that often justify our degradation of nature through pollution, land use, and other habitat destruction. A major challenge is the development of a new, more sustainable economics aimed at delivering both financial and ecological stability – and perhaps a 'prosperity' based on more than simple monetary wealth (Jackson, 2017).

We can divide ecosystem services into several components.

- *Provisioning services* include wild foods such as fish from the ocean and berries from the forest, medicinal herbs, fibre, fuel and drinking water, the pollination of crops by bees, the biological control of pests, as well as the products of cultivation in agro-ecosystems. We can also include the monetary value of ecotourism, billions of pounds each year worldwide, as a provisioning service, since it depends directly on the existence of individual species and the biodiversity of which they are part.
- Nature, though, also contributes the *cultural services* of aesthetic fulfilment and educational and recreational opportunities – the well-being they bring to individuals.
- *Regulating services* include the ecosystem's ability to break down or filter out pollutants, the moderation by forests and wetlands of disturbances such as floods, and the ecosystem's ability to regulate climate (via the capture or 'sequestration' by plants of the greenhouse gas carbon dioxide).
- Finally, and underlying all the others, there are *supporting services* such as primary production, the nutrient cycling upon which productivity is based, and soil formation.

Many people also point to ethical grounds for conservation, with every species being of value in its own right – a value that would still exist even if people were not here to appreciate or exploit the natural world. From this perspective even species with no conceivable economic value require protection.

It would be wrong, though, to see things only from the point of view of conservation. Not that there are really arguments *against* conservation as such. But there are arguments in favour of the human activities that make conservation a necessity: agriculture, the felling of trees, the harvesting of wild animal populations, the exploitation of minerals, irrigation, the discharge of wastes, and so on. To be effective, therefore, the arguments of conservationists must ultimately be framed in cost–benefit terms, because governments will always determine their policies against a background of the money they have to spend and the priorities accepted by their electorates.

15.4.2 Small populations

the classification of risk

When conservation biologists are focused on individual species, those species are almost inevitably rare, and that rarity contributes to the risk of extinction. But how do we define the risk that a species faces? By far the most widely adopted categorisation (Mace *et al.*, 2008) is one proposed by the International Union for Conservation of Nature and Natural Resources (IUCN, 2012). A species may be placed into a particular category by satisfying one of several sets of criteria – some including rates of decline, some including habitat range, some including abundance estimates, and some based directly on estimated extinction risks. This is necessary because different types and amounts of information are available for different species. The range of categories can be illustrated by focusing on extinction risk, where a species is described as *critically endangered* if there is considered to be more than a 50% probability of extinction in 10 years or three generations, whichever is longer (Figure 15.24); *endangered* if there is more than a 20% chance of extinction in 20 years or five generations; *vulnerable* if there is a greater than 10% chance of extinction in 100 years; *near threatened* if the species is close to qualifying for a threat category or judged likely to qualify in the near future; or *of least concern* if it does not meet any of these threat categories (Rodrigues *et al.*, 2006). Based on the these and the related criteria, 13% of bird species, 25% of mammals, 33% of reef building corals, 41% of amphibians and 63% of cycads (palm-like tropical plants) are threatened with extinction, being critically endangered, endangered or vulnerable (IUCN, 2015).

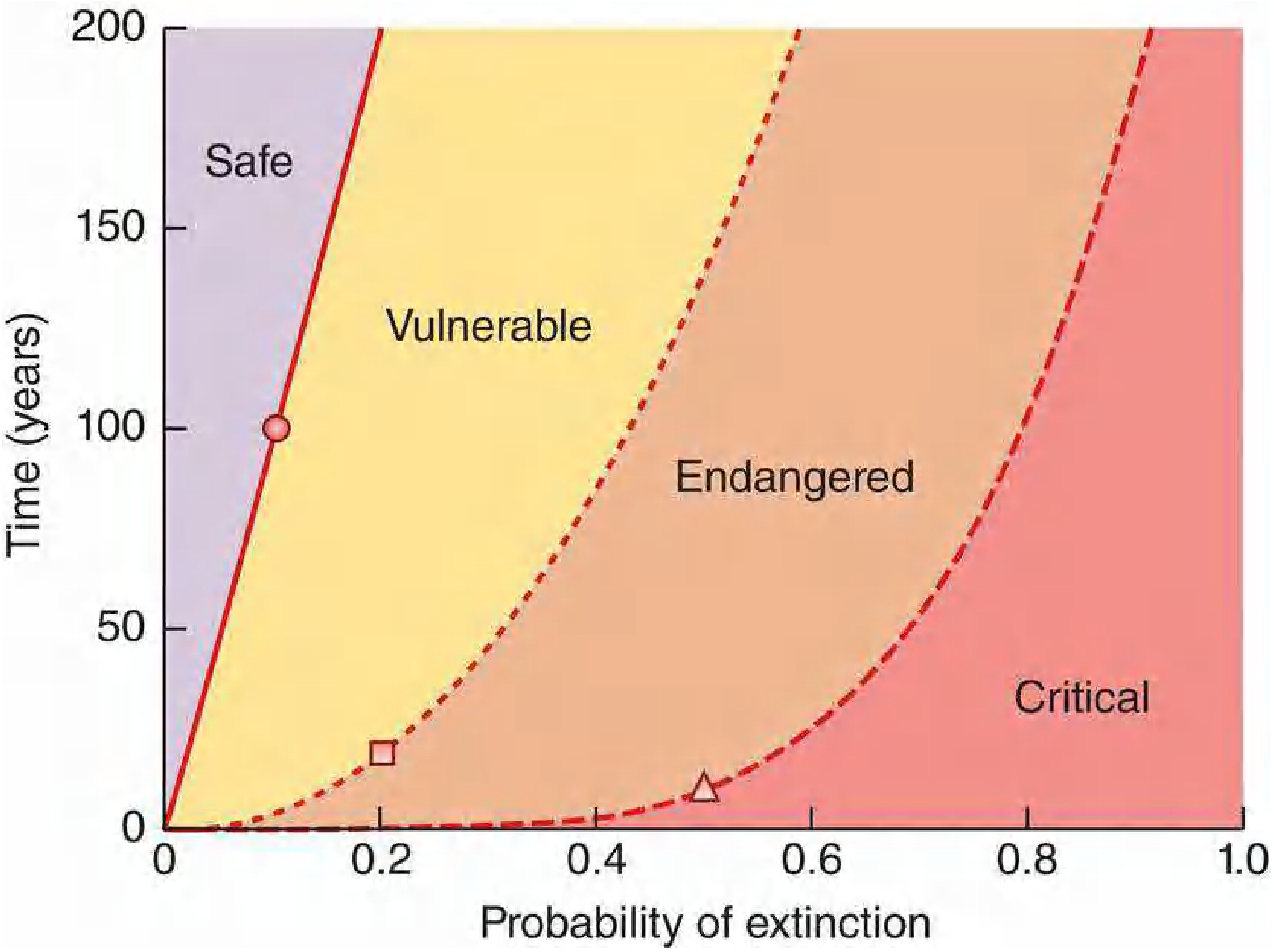


Figure 15.24 The categorisation of a species' risk of extinction changes as its probability of extinction varies as a function of time. The red circle represents a 10% probability (0.1) of extinction in 100 years (minimum criterion for a population to be designated 'vulnerable'). The square represents a 20% probability of extinction in 20 years (minimum criterion for the designation 'endangered'). The triangle represents a 50% probability of extinction in 10 years (minimum criterion for the designation 'critically endangered').

Source: After Akçakaya (1992).

demographic risks associated with small populations

Every one of the IUCN criteria makes reference either to how small a population is, or to a proxy of population size (like area occupied), or to a consequence of population size (like extinction risk). There is thus a pressing need to understand the dynamics of small populations. These are governed by a high level of uncertainty. *Demographic uncertainty* is the term we apply to random variations in the number of individuals that are born male or female, or in the number that happen to die or reproduce in a given year, or in the genetic quality of the individuals in terms of survival/reproductive capacities – all features that can be of crucial importance to the fate of small populations. Suppose a breeding pair produces a clutch consisting entirely of females. Such an event would go unnoticed in a large population but could be the last straw for a species down to its last pair. *Environmental uncertainty*, then, refers to unpredictable changes in environmental factors, whether disasters (such as floods, storms or droughts of a magnitude that occurs very rarely) or relatively minor events (year-to-year variation in average temperature or rainfall). These can also seal the fate of a small population. The need, frequently, to incorporate these uncertainties into stochastic population models was discussed in Section 5.6.6.

As an illustration, take the demise in North America of the heath hen (*Tympanuchus cupido cupido*). This bird, once extremely common from Maine to Virginia, was highly edible, easy to shoot, susceptible to introduced cats, and affected by conversion of its grassland habitat to farmland. It is perhaps not surprising then that by 1830 it had disappeared from the mainland and was found only on the island of Martha's Vineyard. In 1908 a reserve was established for the remaining 50 birds, and by 1915 the population had increased to several thousand. But 1916 was a bad year. Fire (a disaster) eliminated much of the breeding ground, there was a particularly hard winter coupled with an influx of goshawks (environmental uncertainty), and finally poultry disease arrived on the scene (another disaster). At this point, the remnant population was likely to have become subject to demographic uncertainty. For example, of the 13 birds remaining in 1928, only two were females. A single bird was left in 1930, and the species went extinct in 1932.

genetic problems in small populations

Further problems may arise in small populations through loss of genetic variation. Genetic variation is determined primarily by the joint action of natural selection and genetic drift (where the frequency of genes in a population is determined by chance rather than evolutionary advantage). The relative importance of genetic drift is higher in small isolated populations that, as a consequence, are expected to lose genetic variation. The rate at which this happens depends on the *effective population size* (N_e), rather than simply on the number of individuals present (N). Effective population size is the size of the ideal population to which the actual population is equivalent in genetic terms, where 'ideal' means the sex ratio is 1 : 1, the distribution of numbers of offspring among parents is random, and the population size remains constant. With unequal sex ratios, most progeny produced by relatively few individuals, and varying population size, all commonplace, N_e is usually less, and often much less than N (Lande & Barrowclough, 1987). With the advent of affordable molecular tools for obtaining detailed genetic data from natural populations, effective population sizes are increasingly being estimated from the signatures they leave in the population's genetic architecture (Gilbert & Whitlock, 2015).

Genetic diversity is important in the first place because of the long-term evolutionary potential it provides. Rare forms (*alleles*) of a gene, or combinations of alleles, may confer no immediate advantage but could turn out to be well suited to changed environmental conditions in the future. Small populations tend to have less variation and hence lower evolutionary potential. A more immediate potential problem arises because when populations are small, related individuals tend to breed with one another. All populations carry recessive alleles that can be harmful, even lethal, to individuals when homozygous. Individuals that breed with close relatives are more likely to produce offspring that receive harmful alleles from both parents. The deleterious effects that result are known as *inbreeding depression*. There are many examples of inbreeding depression – breeders of domesticated animals and plants, for example, have long been aware of reductions in fertility, survivorship, growth rates, and resistance to disease, and they commonly seek to prevent close relatives breeding with one another.

inbreeding depression in the Glanville fritillary

One study that has demonstrated the importance of genetic effects in small populations is that of the Glanville fritillary butterfly, *Melitaea cinxia*, on the Åland islands off the coast of Finland that we examined in detail in Section 6.7.4. Saccheri *et al.* (1998) related the probability of extinction of populations of the butterfly between late summer 1995 and late summer 1996 both to ecological variables that had previously been established as significant determinants of extinction – for example, population size, area occupied and so on – and to the level of genetic variability in each population: the average number of heterozygous loci amongst seven that were tested. As expected, the ecological variables were important, accounting, between them, for 56% of the total deviance in a statistical model fitted to account for whether or not populations went extinct. Indeed, the smallest populations tended to be the least diverse. However, including genetic variability itself accounted for a further 26% of the deviance (Figure 15.25a): the least variable populations were the most likely to go extinct.

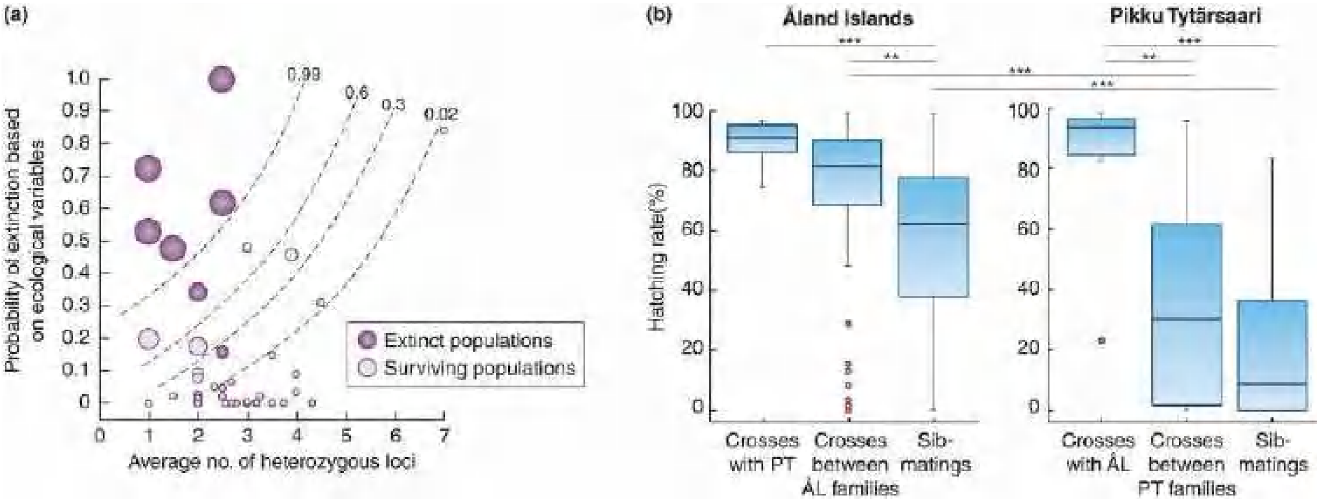


Figure 15.25 Populations of Finnish butterflies are more likely to go extinct when they have lower genetic diversity, which they do in smaller populations, because the individuals in such populations have lowered fitness as a result of inbreeding depression. (a) For 42 populations of the Glanville fritillary butterfly, *Melitaea cinxia*, in the Åland islands, Finland, each dot shows its estimated probability of extinction based on ecological parameters and the average number of heterozygous loci (out of seven) in the population. The size of each dot represents the probability of extinction when ecological and genetic determinants are combined. The lines are isoclines of extinction risk, as indicated. (b) Egg hatching rate for the butterflies in three types of crosses, as indicated, for populations from the Åland islands (ÅL) and Pikku Tytärsaari (PT). Boxes and whiskers indicate medians, first and third quartiles, minima and maxima, and outliers. The lines above indicate significant differences between pairs: $^{**}P < 0.01$, $^{***}P < 0.001$.

Source: (a) After Saccheri *et al*. (1998). (b) After Matilla *et al*. (2012).

A later study then investigated inbreeding depression in a small isolated population of the butterfly (N_e approximately 100) on the island of Pikku Tytärsaari (PT), 400 km from the Åland islands (Matilla *et al*., 2012). The estimated genetic diversity on PT (heterozygosity over all loci) was 0.37 compared with 0.62 for the Åland island population, but most significantly in the present context, this was reflected in individuals' fitness from the different populations. For example, egg hatching rate was threateningly low when PT individuals were mated with their own sibs, and even when they were mated with individuals from other PT families, but rebounded to almost 100% when they were outcrossed to individuals from other regions, including Åland (Figure 15.25b). The same pattern was apparent for the Åland island population, but the rates were nothing like as low. We can see, therefore, how small populations lack genetic diversity, increasing the probability of individuals being homozygous for deleterious mutants, reducing their fitness, and threatening the whole population with extinction.

recovery of the pink pigeon

It is, though, possible to reduce and perhaps even solve the genetic problems in small populations. The pink pigeon (*Columba mayeri*) provides a good example. Once widespread on the island of Mauritius, it had declined to a population of only nine or 10 birds by 1990. Other animals had been maintained in captivity, but that captive population was originally descended from just 11 founder individuals, augmented in 1989–94 by adding 12 more (offspring of the remaining wild individuals). The aim, therefore, was to manage matings in captivity, so as to retain high levels of genetic diversity and minimise inbreeding, and then release offspring back into the wild. Between 1987 and 1997, scientists reintroduced 256 birds on Mauritius, wherever possible selecting birds with minimal inbreeding (based on records in breeding 'stud books') and releasing them in groups with good representation of the different founder ancestries. As a result, the population had swelled to 355 free-living individuals by 2003 (plus more in captivity) and has remained at or above that level since. Furthermore, preoccupation with avoiding inbreeding has been justified by studies in which the genetics and ecological success of both captive and wild pink pigeon populations have been carefully monitored. Inbreeding reduced egg fertility and survival of nestlings (Figure 15.26), but effects were strongly marked only in the most inbred birds.

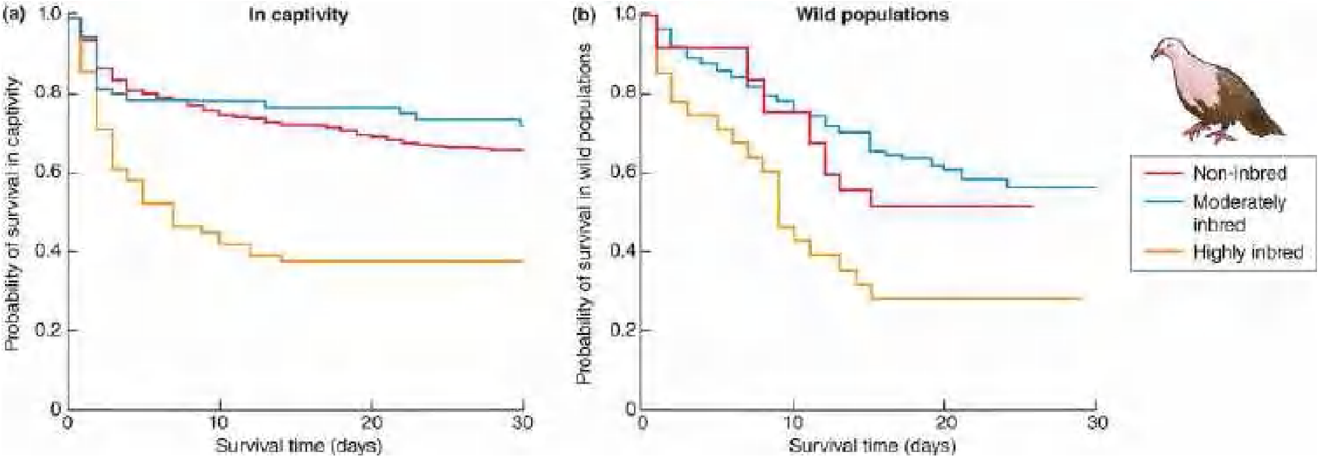
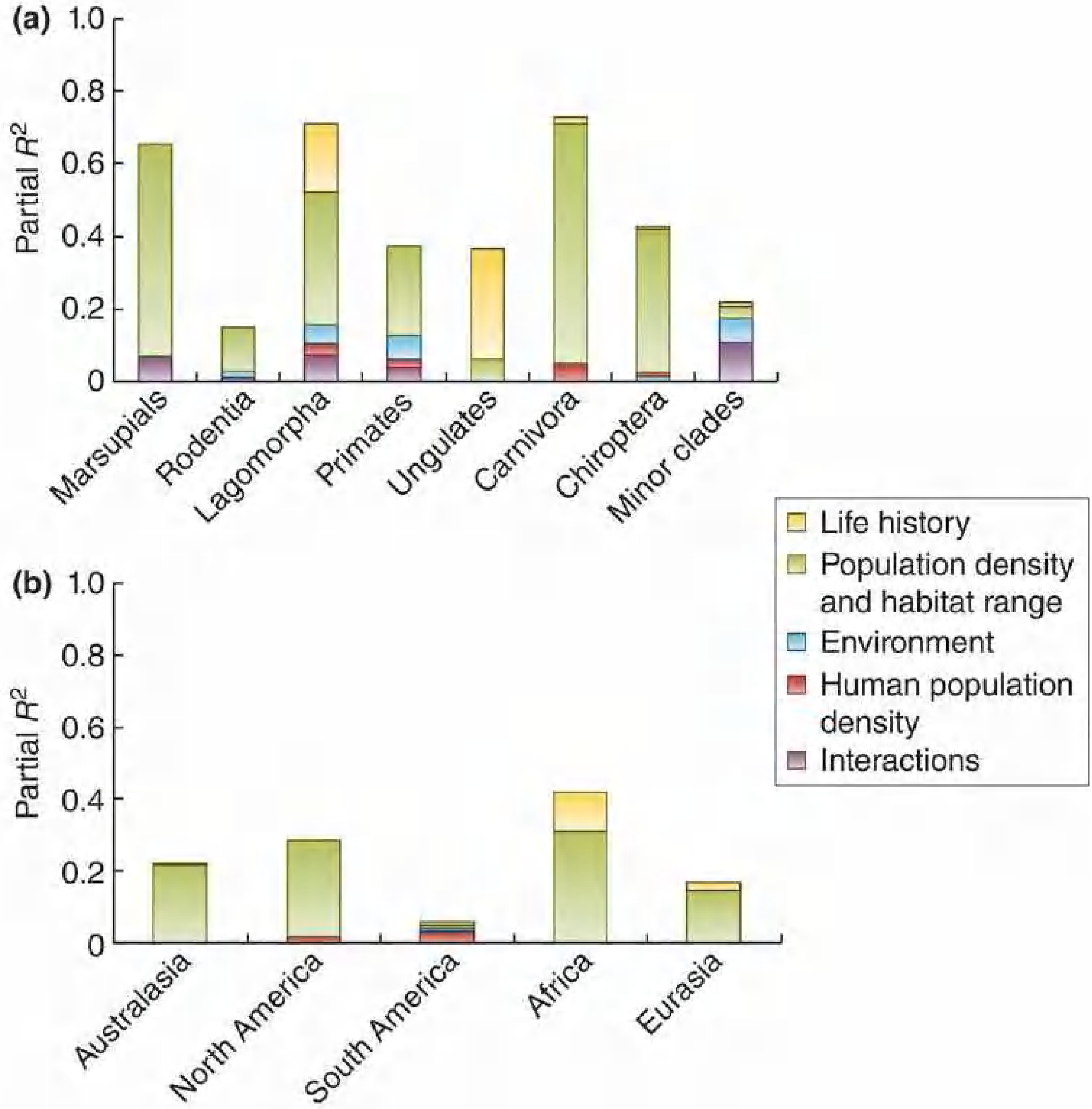


Figure 15.26 Highly inbred pink pigeons from Mauritius have significantly reduced survival. Effect of inbreeding on probability of survival to 30 days of age of pink pigeon nestlings (a) in captivity and (b) in the wild population. Inbreeding is expressed as an index derived from known ancestry in relationship to 23 founder individuals.

Source: After Swinnerton *et al*. (2004).

habitat reduction

Both demographic and genetic factors, then, can make small populations more vulnerable to extinction. Generally, though not of course invariably, species suffering a reduction in the area they are able to occupy will, as a consequence, also suffer a reduction in abundance. It is perhaps not surprising, therefore, that an analysis of species-level extinction risk across the mammals found that in the best statistical model, population density and especially geographical range size were the most consistently powerful predictors of risk both across different mammal groups and across different continents (Figure 15.27).



[Figure 15.27](#) **Population density and habitat range are the most powerful predictors of extinction risk in mammals** (a) The partitioning of variance in extinction risk in mammals among four different types of predictors (and interactions between them) when separate models were fitted to different taxonomic groups, as indicated. (b) The equivalent analysis when the data are broken down by the continent on which the study was carried out. 'Life history' includes mass, litter size, litters per year, etc. 'Environment' includes precipitation, transpiration and latitude.

Source: After Cardillo *et al*. (2008).

15.4.3 Causes of extinction

Clearly, populations on the way to extinction are small, whether they travel that road all the way to its destination, or are stalled, or even turned back. But what sets a population on this path? We turn next to some of the more important factors that threaten the future of species populations and hence of biodiversity more generally.

overexploitation: the case of elephant ivory

We discussed overexploitation as an integral part of the managed exploitation of populations throughout the previous Section. Such problems are not new. In prehistoric times, it seems likely that humans were responsible for the extinction of many large animals, the so-called megaherbivores, by overhunting them, and in much the same way, populations of the great whales and many commercial fish species are in danger of following a similar path.

In many cases, as we have seen, the endangered species is being exploited for food. Often, though, the 'food' value of the species is exotic and imaginary (for instance, the supposed medicinal powers of rhino horn in some cultures), or the species are exploited for 'aesthetic' reasons, whether for their body parts or as exotic pets. In these cases, their value to collectors goes up as they become rarer. Thus, instead of the normal safeguard of a density-dependent reduction in consumption rate at low density (see Chapter 9), increased rarity may actually increase the rate of exploitation, unless hunting and other forms of removal are strictly regulated.

Elephant ivory is a good example. African elephants have been exploited for ivory for many years, and as they have become rarer, the resource itself has become more valuable. For example, the estimated numbers of elephants in Africa was 1.3 million in 1979 but only 600 000 in 1989. Over the same period, the value of uncarved ivory in Kenya was $2.5 per pound in 1969, $34 per pound in 1978 and more than $90 per pound in 1989 (Lemieux & Clarke, 2009). Between 1979 and 1989, the two countries suffering the greatest reduction in elephant numbers were the Democratic Republic of Congo and Tanzania (Figure 15.28a). The consequent rarity inevitably increased the value of what remained, so conservation of elephants would have been possible only with well-resourced and well-organised control of elephant hunting (poaching).

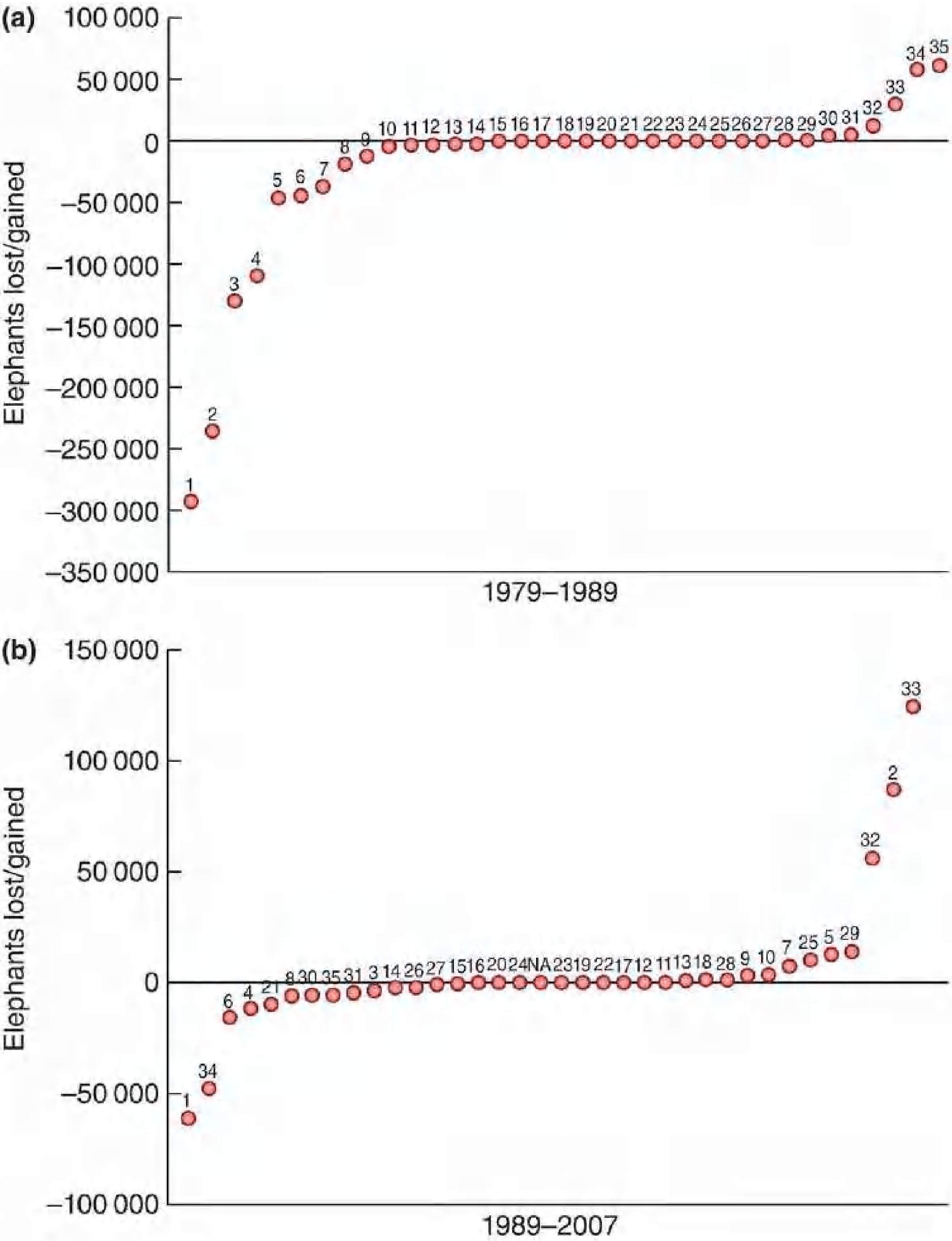

(a)
Elephants lost/gained
100 000
50 000
0
−50 000
−100 000
−150 000
−200 000
−250 000
−300 000
−350 000
1979–1989
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
(b)
Elephants lost/gained
150 000
100 000
50 000
0
−50 000
−100 000
1989–2007
1
34
6
4
21
8
30
35
31
3
14
26
27
15
16
20
24
NA
23
19
22
17
12
11
13
18
28
9
10
7
25
5
29
32
2
33

Figure 15.28 Loss of elephants to ivory poachers can be successfully resisted, as in Tanzania, but inaction can see further deterioration, as in the Democratic Republic of the Congo. (a) Changes in the abundances of elephants in African countries for which data were available, 1979–89, with countries ranked from greatest fall to greatest rise. ‘1’ is the Democratic Republic of Congo; ‘2’ is Tanzania. (b) Corresponding changes for 1989–2007, but with rank numbers retained from (a), plus one additional country (NA).

Source: After Lemieux & Clarke (2009).

From the 1990s, Congo has suffered repeated civil conflict, had one of the highest corruption perception indexes in Africa, and has supported an unregulated ivory market. With no way to control elephant poaching, the increasing value of ivory drove elephant numbers still lower (Figure 15.28b). By contrast, Tanzania had no civil conflict, a low level of corruption, and no unregulated ivory market. It managed to jump from the second largest fall to the second largest rise in elephant numbers in the period leading to 2007 (Figure 15.28b). The rarity value vicious circle is a problem, but one that can apparently be overcome.

habitat destruction, degradation and disturbance

Habitats can be adversely affected by human influence in three main ways. Firstly, a proportion of the habitat available to a particular species may simply be *destroyed*, for urban and industrial development, or for the production of food and other natural resources such as timber. Of these, forest clearance has been, and still is, the most pervasive type of habitat destruction (Figure 15.29a, b). Much of the native temperate forest in the developed world was destroyed long ago, but current rates of deforestation remain high in all biomes (Figure 15.29c), though it is likely that whereas in temperate zones more of this forest loss currently is natural (for example forest fires), in tropical zones more is human-induced. The process of habitat destruction also often makes the habitat available to a particular species more fragmented than in the past, with potentially serious effects for the populations concerned (see Section 15.4.5).

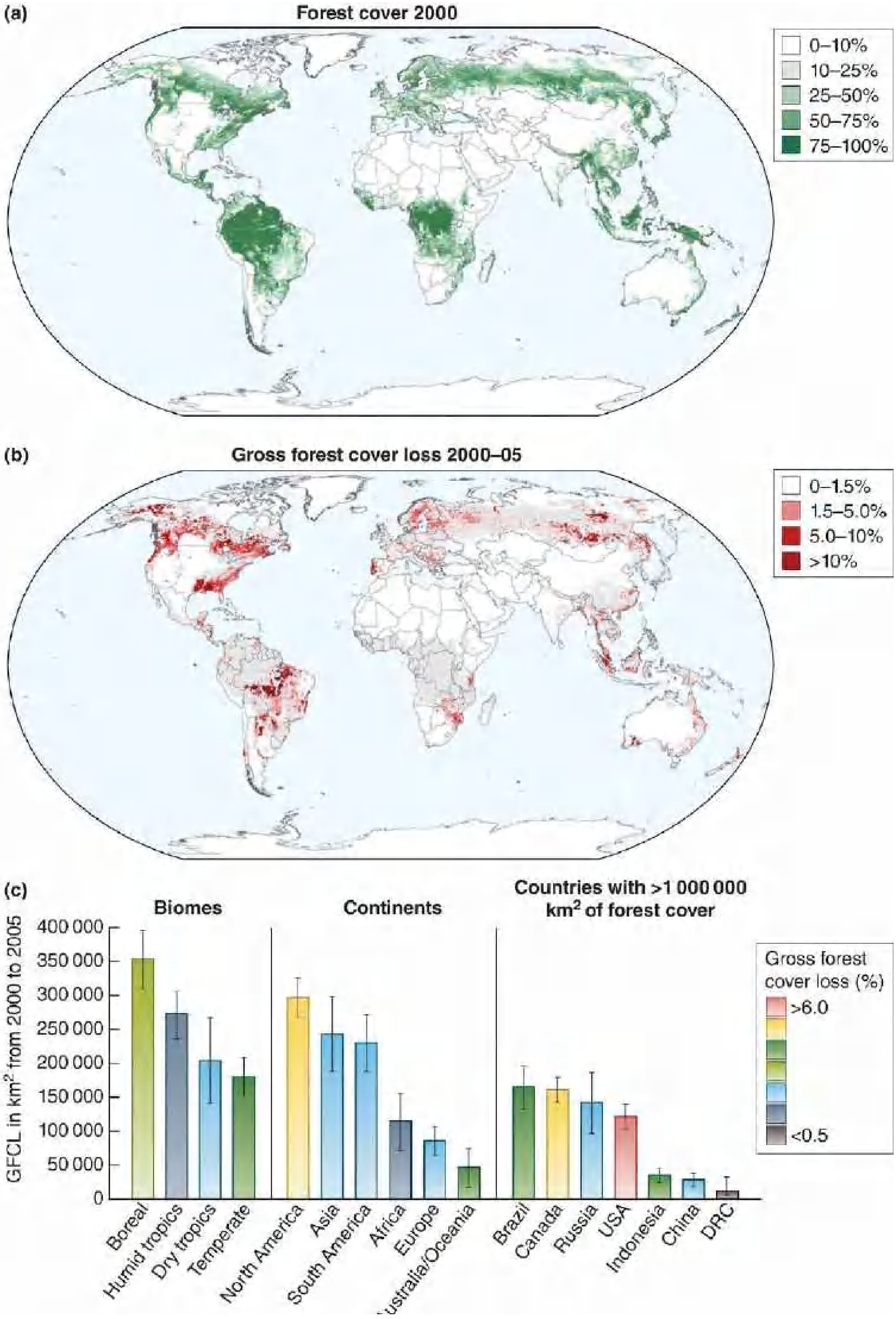

(a)
Forest cover 2000
0–10%
10–25%
25–50%
50–75%
75–100%
(b)
Gross forest cover loss 2000–05
0–1.5%
1.5–5.0%
5.0–10%
>10%
(c)
Biomes
Continents
Countries with >1 000 000 km² of forest cover
GFCL in km² from 2000 to 2005
400 000
350 000
300 000
250 000
200 000
150 000
100 000
50 000
0
Boreal
Humid tropics
Dry tropics
Temperate
North America
Asia
South America
Africa
Europe
ustralia/Oceania
Brazil
Canada
Russia
USA
Indonesia
China
DRC
Gross forest cover loss (%)
>6.0
<0.5

Figure 15.29 Forest cover is being lost throughout the world and from all biomes. (a) The estimated percentage forest cover throughout the world in 2000. (b) The estimated percentage gross forest cover loss (GFCL) throughout the world between 2000 and 2005. (c) GFCL 2000–05 in absolute terms for different biomes, continents and countries with large areas of forest cover. Bars are 95% CIs.

Source: After Hansen *et al.* (2010).

Secondly, habitat may be *degraded* by pollution to the extent that conditions become untenable for certain species. Degradation by pollution can take many forms, from the application of pesticides that harm non-target organisms (discussed in Section 15.2.2), to acid rain (see Section 22.3), to the run-off of fertilisers from agricultural land and the eutrophication of water bodies to which it can give rise (see Section 22.4).

Thirdly, habitat may simply be *disturbed* by human activities to the detriment of some of its occupants. Habitat disturbance is not such a pervasive influence as destruction or degradation, but certain species are particularly sensitive. For example, diving and snorkelling on coral reefs, even in marine protected areas (see Application 19.4), can cause damage by means of direct physical contact with hands, body, equipment and fins. Often individual disturbances are minor, but these can result in serious cumulative damage. In one analysis of 214 divers in a marine park on Australia's Great Barrier Reef, 15% of divers damaged or broke corals, mostly by fin flicks (Rouphael & Inglis, 2001). Specialist underwater photographers caused more damage on average (1.6 breaks per 10 minutes) than divers without cameras (0.3 breaks per 10 minutes). Impacts were also much more likely to be caused by male than female divers. Nature recreation, ecotourism, and even ecological research are not without risk of disturbance and the decline of the populations concerned.

global environmental change: the case of Australian butterflies

When the environment changes, species may find themselves without a home, or with a home that is no longer large enough to accommodate them, or with a home that has moved too far or too fast for them to follow. We discuss this in detail in Section 22.2. One way to assess the extinction risk of species under global climate change is to estimate the loss in area of key habitats on the basis of predicted changes to temperature and rainfall. Beaumont and Hughes (2002) used predicted climate changes to model the future distributions of 24 Australian butterfly species. Under even a moderate set of future conditions (temperature increase of 0.8–1.4°C by 2050), the distributions of 13 of the species decreased by more than 20% (Figure 15.30). Most at risk are butterflies like *Hypochrysops halyaetus*, that not only have specialised food plant requirements, but that also depend on the presence of ants for a mutualistic relationship (see Section 13.4.2). This species is predicted to lose 58–99% of its current range. Moreover, only one-quarter of its predicted future distribution occurs in locations that it currently occupies. This highlights a very important general point: regional conservation efforts and current nature reserves may turn out to be in the wrong place in a changing world.

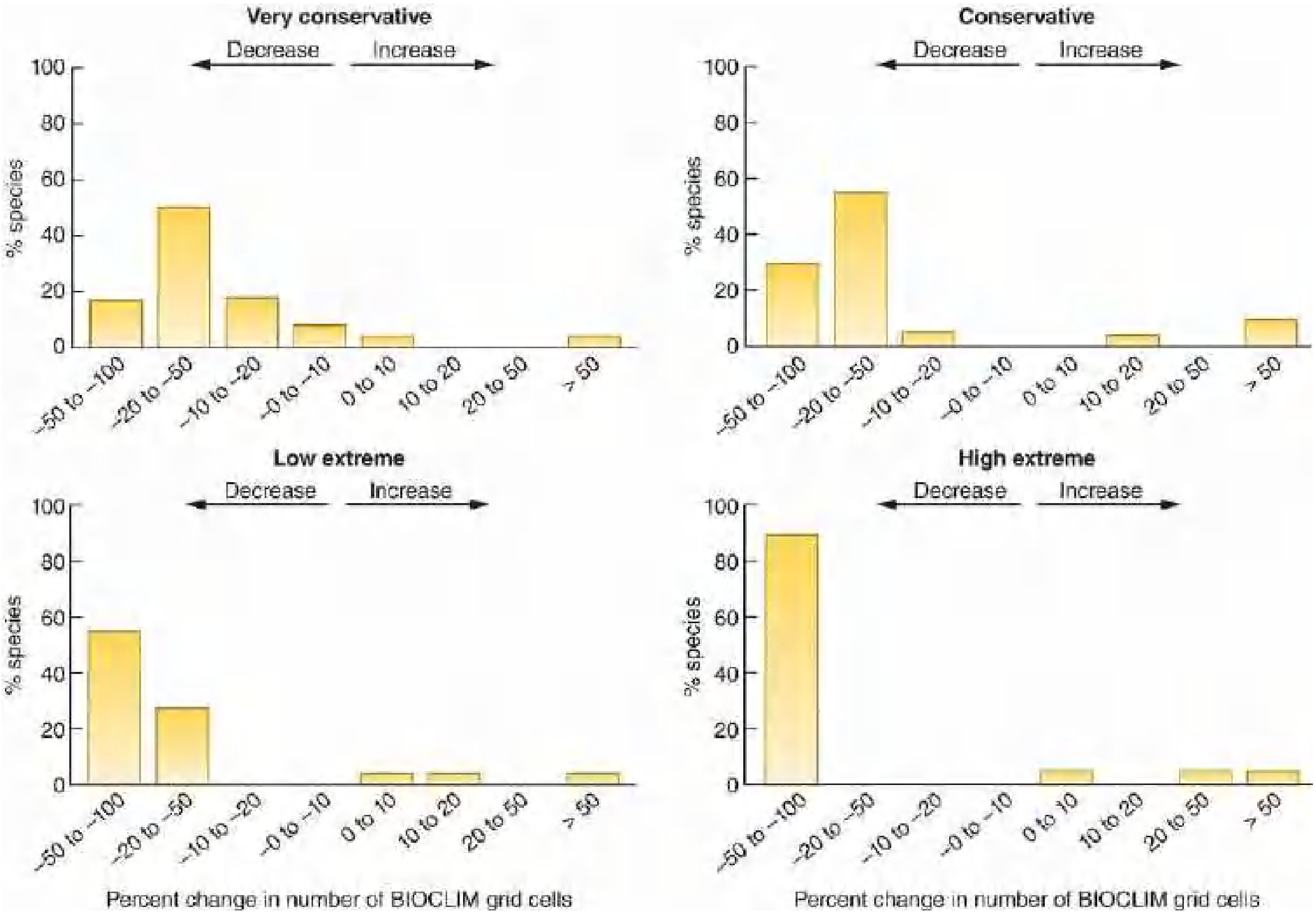


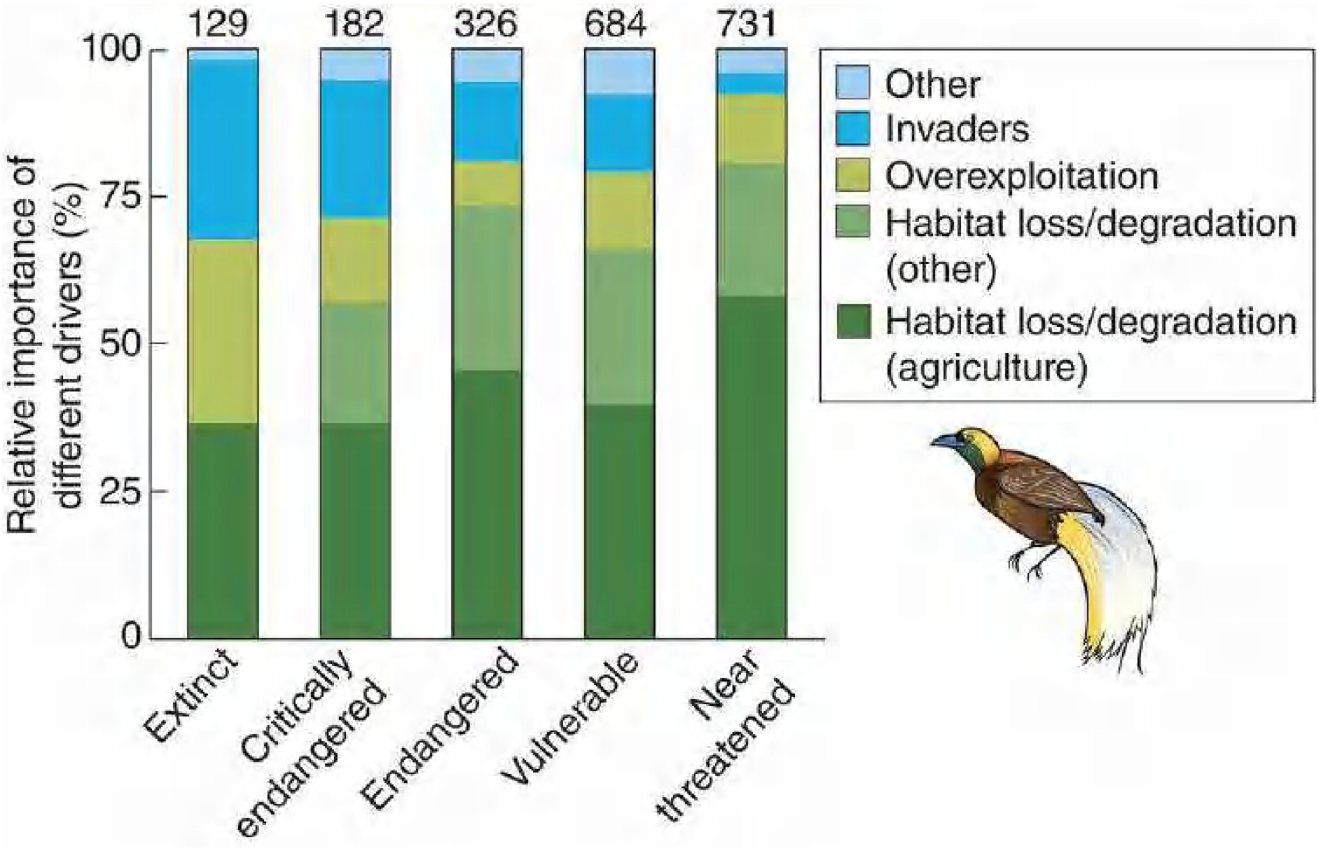
Figure 15.30 Even with conservative predictions, many Australian butterfly species are threatened with extinction by climate change. The effects of climate change to 2050, as predicted by the bioclimatic computer program BIOCLIM, for 24 species of Australian butterfly subjected to four climate change scenarios ranging from 'very conservative' (CO_2 increase to 479 ppm, Australian mean annual temperature (AMAT) increase 0.8–1.4°C, Australian mean annual precipitation (AMAP) increase 0–18%) to 'high extreme' (CO_2 increase to 559 ppm, AMAT increase 2.1–3.9°C, AMAP increase 0–59%). The program measures habitable area as the number of BIOCLIM grid cells.

Source: After Beaumont & Hughes (2002).

invasive species

Invasions of exotic species into new geographic areas sometimes occur naturally and without human agency. However, human actions have increased this trickle to a flood. Human-caused introductions may occur either accidentally as a consequence of human transport, or intentionally but illegally to serve some private purpose, or legitimately to procure some hoped-for public benefit by bringing a pest under control, producing new agricultural products, or providing novel recreational opportunities. Many introduced species are assimilated into communities without much obvious effect. However, some have been responsible for dramatic changes to native species and natural communities.

We discuss invasions more fully in Section 22.7, but we can get some sense of the scale of the problem of invasive or introduced species by noting that these rank second after habitat degradation among the factors threatening bird biodiversity (Figure 15.31), and that of the 958 species classified as 'imperilled' in the USA (Wilcove *et al.*, 1998) 42% are at risk primarily because of introduced or invasive species.



[**Figure 15.31**](#) **Invasive species are the second greatest threat to bird species in the USA.** Relative importance of different drivers responsible for the loss or endangerment of bird biodiversity. Patterns are shown for five categories of extinction threat (see [Section 15.4.2](#)). The values above each histogram are the numbers of species in each threat category overall. Habitat loss/degradation poses a much bigger risk now than in the past (compare histograms for endangered and vulnerable categories with extinct birds), and this is set to increase in the future, in particular via agricultural expansion (histogram for near-threatened species).

Source: After Balmford & Bond ([2005](#)).

infectious disease: the case of chytrid disease of amphibians worldwide

We saw earlier, when we described the final years of the heath hen, how infectious disease may play its part in the decline of a species towards extinction. This integral role, embedded in a range of factors, is probably typical. Recently, however, the power of infectious disease in its own right in driving species to extinction has been emphasised by the plight of amphibian species worldwide (see [Figure 15.22](#)a). An estimated 50% of all amphibian species are currently facing an extinction crisis, with more than 100 already classified as extinct or possibly extinct. One major cause is chytrid disease, caused by the fungus *Batrachochytrium dendrobatidis*, which has been implicated as the immediate cause of catastrophic declines in more than 200 of the 350 species it is known to infect (Fisher *et al*., [2009](#)). Although it was identified for the first time only in 1997, its effects undoubtedly go back further than that.

One example is illustrated in [Figure 15.32](#)a, which shows the moving wave of amphibian declines in Central America from 1987 to 2004. A study at what turned out to be the front of that wave in 2004 had been monitoring the abundance of frogs going back to 1998. The precipitous decline in amphibian abundance in 2004 is shown in [Figure 15.32](#)b. Prior to the die-off, none of more than 1500 frog samples from 43 species showed any sign of the infection. In 2004, the prevalence of infection in 879 samples from 48 species was around 50% (Lips *et al*., [2006](#)).

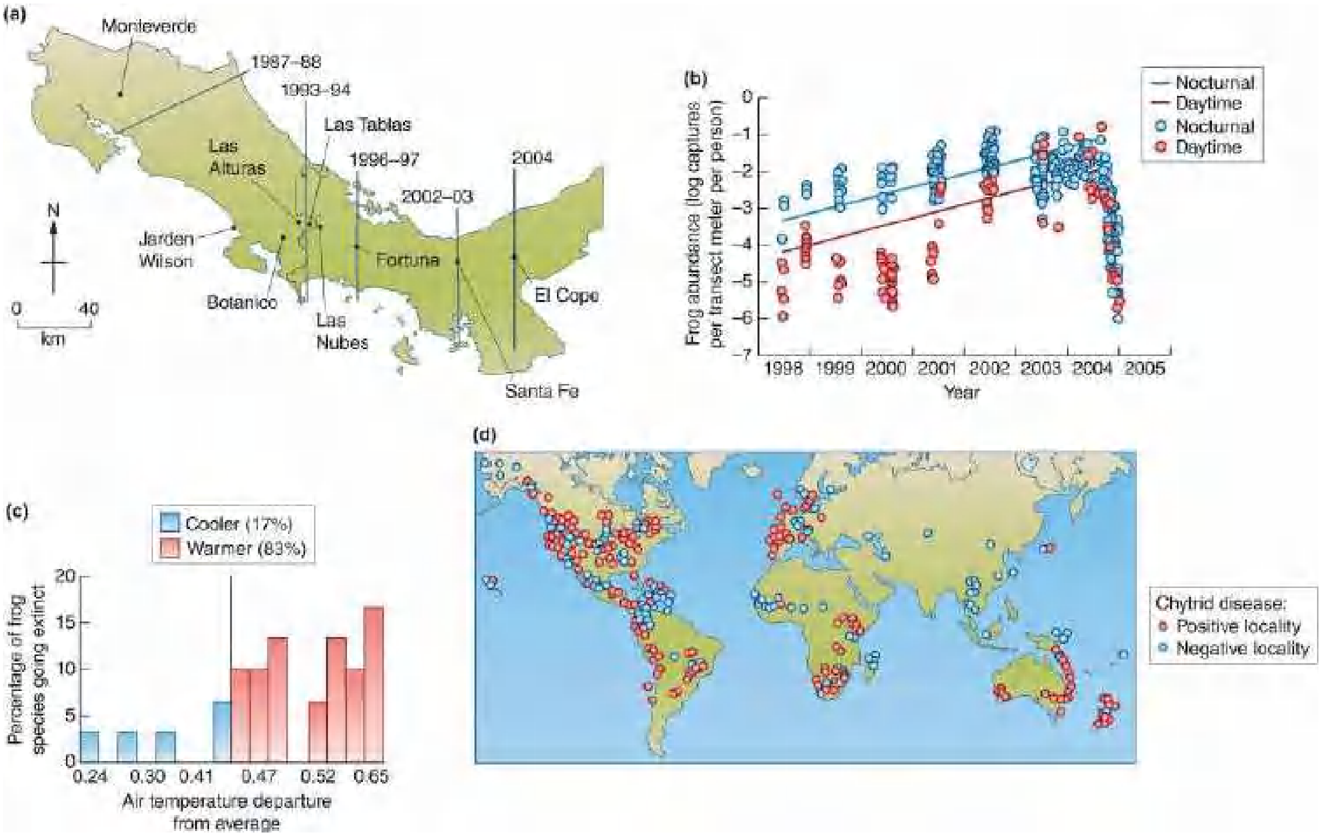


Figure 15.32 Chytrid disease in amphibians spread through Central America from the late 1980s and has been spreading almost everywhere in the world. (a) A map of Central America, with sites of reported major declines in amphibian populations, and lines showing the approximate location of a wave of declines as it moved south-east. (b) Frogs observed between 1998 and 2005 in daytime and nocturnal samples taken at El Copé (see map in (a)). Statistical tests suggest a significant shift in the direction of the relationship ($P < 0.001$) on 4 September 2004 coinciding with the first appearance of chytrid disease. (c) The percentage of species of harlequin frog (*Atelopus* spp.) going extinct in the face of chytrid disease in Costa Rica at different temperatures (measured as departures in air temperature from the 1950–79 average). (d) The current distribution of chytrid disease in amphibians worldwide.

Source: (a, b) After Lips *et al.* (2006). (c) After Pounds *et al.* (2006). (d) After Fisher *et al.* (2009).

In fact, the spread of the disease seems itself to be linked to global environmental change. Certainly in Costa Rica at least, the loss of species is closely linked to unusually warm conditions that favour its spread (Figure 15.32c). Overall, therefore, the threat to amphibians is grave. The infection is already found throughout the world (Figure 15.32d) and is likely to spread further still.

extinction vortices: the case of the southern dunlin

Some species are at risk for a single reason, but often, as in the case of the heath hen discussed earlier, a combination of factors is at work, perhaps occurring sequentially – a so-called *extinction vortex* (Figure 15.33a). The declining population of a shore bird in Sweden, the southern dunlin, *Calidris alpina schinzii*, appears to be entering such a vortex. During the study period, 1993–2004, the population declined steadily (Figure 15.33b), but habitat loss that had previously threatened the species, and reduced the size of its populations, had been arrested by this time, so it cannot have been the cause of the decline. However, there was clear evidence that close relatives were increasingly breeding with one another (Figure 15.33c), and that reduced heterozygosity, an inevitable consequence of inbreeding, was leading to an increase in the failure of dunlin eggs to hatch (Figure 15.33d). The population has continued to decline, despite active measures to increase breeding success, including preventing the trampling of nests by cattle. It seems that earlier habitat loss may have propelled the dunlins into an extinction vortex from which inbreeding is now making it difficult to escape.

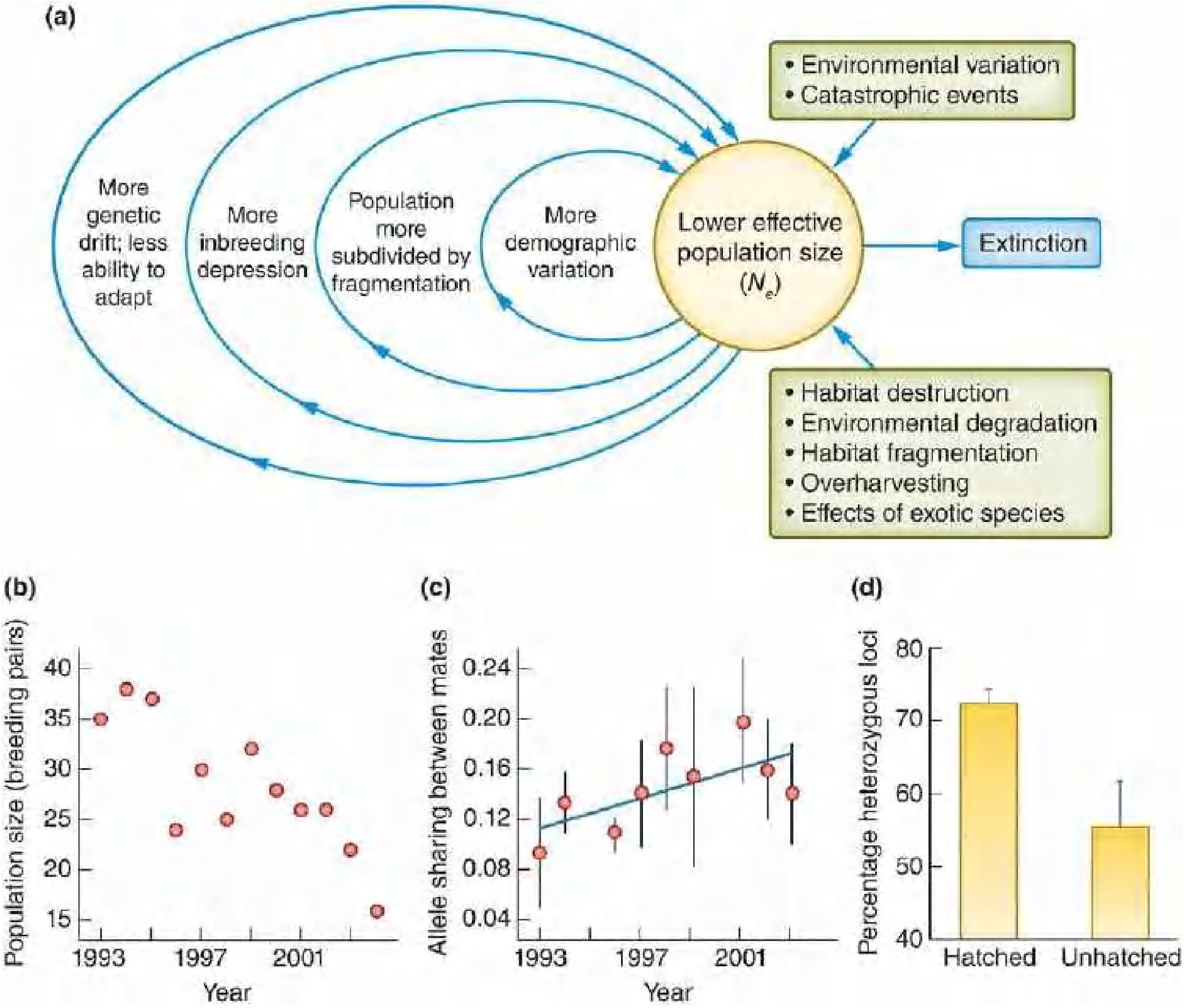


Figure 15.33 The extinction vortex in principle and in action in Swedish southern dunlins. (a) The manner in which extinction vortices may progressively lower population sizes leading inexorably to extinction. (b) The declining size of a population of the southern dunlin in south-west Sweden between 1993 and 2004. (c) The increasing genetic similarity between mating pairs of the dunlins over the same period, as measured by the proportion of gene alleles they share (bars are SEs, $P < 0.05$). (d) The percentage of gene loci that are heterozygous among dunlin chicks that hatched or failed to do so (bars are SEs).

Source: (a) After Primack (1993). (b) After Blomqvist *et al.* (2008).

chains of extinctions: the case of pollinators on Guam

So far, we have looked at individual species, treating them as though they were largely independent entities. As we shall discuss in Chapters 16 and 17, however, species are typically bound together in a web of interactions, and a chain of extinctions may follow inexorably from the extinction of particular species, which therefore deserves special attention.

Flying foxes (large fruit bats) in the genus *Pteropus*, which occur on many South Pacific islands, are the major, and sometimes the only, pollinators and seed dispersers for hundreds of native plants (many of which are of considerable economic importance, providing medicines, fibre, dyes, prized timber and foods). Flying foxes are highly vulnerable to human hunters and there is widespread concern about declining numbers. On the island of Guam, the two indigenous flying fox species are either extinct or virtually so, and there are already indications of reductions in fruiting and dispersal. Guam is also now home to the accidentally introduced brown tree snake, *Boiga irregularis*, which may have contributed to the decline of the flying foxes. Its most devastating effects, however, have been on indigenous birds, many species of which are also important pollinators, and here too, the cascading effects from snake to pollinating birds to plants are apparent.

A comparison with the nearby but nearly snake-free island of Saipan for two important tree species showed, first, that visits by pollinating birds had ceased on Guam and pollinating visits overall were significantly down (Figure 15.34a, b), but also that recruitment of the trees (i.e. appearance of seedlings and saplings) was failing (Figure 15.34c). These are only examples of tree species on Guam pollinated mainly by birds. If urgent conservation measures are not taken to protect the plants by protecting the pollinators, this lowered recruitment may in due course be fatal.

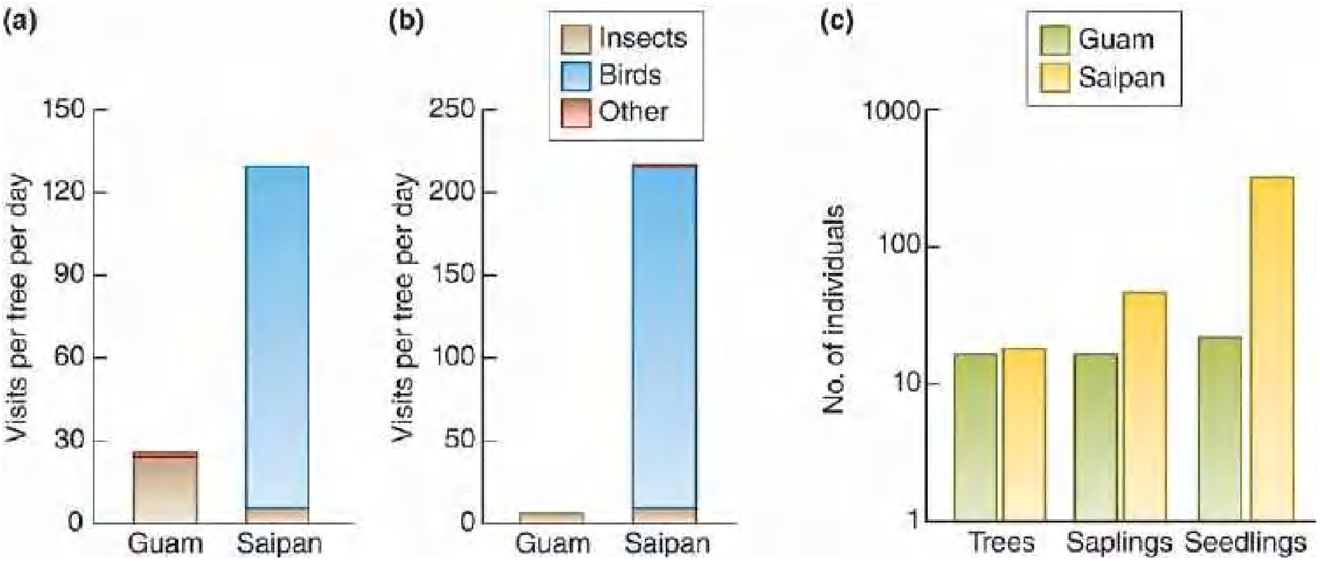


Figure 15.34 Few pollinators, no bird pollinators and low recruitment of new trees on the brown tree snake-infested island of Guam. (a) Average daily number of pollinator visits per tree to the mangrove tree *Bruguiera gymnorrhiza* by insects, birds and other visitors on the nearby islands of Guam and Saipan. (b) Equivalent results for the forest tree *Erythrina variegata*. (c) Mature trees, saplings and seedlings of *B. gymnorrhiza* present in study sites on Guam and Saipan, as indicated.

Source: After Mortensen *et al*. (2008).

15.4.4 Minimum viable populations and population viability analysis

MVPs

Given the dangers of the extinction vortex, conservation at the species level is often focused on how close a population is to the brink of such a vortex. Considerations like this have led to the concept of a *minimum viable population* (MVP) – a species-specific abundance threshold, below which the threat of extinction (being sucked into the vortex) is dangerously or unacceptably high (Shaffer, 1981). Note immediately, firstly, that what defines an MVP is therefore a matter of judgement. We might suggest, for example, that an MVP will have at least a 95% probability of persistence for 100 years, but there is no objective or consensus definition of 'dangerous' or 'unacceptable'. Secondly, while the concept was originally focused on populations, it has often been broadened to include the threat facing whole species. The distinction between the two is important from a practical point of view, since, clearly, a species may survive the loss of one or more of its populations. It is also important in terms of conservation priorities. There is often considerable, and understandable, concern about local extinctions – the loss of a particular population: perhaps the only one in a country or region – but there can be little doubt that the loss of a species globally is more concerning than this.

triage, complacency and fatalism

The use of MVPs in determining conservation priorities can be double-edged. In desperate times, painful decisions have to be made. Thus, wounded soldiers arriving at field hospitals in the First World War were subjected to a *triage* evaluation, distinguishing first those who were likely to survive but only with rapid intervention, second those who were likely to survive without rapid intervention, and third those who were likely to die with or without intervention, and on whom intervention was therefore likely to be wasted. Conservation managers are often faced with the same kind of choices and need to demonstrate some courage in giving up on hopeless cases, and prioritising those species and habitats where something can be done. Determining MVPs, and how close species are to their MVP, can be crucial in making such decisions. But there are dangers both ways. Underestimating an MVP can lead to dangerous complacency: failing to act to conserve a species before it is too late to do so. But overestimating an MVP can lead to equally dangerous fatalism: giving up on a species or population that our efforts could rescue.

genetics – a 50/500 rule?

There have been both generic and more customised approaches to judging how close populations are to their MVP. On the one hand, *population viability analyses* (PVAs) combine data and models to determine how close a population is to extinction over a specified time frame and on the basis of stated assumptions. We turn to these shortly. On the other hand, there have been efforts to come up with more generic rules of thumb that can be applied to all populations, particularly from the perspective of genetic and evolutionary viability. As Frankham *et al.* (2014) describe, this goes back to Franklin (1980), in particular, who reviewed the evidence available at the time and proposed that an effective population size of at least 50 was required to avoid deleterious effects of inbreeding in the short term (five generations), while one of 500 was required to maintain adaptive evolutionary potential in perpetuity. These figures have taken on an almost doctrinal status. Subsequent empirical studies, however, and reconsiderations of the theoretical arguments, have suggested that problems may be detectable in larger populations than this. Frankham *et al.* (2014) therefore propose that the figures should be doubled to 100 and 1000, and that the guideline figures for assigning the IUCN statuses, 'vulnerable', 'endangered' and 'critically endangered' should also be doubled.

clues from long-term studies

The focus of PVA differs from many of the population models developed by ecologists because its aim, ultimately, is to predict an extreme event, extinction, rather than central tendencies such as mean population sizes. Increasingly, however, PVA has moved away from the simple estimation of extinction probabilities and times to extinction, to focus on the comparison of likely outcomes of alternative management strategies. Arguably the most straightforward approach to predicting future extinctions is to do so directly from what has happened in the past. Figure 15.35 illustrates data on the survival of populations of bighorn sheep, *Ovis canadensis*, in desert areas of North America. If we define an MVP as one that will give at least a 95% probability of persistence for 100 years, we can explore these data to provide an approximate answer. Populations of fewer than 50 individuals all went extinct within 50 years whilst only 50% of populations of 51–100 sheep lasted for 50 years. Evidently, for our MVP we require more than 100 individuals; in the study, such populations demonstrated close to 100% success over the maximum period studied of 70 years. A similar analysis of long-term records of birds on the Californian Channel Islands indicates an MVP of between 100 and 1000 pairs of birds needed to provide a probability of persistence of between 90% and 99% for the 80 years of the study (Thomas, 1990). However, studies such as these are rare. The long-term data are available because of the extraordinary interest people have in hunting (bighorn sheep) and ornithology (Californian birds).

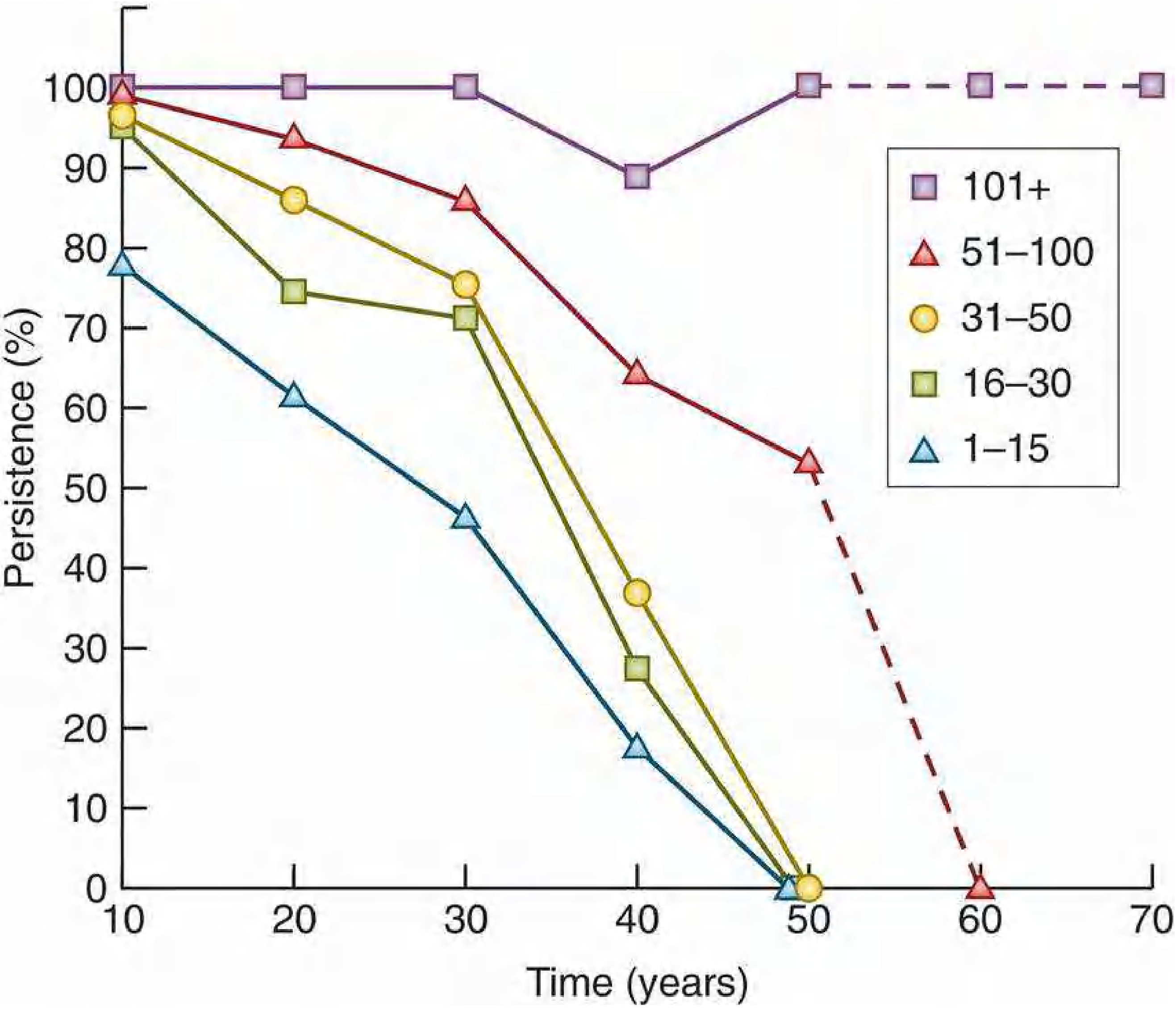


Figure 15.35 Persistence times of different-sized bighorn sheep populations suggest a minimum viable population. The percentage of populations of bighorn sheep in North America that persists over a 70-year period with various initial population sizes, as indicated.

Source: After Berger (1990).

Thus, most PVAs take contemporary data and then use models such as population projection matrices (PPMs) or integral projection models (IPMs; Section 4.8) to predict future trajectories for the focal populations and hence their extinction probabilities. These are often embedded in computer programs designed for PVA, such as RAMAS, which can also incorporate spatial information (Akcakaya & Root, 2013), or VORTEX, which can also incorporate genetic data (Lacy

& Pollack, 2013), They do so by stochastic simulation. That is, models are run – populations are projected forward – multiple, typically thousands of times, each time with a slightly different set of parameter values chosen at random from a realistic distribution of values centred on the best estimates of those parameters, reflecting the inevitable uncertainty in those parameter estimates and the stochasticity in the underlying processes. Hence, the precise outcome is different in each run, allowing, for example, the percentage of cases in which the population goes extinct within 100 generations to be calculated, which then becomes an estimate of its probability of going extinct within 100 generations.

PVA of an endangered plant, using RAMAS

We saw in Application 4.3, for example, how an IPM was used to identify populations of the northern wheatear in the Netherlands under threat of extinction and what management options might be used to counteract that threat. Another example is provided by the royal catchfly, *Silene regia*, a long-lived prairie perennial plant that has suffered a dramatic reduction in its range. Menges and Dolan (1998) collected demographic data for up to seven years from 16 populations in the US Midwest. The species has high survivorship, slow growth, frequent flowering and non-dormant seeds, but has very episodic recruitment: most populations in most years fail to produce seedlings. The populations, whose total adult numbers ranged from 45 to 1302, had been subject to different management regimes. A projection matrix was produced for each population in each year; one is illustrated in Table 15.2. Multiple simulations were then run in RAMAS for every matrix to determine the finite rate of increase (λ; also often referred to as *R*, see Section 4.7) and the probability of extinction in 1000 years. Figure 15.36 shows the median λ for the 16 populations, grouped into cases where particular management regimes were in place, for years when recruitment of seedlings occurred and for years when it did not. All sites where λ was greater than 1.35 when recruitment took place were managed by burning and some by mowing as well; none of these were predicted to go extinct during the modelled period. On the other hand, populations with no management, or whose management did not include fire, had lower values for λ and all except two had predicted extinction probabilities (over 1000 years) of from 10% to 100%. The obvious management recommendation is to use prescribed burning to provide opportunities for seedling recruitment. Low establishment rates of seedlings in the field may be due to frugivory by rodents or ants and/or competition for light with established vegetation (Menges & Dolan, 1998) – burnt areas probably reduce one or both of these negative effects. Also, while management regime was by far the best predictor of persistence, it is of interest that populations with higher genetic diversity also had higher median values for λ.

Table 15.2 An example of a projection matrix for a particular *Silene regia* population from 1990 to 1991, assuming recruitment. Numbers represent the proportion changing from the stage in the column to the stage in the row (bold values represent plants remaining in the same stage). 'Alive undefined' represents individuals with no size or flowering data, usually as a result of mowing or herbivory. Numbers in the top row are seedlings produced by flowering plants. The finite rate of increase λ for this population is 1.67. The site is managed by prescribed burning.

Source: After Menges & Dolan (1998).

	Seedling	**Vegetative**	**Small flowering**	**Medium flowering**	**Large flowering**	**Alive undefined**
Seedling	–	–	5.32	12.74	30.88	–
Vegetative	0.308	**0.111**	0	0	0	
Small flowering	0	0.566	**0.506**	0.137	0.167	0.367
Medium flowering	0	0.111	0.210	**0.608**	0.167	0.300
Large flowering	0	0	0.012	0.039	**0.667**	0.167
Alive undefined	0	0.222	0.198	0.196	0	**0.133**

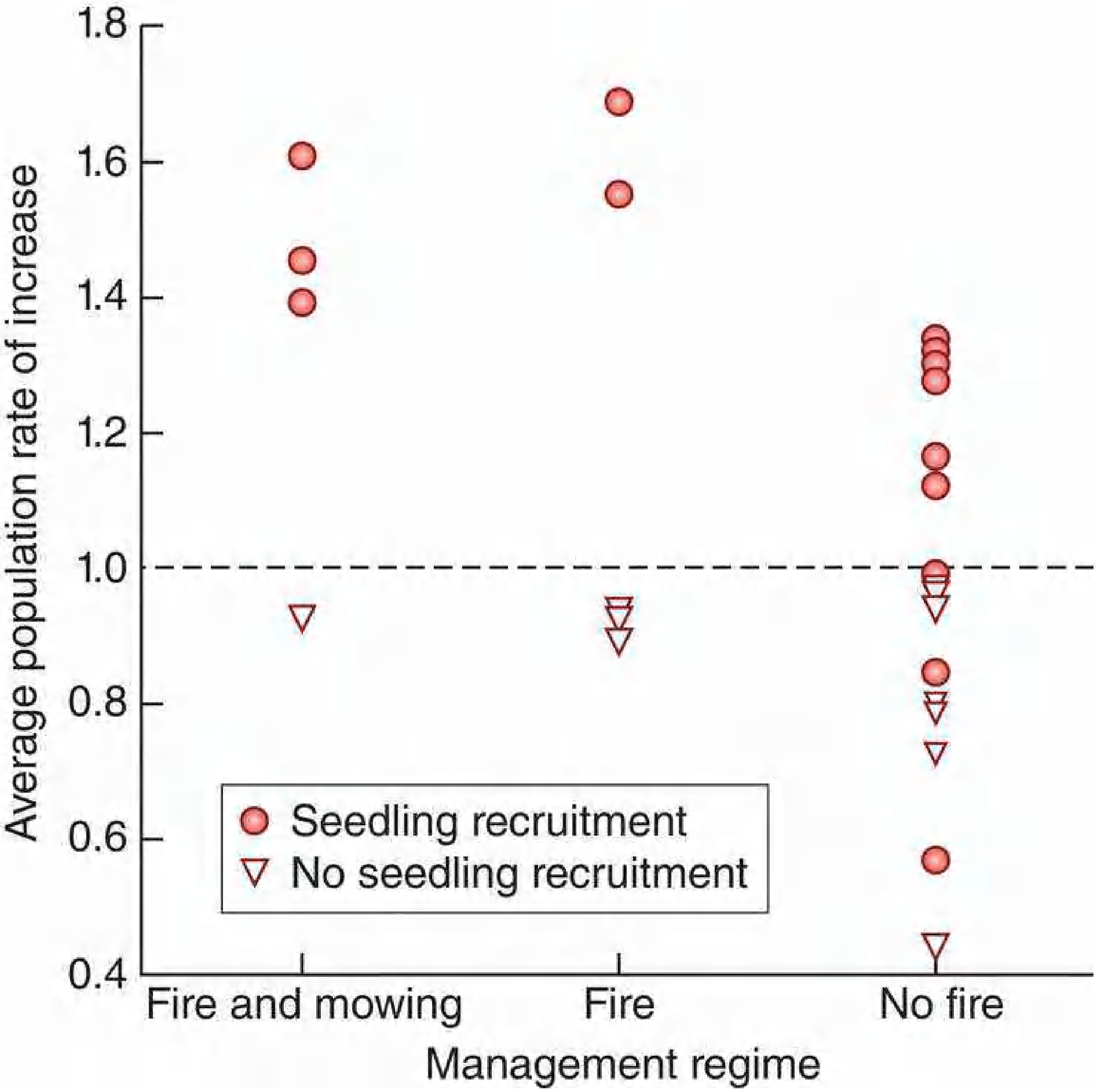


Figure 15.36 Populations of *Silene regia* not managed by burning are most likely to be *en route* to extinction. Median finite rates of increase of *Silene regia* populations as a function of management regime, for years with seedling recruitment and without. Unburned management regimes include just mowing, herbicide use or no management.

Source: After Menges & Dolan (1998).

a PVA for koalas, using VORTEX

As a second example, koalas (*Phascolarctos cinereus*) are regarded as 'near threatened' in Australia, with populations in different parts of the country varying from secure to vulnerable or extinct. Penn *et al.* (2000) used the PVA program VORTEX to model two populations in Queensland, one thought to be declining (Oakey) and the other secure (Springsure). Koala breeding commences at two years in females and three years in males. The other demographic values were derived from extensive knowledge of the two populations. The Oakey population was modelled from 1971 and the Springsure population from 1976, when first estimates of density were available. We can see that the model trajectories were indeed declining for Oakey and stable for Springsure (Figure 15.37). Over the modelled period, therefore, the probability of extinction of the Oakey population was 0.380 (380 of 1000 simulations went extinct), while that for Springsure was only 0.063. Conservation efforts should clearly be focused mostly on Oakey. Indeed, Penn *et al*. (2000) were able to compare the predictions of their PVAs with real population trajectories, because the koala populations have been continuously monitored since the 1970s (Figure 15.37). The predicted trajectories were close to the actual population trends, particularly for the Oakey population, and this gives us added confidence in the modelling approach.

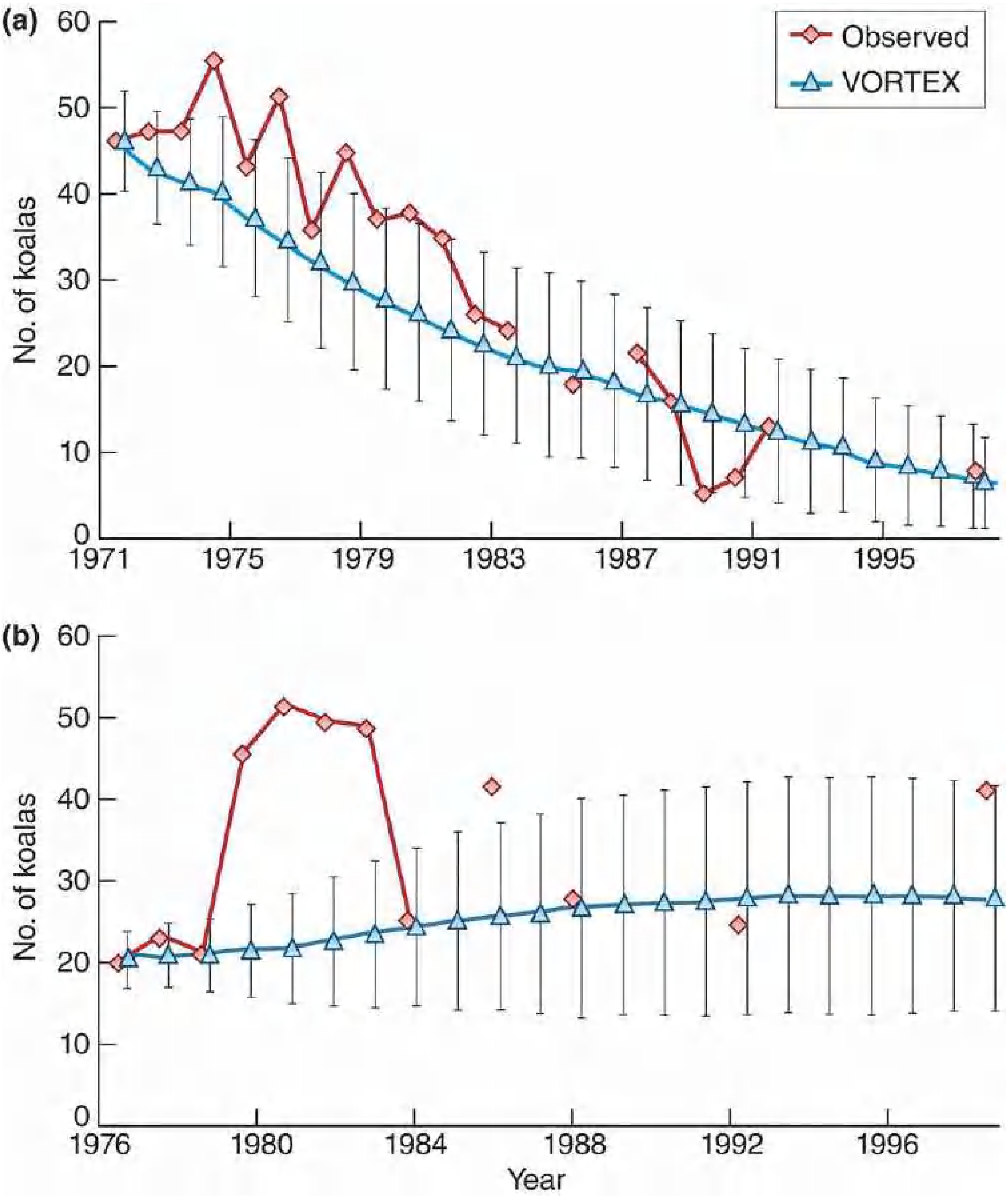


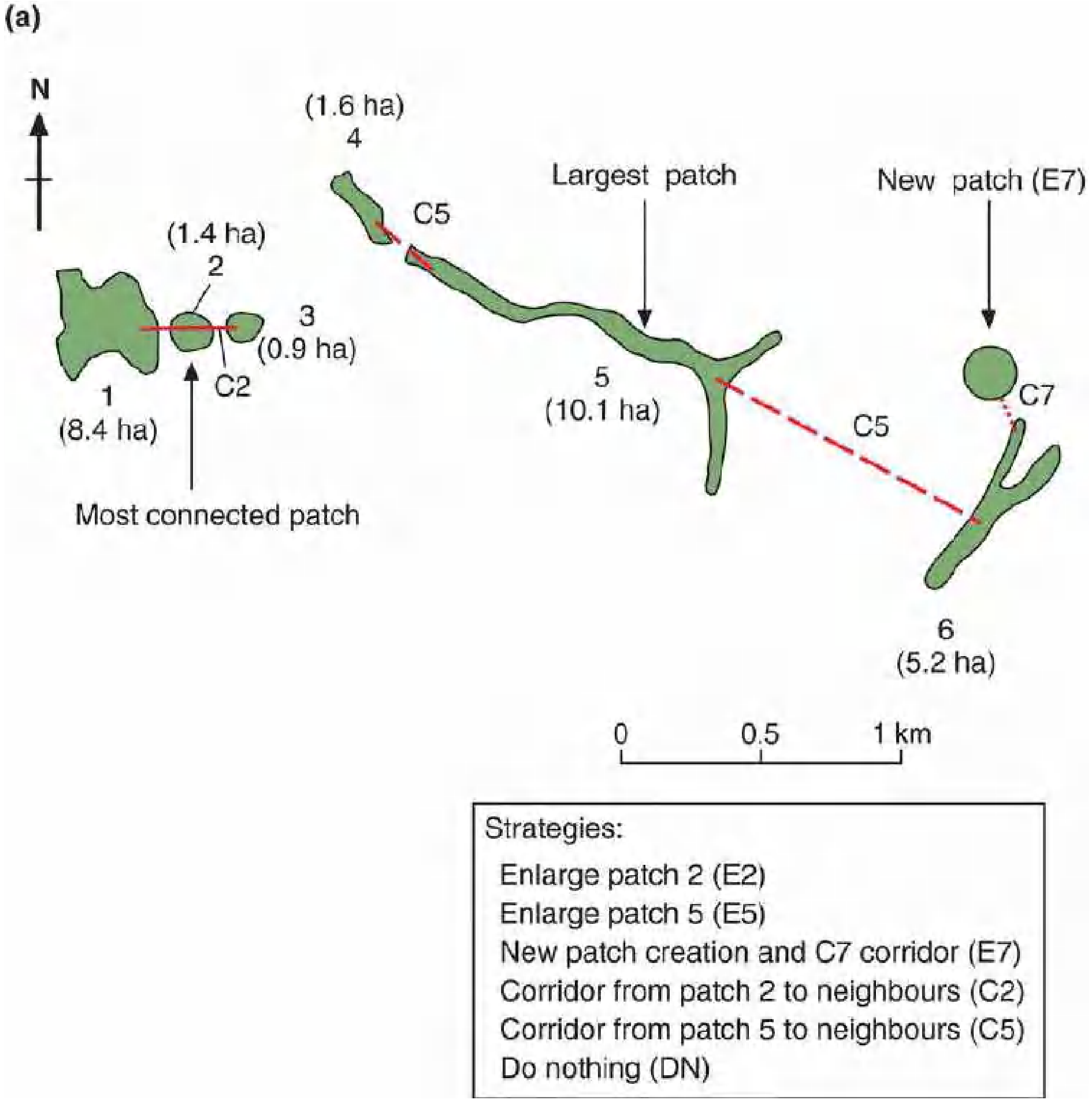
Figure 15.37 The VORTEX model correctly predicts declining and stable populations of koalas. Observed koala population trends (red diamonds) compared with predicted population performance (blue triangles, ±1 SD) based on 1000 repeats of the VORTEX modelling procedure at (a) Oakey and (b) Springsure. Real population censuses were not performed every year.

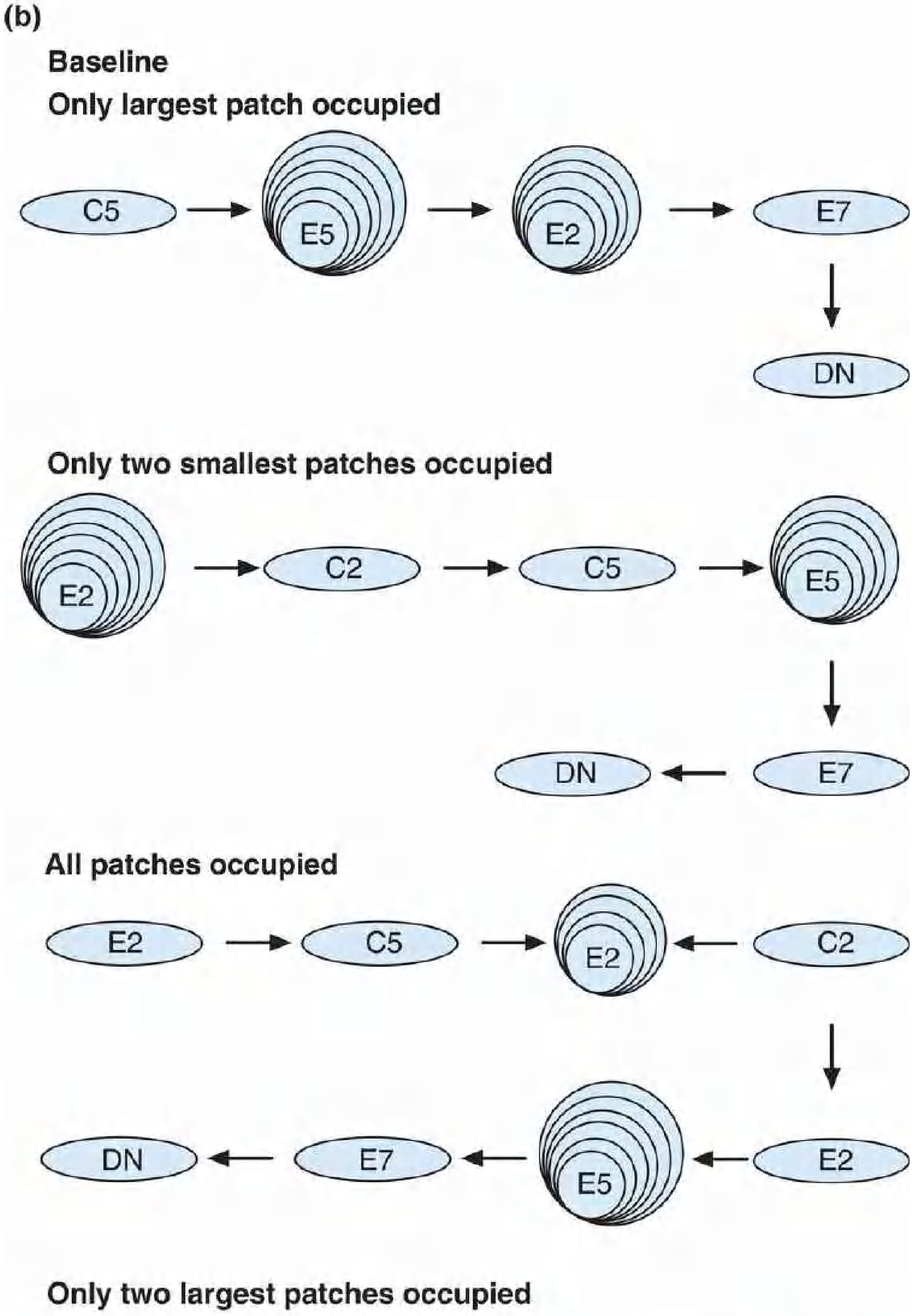
Source: After Penn *et al*. (2000).

15.4.5 Conservation of metapopulations

the case of the southern emu-wren: comparing the cost of different strategies

There is a clear link between the distinction between species and population conservation, on the one hand, and between population and metapopulation conservation on the other. For both whole species and metapopulations, we may need to consider explicitly the spatial distribution of individual populations and the connectivity between them. For example, Westphal *et al.* (2003) built a stochastic patch occupancy model based on realistic extinction and recolonisation matrices for the critically endangered southern emu-wren (*Stipiturus malachurus intermedius*) to find optimal solutions for its future management. The bird occupied a metapopulation in the Mount Lofty ranges of South Australia comprising just six remaining patches of dense swamp habitat (Figure 15.38a). The management strategies evaluated were the enlargement of existing patches, linking patches via corridors (emu-wrens are poor flyers), and creating a new patch. The optimisation modelling then compared individual management actions, and also a variety of successive actions, to find those that reduced the 30-year extinction risk to the greatest extent.


(a)
N
(1.6 ha)
4
Largest patch
New patch (E7)
C5
(1.4 ha)
2
3
(0.9 ha)
C2
1
(8.4 ha)
Most connected patch
5
(10.1 ha)
C7
C5
6
(5.2 ha)
0
0.5
1 km
Strategies:
Enlarge patch 2 (E2)
Enlarge patch 5 (E5)
New patch creation and C7 corridor (E7)
Corridor from patch 2 to neighbours (C2)
Corridor from patch 5 to neighbours (C5)
Do nothing (DN)


(b)
Baseline
Only largest patch occupied
C5
E5
E2
E7
DN
Only two smallest patches occupied
E2
C2
C5
E5
E7
DN
All patches occupied
E2
C5
E2
C2
E2
E5
E7
DN
Only two largest patches occupied

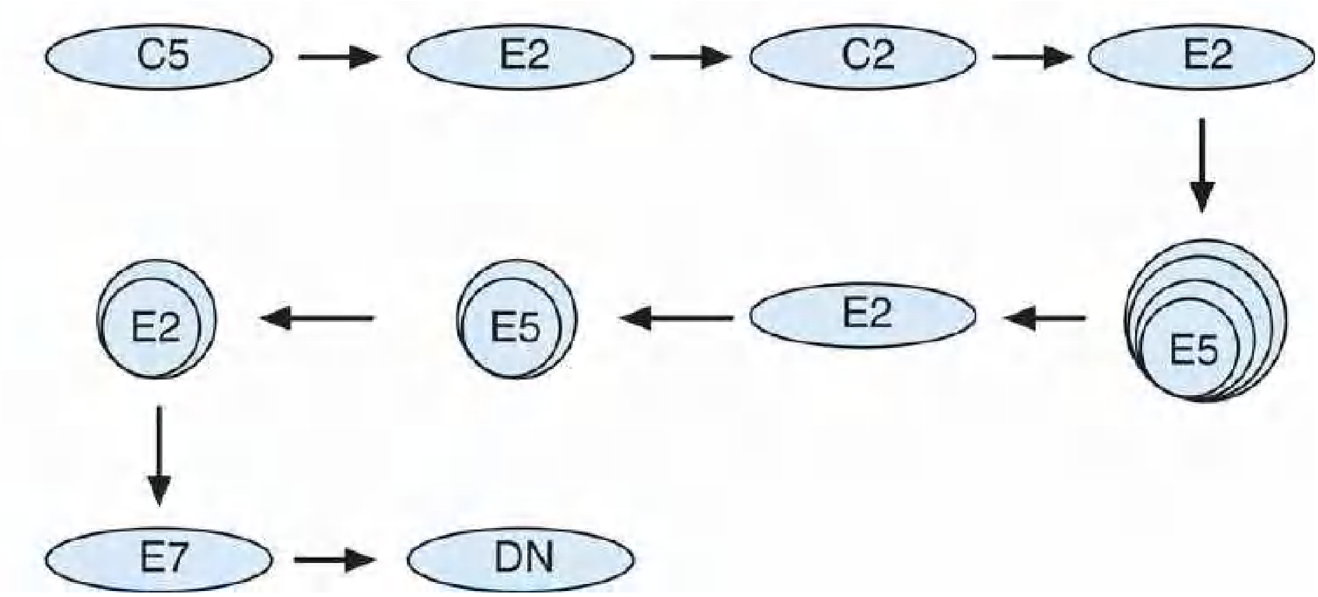


Figure 15.38 Optimal strategies for conserving an emu-wren metapopulation depend on the starting point. (a) The southern emu-wren metapopulation, showing the size and location of patches (green) and corridors (red lines). For further details, see text. (b) The optimal management scenario trajectories for different starting configurations of the southern emu-wren metapopulation. Each circle represents one action. Concentric circles show the repeated execution of a strategy before the next strategy is implemented. Note that each trajectory ends with a do-nothing action when the metapopulation state is such that lack of management action does not produce a probability of extinction that is significantly worse than any active strategy

Source: After Westphal *et al.* (2003).

Optimal decisions depended on the current state of the metapopulation (Figure 15.38b). For example, when only a more extinction-resistant large patch is occupied, connecting it to neighbouring patches is optimal (strategy C5). However, if only the two smallest patches are occupied, the optimal single action is to enlarge one of them (strategy E2). The best of these individual strategies reduced the 30-year extinction probabilities by up to 30%. However, the application of chains of different actions taken over successive time periods reduced extinction probabilities by 50–80% compared with no-management models. Again, the optimal strategy varied according to the starting state of the metapopulation (Figure 15.38b). Clearly, optimal decisions are highly state dependent, relying on knowledge of current patch occupancy. It will be hard to come up with simple rules of thumb for metapopulation management.

predicting the future for cougars

Cougars, *Puma concolor*, were absent from most of Midwestern USA for more than 100 years but have recently become a conservation success story as their populations, structured as a metapopulation, have expanded and moved eastwards since the 1990s. LaRue and Nielsen (2016) therefore utilised the PVA program RAMAS-GIS, which incorporates dispersal between patches, to ask which parts of the Midwest are likely to be occupied by cougars in 25 years' time, and what effect hunting of the cougars, currently occurring where they are established, could have on this. They used empirical estimates for fecundity, survival and dispersal, but also sought expert opinion (see Section 15.4.6) to determine habitat suitability. Focusing on eight main patches of suitable habitat (Figure 15.39), they calculated that the population as a whole would expand (λ = 1.08) and that all eight patches would be occupied in the absence of hunting and seven out of eight if it were continued, suggesting more broadly that cougars are likely to establish themselves over large areas of the North American Midwest over the next 25 years, irrespective of hunting pressure.

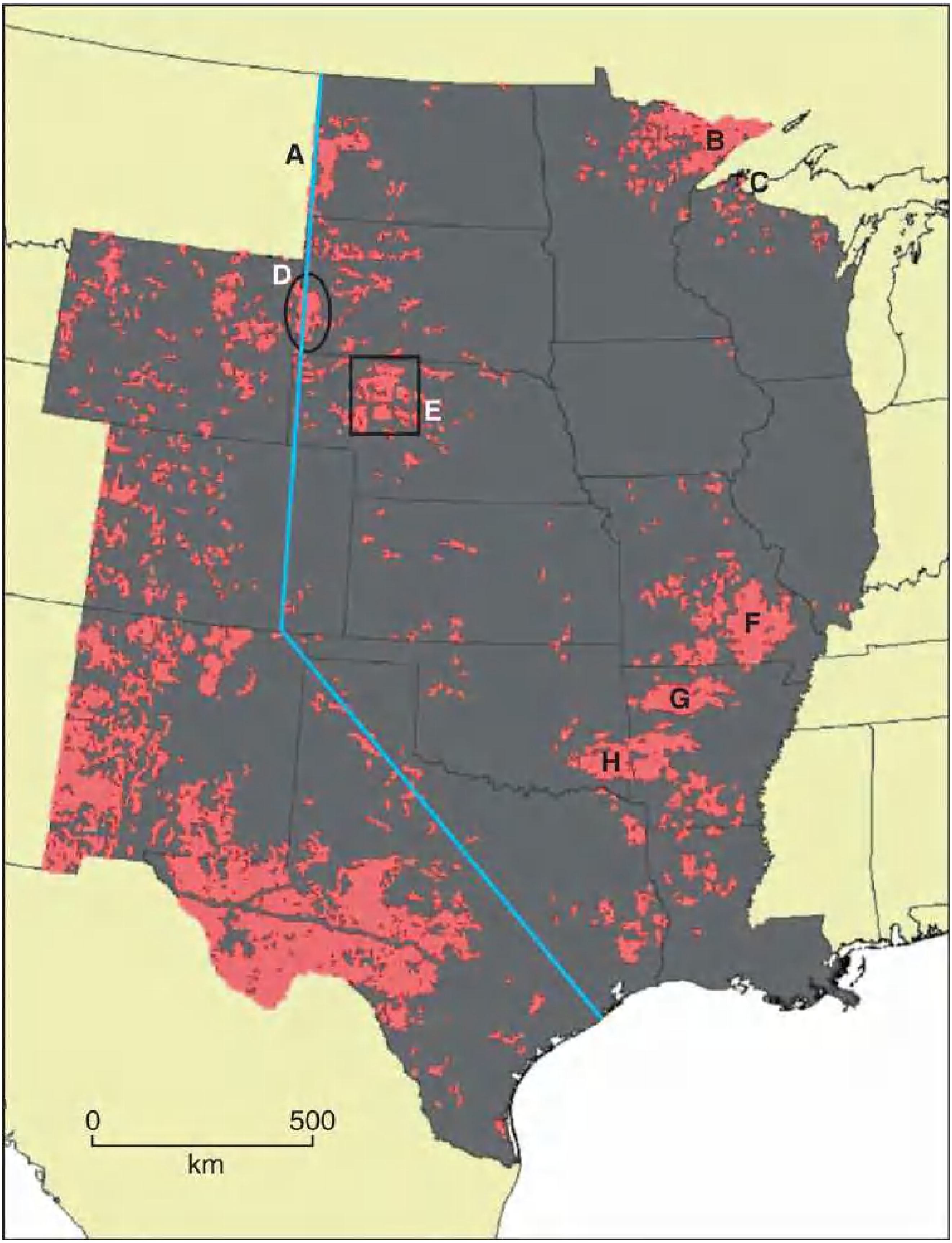

A
B
C
D
E
F
G
H
0
500
km

Figure 15.39 Large areas in the east of Midwestern USA are likely to be occupied by cougars in the next 25 years. A map showing areas of habitat suitable for cougars, *Puma concolor*, marked in red. By 2014, cougars had been sighted at dispersed locations to the east of the blue line. There are eight larger patches to the east that cougars may occupy in the next 25 years as follows: A, North Dakota; B, Northern Minnesota; C, Northern Wisconsin; D, Black Hills; E, Nebraska; F, Missouri; G, Ozarks; H, Ouachita. Simulations in the PVA program RAMAS-GIS suggest that all will be occupied.

Source: After LaRue & Nielsen (2016).

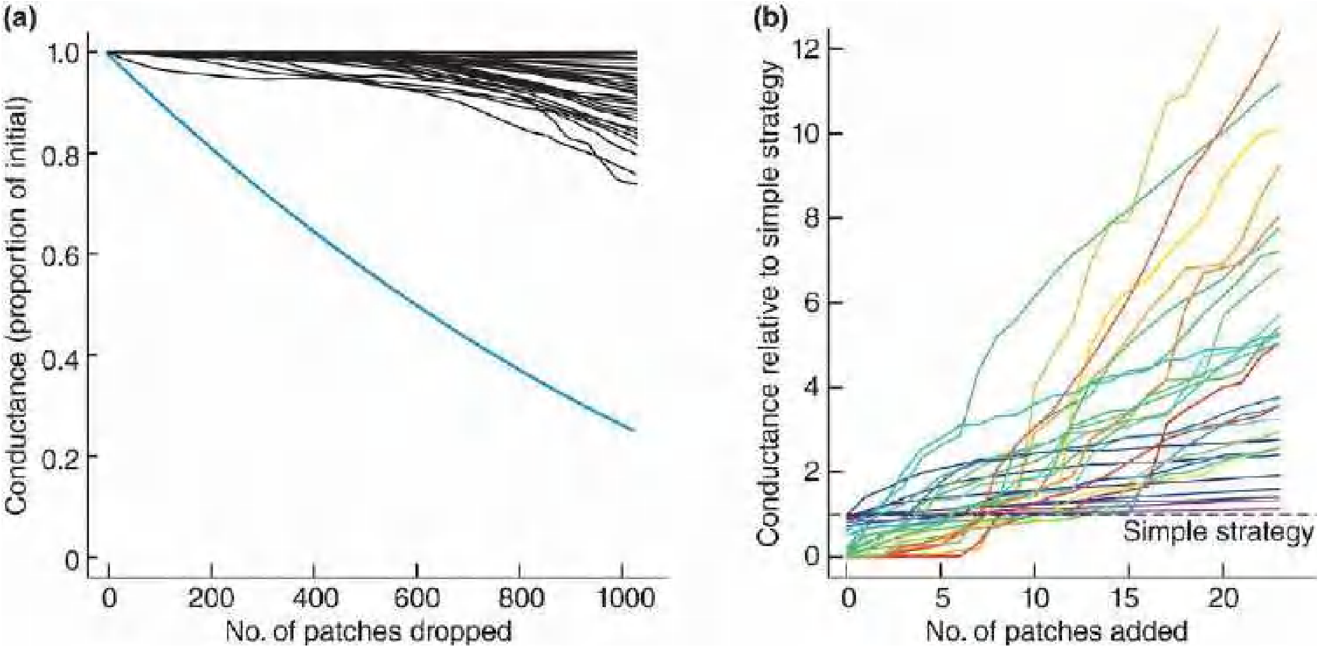


Figure 15.40 Strategic dropping of patches from or adding of patches to a network preserves or enhances the viability of a species that needs to move across the landscape in order to survive in the face of a changing climate. (a) The performance of a strategic algorithm iteratively removing patches in order to sustain conductance until half the patches on a 2048-patch network had been dropped (black lines, 30 networks) relative to the loss of conductance when patches are lost at random (blue line). (b) The performance of a strategic algorithm iteratively adding patches according to their power, relative to the simple strategy of adding patches in a straight line from source to target. Rainbow colours (from red to violet) represent different initial values of conductance in the network. The power-based strategy achieves better results with fewer patches.

Source: After Hodgson *et al*. (2016).

optimising range shifts

Programs like RAMAS-GIS are valuable for predicting future trajectories for populations with an explicit spatial structure, but they are not designed to address the key question of what spatial structure is most likely to enhance the viability of a population when, for example, it moves across a fragmented landscape in a range shift that tracks the changing global climate. Given that there will inevitably be constraints on available funds, what is the optimal allocation of those funds to habitat protection or enhancement that would allow that range shift to be successful? One approach to this has drawn an analogy between a spatially explicit metapopulation and an electrical circuit, with individuals moving between patches just as current moves between circuit nodes (Hodgson *et al*., 2016). On this basis, two methods for optimising flow were proposed and tested in simulations on a variety of artificial but realistic networks. Firstly, high ‘current’ (numbers of individuals) flowing through a patch indicates that it plays a valuable role in colonisation (and hence survival of the species in the face of change), suggesting a strategy in which the least valuable patches are iteratively dropped from a large network until a network that can be afforded remains. Alternatively, high ‘power’ in a link between two patches (high flow but also high resistance) indicates a bottleneck in the flow of individuals, suggesting a strategy in which patches that ease bottlenecks are iteratively added to a small network until costs become prohibitive. Both strategies perform significantly better than simply either dropping patches at random or adding patches in straight lines from a source population to a target (Figure 15.40).

Based on this work, a computer program, Condatis, has been developed to help preserve and enhance connectivity in real world habitat networks (Wallis & Hodgson, 2018). An example is shown in Figure 15.41. A UK regional agency, Warwickshire County Council, have used Condatis, alongside a classic metapopulation connectivity measure (based on Hanski’s models – see Section 6.7.4), to inform their work on ‘green infrastructure’, and in particular their decisions on whether to approve or reject planning applications. They also make the analysis results freely available online so that, for example, local community groups can use them to object to a planning application.

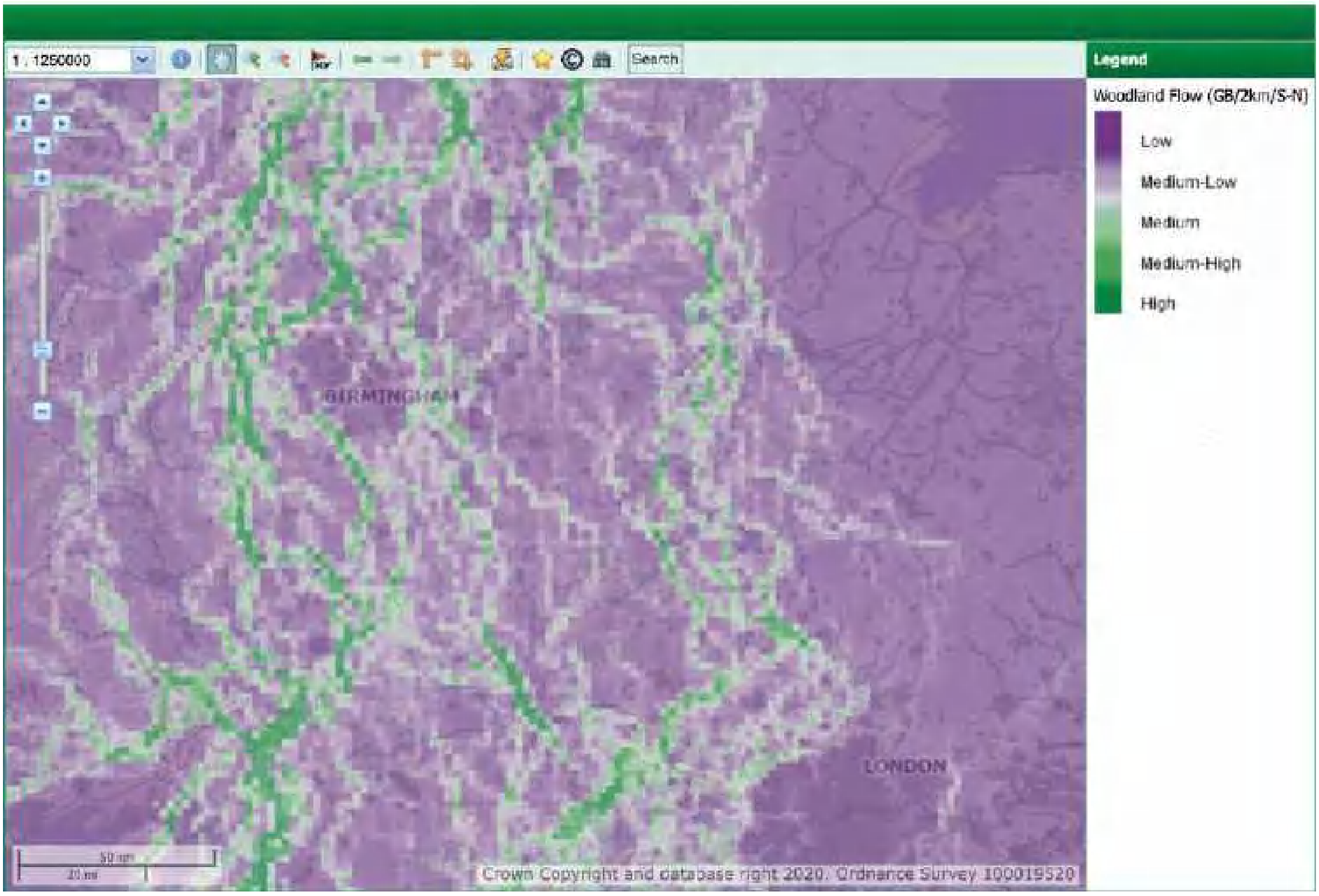
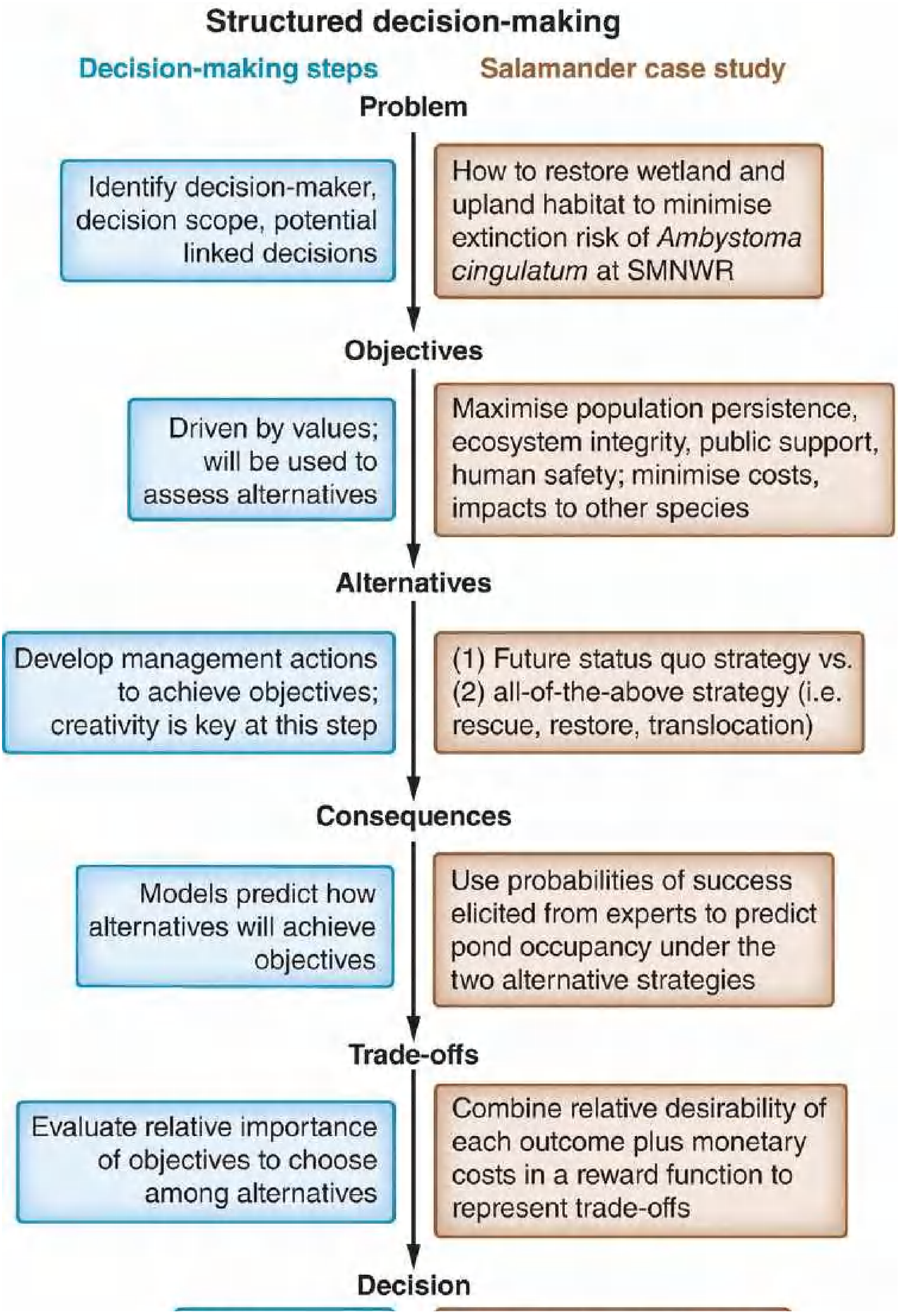


Figure 15.41 New building can be discouraged in habitats where the flow of species migrating to new habitats is greatest. The web service of a UK regional agency, Warwickshire County Council, comprising maps of important green infrastructure: http://maps.warwickshire.gov.uk/greeninfrastructure/. The layer of south–north flow of various species through woodland, calculated using the computer program Condatis, is shown overlain onto the UK Ordnance Survey map. It is a priority to preserve woodland, and hence discourage new building, in the routes where the flow is concentrated, for example both sides of Birmingham and in the Cotswolds to the south-west.

15.4.6 Decision analysis

In practice, many conservation decisions have to be taken in the absence of the sorts of dataset that are necessary to carry out a formal PVA. One way forward in such cases is to elicit the opinions of experts, often at meetings convened for the purpose, and to combine these where possible with such data as are available, using a variety of approaches described collectively as 'knowledge synthesis' techniques (Pullin *et al*., 2016). These often incorporate, or are combined with, structured decision making processes referred to as 'decision analysis' techniques, especially multicriteria decision analysis (Adem Esmail & Geneletti, 2018), which allow analyses of the performance of alternative conservation strategies to be combined with the preferences and priorities of stakeholders in a transparent way that can therefore be discussed and, where appropriate, modified, perhaps in the light of additional information. One example comes from the conservation of an endangered species of salamander, *Ambystoma cingulatum*, in Florida, USA (O'Donnell *et al*., 2017). A workshop was held, attended by conservationists and other biologists as well as representatives of United States Fish and Wildlife Service (USFWS), which has ultimate responsibility for the conservation of the species. The aim was particularly to consider the possible role of *ex situ* breeding programmes to supplement the declining populations, but also to consider how this might be combined with translocation of individuals and habitat restoration. The stages of the structured decision-making process are outlined in Figure 15.42. The process starts with an explicit statement of the underlying problem and the objectives that flow from that. Next comes a listing of the available alternatives and an assessment of their chances of success, guided both by expert opinion and available data. This then leads to an assessment of both the desirability and costs of those alternative outcomes, such that the trade-offs between, say, cost and the chance of success can be analysed – is it, for example, worth doubling the cost for a marginal improvement in the likely outcome? And finally, a decision is made. In the case of the salamander, a wide range of actions was implemented, including the development of captive breeding protocols, the establishment of captive populations and habitat modification, but also programmes to collect additional data that the workshop agreed were necessary for better informed decisions to be made.



Problem

Identify decision-maker, decision scope, potential linked decisions

How to restore wetland and upland habitat to minimise extinction risk of *Ambystoma cingulatum* at SMNWR

Objectives

Driven by values; will be used to assess alternatives

Maximise population persistence, ecosystem integrity, public support, human safety; minimise costs, impacts to other species

Alternatives

Develop management actions to achieve objectives; creativity is key at this step

(1) Future status quo strategy vs. (2) all-of-the-above strategy (i.e. rescue, restore, translocation)

Consequences

Models predict how alternatives will achieve objectives

Use probabilities of success elicited from experts to predict pond occupancy under the two alternative strategies

Trade-offs

Evaluate relative importance of objectives to choose among alternatives

Combine relative desirability of each outcome plus monetary costs in a reward function to represent trade-offs

Decision

Figure 15.42 Conservation decisions are often taken, ultimately, as the end point of a structured decision-making process, incorporating both the elicitation of expert opinions and the analysis of data. The general process is shown to the left and its application to the conservation of the salamander, *Ambystoma cingulatum*, at the St Marks National Wildlife Refuge (SMNWR), Florida, USA is shown to the right.

Source: After O'Donnell *et al.* (2017).

decision trees – the Sumatran rhinoceros

Clearly, a key step for the salamanders and in many other cases is the analysis of an optimum strategy in the light of the expert evidence. Decision trees are a technique that may be used to weigh up and then choose between alternatives. To illustrate the approach, we can take a classic example from a workshop on the Sumatran rhinoceros, *Dicerorhinus sumatrensis* (Maguire *et al.*, 1987). At the time, the species persisted only in small, isolated subpopulations in an increasingly fragmented habitat in Sabah (East Malaysia), Indonesia and West Malaysia, and perhaps also in Thailand and Burma – its range has further contracted since. There were only a few designated reserves, which were themselves subject to poaching, and only two individuals were held in captivity. A decision tree for arriving at a conservation strategy is shown in Figure 15.43, based on the estimated probabilities of the species becoming extinct within a 30-year period (equivalent to approximately two rhinoceros generations). The workshop was designed to generate a consensus among the experts on these probabilities. The tree was then constructed in the following way. The two squares are decision points: the first distinguishes between intervention on the rhinoceros' behalf and non-intervention (status quo); the second distinguishes the various management options. For each option, the line branches at a small circle. The branches represent alternative scenarios that might occur, and the numbers on each branch indicate the probabilities estimated for the alternative scenarios. Thus, for the status quo option, there was estimated to be a probability of 0.1 that a disease epidemic would occur in the next 30 years, and hence a probability of 0.9 that no epidemic would occur. If there was an epidemic, the probability of extinction in 30 years (pE) was estimated to be 0.95 whereas with no epidemic the pE was 0.85. The overall estimate of species extinction for an option, E(pE), is then given by: (probability of first option × pE for first option) + (probability of second option × pE for second option), which, for the status quo option, was 0.86. The values of pE and E(pE) for the various intervention options were estimated in a similar way. The final column in Figure 15.43 then lists the estimated costs of the various options.

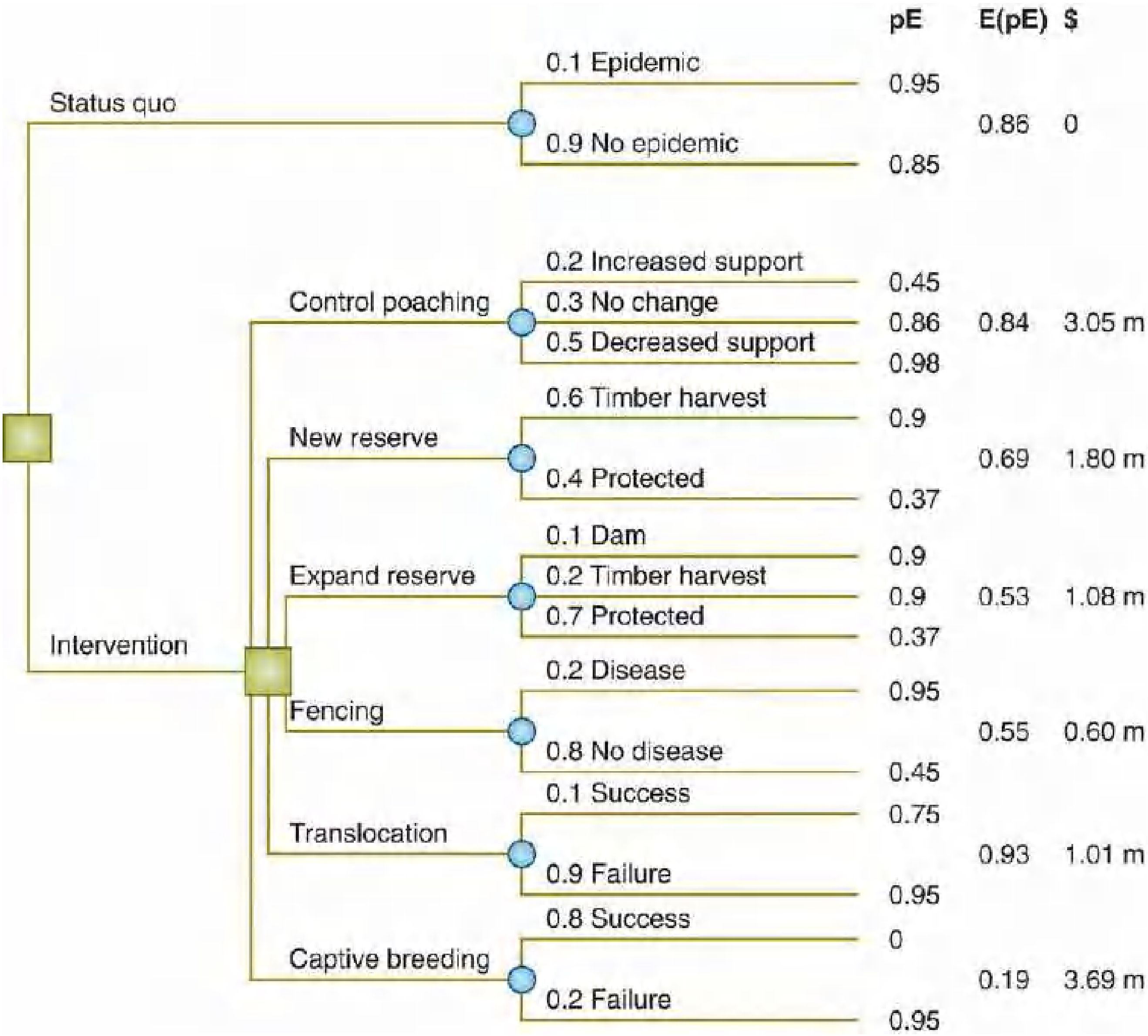


Figure 15.43 A decision tree for the management of the Sumatran rhinoceros gives rise the probability of extinction if various alternative actions are taken. Within the tree, the large green boxes indicate decision points, and the blue circles indicate random events. Probabilities of random events are estimated for a 30-year period; pE, probability of species extinction within 30 years; E(pE), expected value of pE for each alternative. Costs are present values of 30-year costs discounted at 4% per year; m, million.

Source: After Maguire *et al*. (1987).

evaluating management options

Consider two of the interventions. The first was the establishment of a captive breeding programme. Animals would have to be captured from the wild, increasing the pE if the programme failed to an expected 0.95. However, the pE would clearly drop to 0 if the programme succeeded (in terms of the continued persistence of the population in captivity). The cost, though, would be high, since it would involve the development of facilities and techniques in Malaysia and Indonesia (around $2.06 million) and the extension of those that already exist in the USA and Great Britain ($1.63 million). The probability of success was estimated to be 0.8. The overall E(pE) was therefore 0.19.

An alternative was to control poaching, either with more, less, or no change in the level of support. As Figure 15.43 shows, the E(pE) was much higher, 0.84, and the estimated cost not much less than the captive breeding programme.

Which was the best management option? The answer depends on what criteria we use to define ‘best’. Suppose we wanted simply to minimise the chances of extinction, irrespective of cost. The best option would then appear to be captive breeding. In practice, though, costs are most unlikely to be ignored. We would then need to identify an option with an acceptably low E(pE) but with an acceptable cost. In fact, $2.5 million was spent catching Sumatran rhinoceroses for captive breeding, but three died during capture, six died postcapture, and of 21 rhinoceroses taken into captivity none gave birth (Caughley, 1994). The latest IUCN Red List assessment (van Strien *et al.*, 2015) describes limited ongoing attempts at breeding programmes and still only two captive births with support for and coordination of anti-poaching teams being the primary conservation action.

We see from this, therefore, the strengths and weaknesses of subjective expert assessments and the techniques, like decision trees, that they use. The approach makes use of available data, knowledge and experience in a situation when a decision is needed and time for further research is unavailable. It explores the various options in a systematic manner, and it does not duck the regrettable but inevitable truth that unlimited resources will not be available. But in the absence of all necessary data, the recommended best option may simply be wrong, as it seems to have been for the Sumatran rhino. The experts seem to have been far too optimistic about the chances of success of captive breeding. Techniques for weighing up options are improving. An analysis of the decision tree that incorporates uncertainty around the estimated probabilities highlights the particular sensitivity of captive breeding to such uncertainty and downgrades its ‘utility’ accordingly (Regan *et al.*, 2005). Nonetheless, looking forward, decisions based on opinion rather than hard data will be inevitable – as will the mistakes that equally inevitably accompany them.

Chapter 16

Community Modules and the Structure of Ecological Communities

16.1 Introduction

Several earlier chapters (Chapters 8–13, especially) were focused on species interactions and the consequences of these interactions for the abundances of the species concerned. The emphasis, mostly, was on pairs of species – two competitors, a predator and a prey, and so on – even though this is clearly an oversimplification. First, and most obviously, no species pair ever exists in isolation: each is embedded within a wider community of species. But also, there are often additional 'hidden' participants accompanying these pairs. This was a point made several times in Chapter 13 on facilitation, and, for example, when two species compete, the items consumed are often other species that therefore play an integral part in what is superficially a two-party interaction. We pick up this point now as we turn to the structure of ecological communities, and the forces that mould that structure.

community modules

Holt (1997) introduced the concept of *community modules* – small groups of tightly interacting species. These are more than a bridge between the traditional domains of population ecology (single species and species pairs) and community ecology (many species). They are the building blocks from which whole communities are constructed, and studying them is arguably the most productive way of trying to understand the structure of communities as a whole, that is, the repeated patterns we see as we survey the communities around us. Some of the more important community modules are illustrated in Figure 16.1.

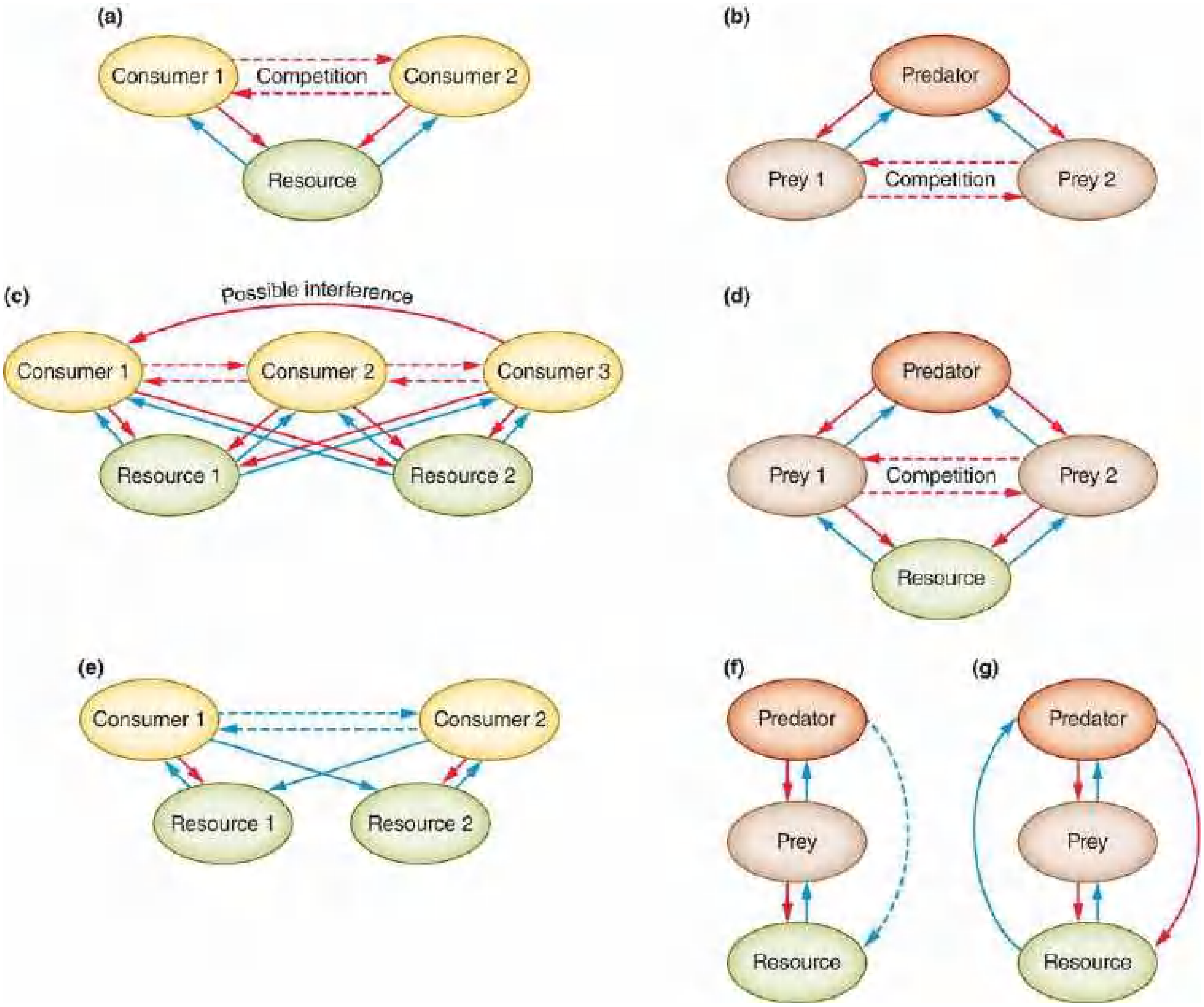


Figure 16.1 **Selected community modules.** In all cases arrows indicate the effect of one species on another: one competitor having a negative effect on another competitor, a predator having a negative effect on its prey, which in turn has a positive effect on the predator and so on. Negative effects are in red; positive effects are in blue. Direct effects have solid arrows; indirect effects have dashed arrows. (a) Exploitative competition. (b) Apparent competition. (c) Three competitors. (d) Two prey sharing both a predator and a resource. (e) Facilitation, of the sort found, for example, in mycorrhizas (see Section 13.9), where each partner provides a resource that the other partner might otherwise lack. (f) A three-level food chain. (g) Omnivory.

Two of the modules were introduced in Chapter 8 – those for exploitative competition and apparent competition (Figure 16.1a, b) – but reprising them here reminds us that they consist typically of three species, not two, and that they are part of a cast of modules that populate the stage on which community ecology is played out. Further modules take us through successive sections of this chapter. In Section 16.2 we ask what role community-wide interspecific competition plays in structuring communities. Figure 16.1c portrays the simplest community module addressing this question: not two species competing for one resource, but three species competing for two. Then, in Section 16.3, we ask what role grazing, predation and parasitism play, focusing on the diamond-shaped module comprising a predator, two prey and those prey's shared resource (Figure 16.1d). As we shall see, the role of predation depends crucially on the relative strengths of the interactions between the predator and its alternative prey and on the balance of competition between them for their shared resource. Finally, we step back to ask what can be said about the relative or combined roles of competition and predation, and also facilitation (Figure

16.1e), in structuring ecological communities, and indeed how these fit in with other, more stochastic processes such as disturbance and the colonisation of empty patches (Section 16.4).

Following this, in the next chapter, we turn to *food chains* (Figure 16.1f) and examine first the interactions that may be generated between species more than one trophic level apart in the chain, and then what may constrain the lengths of the chains. Finally, we look at *food webs* – larger and more complex than community modules, but small enough at least not to require an exhaustive list of all species present.

16.2 The influence of competition on community structure

When interspecific competition was discussed in Chapter 8, the emphasis was on species pairs. We were careful to acknowledge that niche differentiation is not the only route to competitor coexistence, and that competition could exert its influence not only as a contemporary ecological force, but also as an evolutionary force through which coexistence relied on the avoidance of competition, or as an historical ecological force through which species incapable of coexistence with residents had been driven to extinction, locally. These problems become, if anything, more acute when the perspective is broadened to look at community modules (for example, Figure 16.1c) and whole communities. Studies of competition between two species are almost always focused, *a priori*, on species that are good candidates to be competitors. But for whole communities, interspecific competition can be important even if most of the species pairs within it do not compete. We need to know what would count as evidence for the hand of competition in shaping the structure of an ecological community, and we need rigorous methods of evaluating that evidence. Neither has proved easy to develop. Broadly, there have been two approaches – focusing on the process (interspecific competition) itself, and focusing on patterns that may be generated by that process – and we deal with these in turn.

16.2.1 Demonstrable competition between species

literature reviews suggest competition is widespread ...

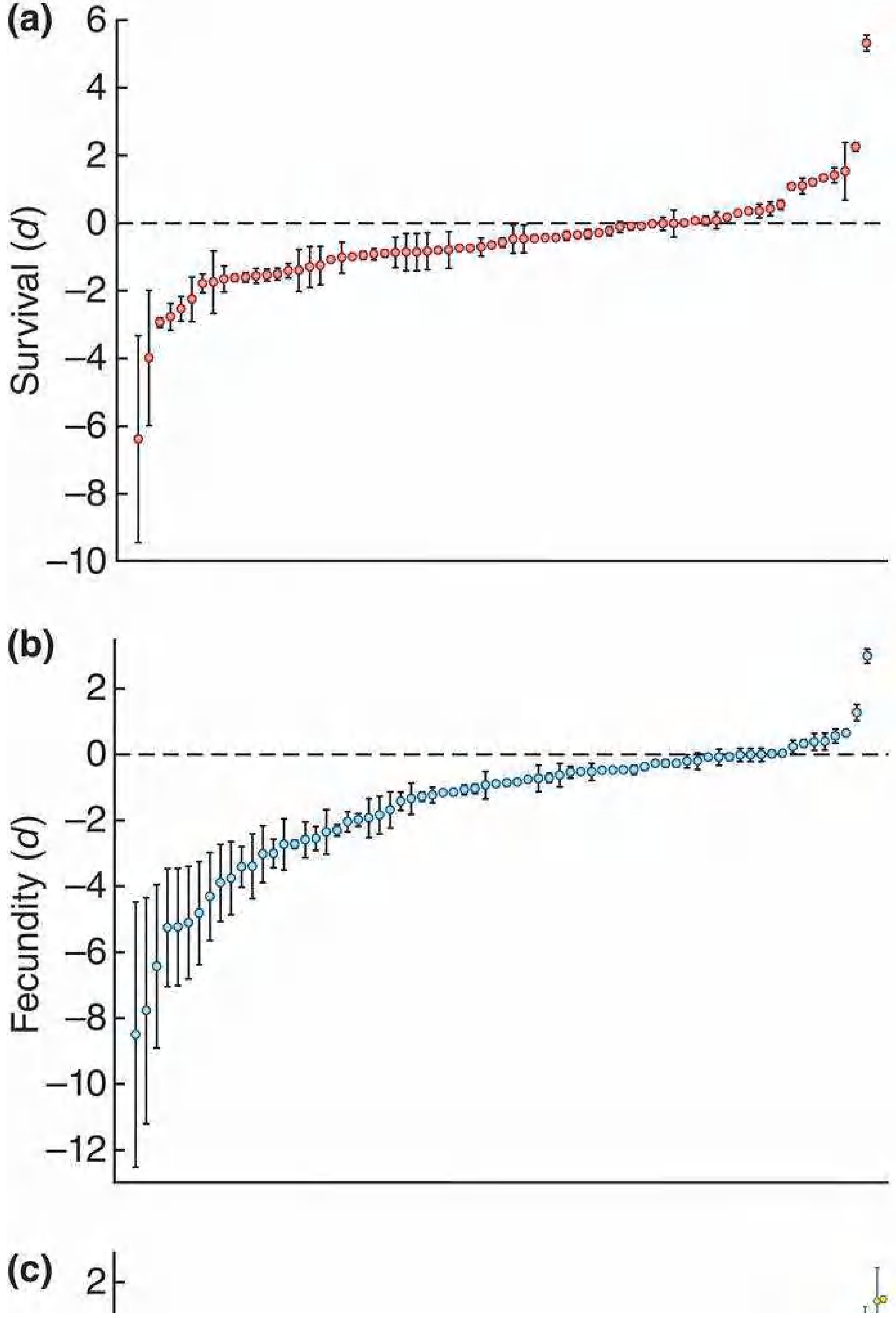
Arguably the most direct way of assessing the importance of interspecific competition in shaping real communities is by determining how common it is, and as noted previously, arguably the best way of doing this is through the analysis of experimental field manipulations. Two historically important surveys of field experiments on interspecific competition were published in 1983. Schoener (1983) examined the results of all the experiments he could find – 164 studies in all. Approximately equal numbers of studies had dealt with terrestrial plants, terrestrial animals and marine organisms, but studies of freshwater organisms were fewer, and amongst the terrestrial studies, most were concerned with temperate regions and relatively few dealt with phytophagous (plant-eating) insects. Any conclusions were therefore subject to limitations imposed by what ecologists had chosen to look at. Nevertheless, Schoener found that approximately 90% of the studies had demonstrated the existence of interspecific competition, irrespective of whether these were terrestrial, freshwater or marine. Moreover, if he looked at single species or small groups of species, rather than at whole studies that may have dealt with several groups of species, 76% showed effects of competition at least sometimes, and 57% showed effects in all the conditions under which they were examined. Connell's (1983) review was less extensive than Schoener's – 72 studies, dealing with a total of 215 species and 527 different experiments – but again, interspecific competition was demonstrated in most of the studies, in more than half of the species and in approximately 40% of the experiments. In contrast to Schoener, Connell found that interspecific competition was more prevalent in marine than in terrestrial organisms, and also that it was more prevalent in large than in small organisms. Taken together, then, Schoener's and Connell's reviews certainly seem to indicate that active, current interspecific competition is widespread.

... but are the data biased?

Connell also found, however, that in studies of just one pair of species, interspecific competition was almost always apparent, whereas with more species the prevalence dropped markedly (from more than 90% to less than 50%). This is to be expected, since, for example, if four species are arranged along a single niche dimension and all adjacent species competed with each other, this would still be only three out of six (50%) of all possible pairwise interactions. But it may also indicate biases in the particular pairs of species studied, and in the studies that are actually reported (or accepted by journal editors). It is highly likely that many pairs of species are chosen for study because they are ‘interesting’ (because competition between them is suspected) and if none is found this is simply not reported. Thus the results of surveys such as those by Schoener and Connell are likely to exaggerate, to an unknown extent, the frequency of competition.

the strength of competition is likely to vary from community to community

As previously noted, phytophagous insects were poorly represented in Schoener’s data, and earlier reviews of phytophagous insects fuelled a belief that competition is relatively rare in the group overall (Strong *et al*., 1984). Indeed, this conforms with the view that herbivores *in general* are seldom food-limited, and are therefore not likely to compete for common resources (Hairston *et al*., 1960; see Section 17.1.3). However, subsequent meta-analyses have suggested that the prevalence of interspecific competition in phytophagous insects is not greatly out of line with those in other groups (62%; 205 of 333 interactions), with demonstrable effects on survival, fecundity and abundance (Figure 16.2) (Kaplan & Denno, 2007).


(a)
Survival (*d*)
6
4
2
0
−2
−4
−6
−8
−10
(b)
Fecundity (*d*)
2
0
−2
−4
−6
−8
−10
−12
(c)
2

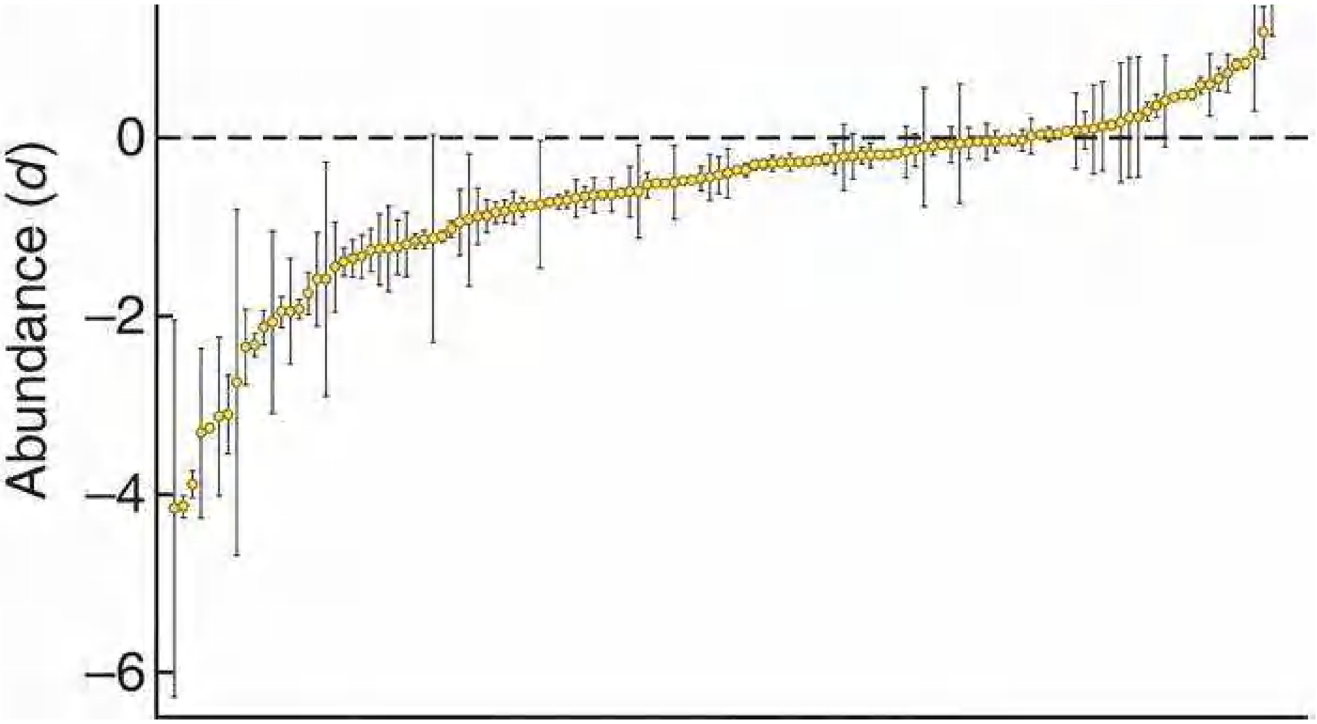


Figure 16.2 Interspecific competition ($d < 0$) affects phytophagous insects in more than 60% of studies. (a) Effects sizes from 69 studies that examined the effects of interspecific competition on the survival of phytophagous insects, arranged in order from the most negative to the most positive. The means and 95% CIs are shown for each study. Effects sizes are measured by Hedges' d statistic. A value of zero indicates no effect, and hence negative values imply interspecific competition. Mean $d = -0.45$, CI −0.73 to −0.14. (b) The equivalent for 70 studies on fecundity. Mean $d = -1.04$, CI −1.36 to −0.72. (c) The equivalent for 125 studies on abundance. Mean $d = -0.54$, CI −0.73 to −0.37.

Source: After Kaplan & Denno (2007).

16.2.2 The structuring power of competition

expectations from competition theory

Taken overall, therefore, field experiments confirm that contemporary interspecific competition is widespread. Clearly, competition cannot be dismissed as a structuring force. But the experiments also suggest that its prevalence varies with taxon, trophic level and habitat, and it is unlikely that we will ever be able to put an accurate figure on its prevalence. Indeed, even if we could, it is not clear how useful that figure would be, since as noted already, and as we saw in Chapter 8, the patterns that interspecific competition may be expected to generate are largely the same whether they are the result of contemporary ecological, historical ecological or evolutionary forces. In other words, the intensity of current competition may sometimes be linked only weakly to the structuring power of competition within the community. This in turn invites us to ask what a community would look like if interspecific competition was shaping it or had shaped it in the past, and then to examine real communities to see whether they conform to these predictions. Three predictions in particular emerge readily from conventional competition theory (see Chapter 8), though as we shall see (in, for example, Section 16.2.4) 'conventional' theory may itself be challenged.

1. Potential competitors that coexist in a community should, at the very least, exhibit niche differentiation (discussed in Sections 16.2.3 and 16.2.4).

2. This niche differentiation will often manifest itself as morphological differentiation, often known as character displacement, insofar as an organism's morphology typically reflects the resources it exploits or the manner in which it exploits them (see Section 16.2.5).
3. Potential competitors with little or no niche differentiation are much less likely to coexist. Their distributions should therefore be negatively associated: each should tend to occur only where the other is absent (Section 16.2.6).

16.2.3 Evidence from community patterns: niche differentiation

Patterns of niche differentiation in animals and plants were outlined in Chapter 8. On the one hand, resources may be utilised differentially. This may express itself directly within a single habitat, or, if the resources are themselves separated spatially or temporally, as a difference in the microhabitat, geographic distribution or temporal appearance of the species. Alternatively, species' competitive abilities may vary with environmental conditions. This too can express itself as either microhabitat, geographic or temporal differentiation, depending on the manner in which the conditions themselves vary. Note, though, that even the simple step up to a three-competitor two-resource community module (Figure 16.1c) opens up new possibilities. One is *niche complementarity*, whereby niche differentiation overall is achieved by species that are not differentiated along one dimension being differentiated along another.

niche complementarity – anemone fish in Papua New Guinea

An example is provided by a study of a number of species of anemone fish near Madang in Papua New Guinea (Elliott & Mariscal, 2001). This region has the highest reported species richness of both anemone fishes (nine) and their host anemones (ten), forming behavioural mutualistic relationships (Section 13.3) in which the fish keep the anemones clean of debris while being provided with a protected habitat that other fish cannot occupy because of the anemones' stinging tentacles. Anemones seem to be a limiting resource for the fishes, because almost all the anemones were occupied, and when some were transplanted to new sites they were quickly colonised and the abundance of adult fish increased. However, each individual anemone is typically occupied by individuals of just one fish species, because the residents aggressively exclude intruders, except that aggressive interactions are less frequently observed between anemone fish of very different sizes, suggesting perhaps that different sized species of fish are utilising the resources provided by the anemones in different ways. Surveys at three replicate reef sites in four zones (nearshore, mid-lagoon, outer barrier reef and offshore reef: Figure 16.3a) showed that each anemone fish was primarily associated with a particular species of anemone and each showed a characteristic preference for a particular zone (Figure 16.3b). However, it was only when anemone species, zone and fish size were considered together that the fish were fully differentiated in their resource use. Different anemone fish that lived with the same anemone were typically associated with different zones. Thus, for example, *Amphiprion percula* occupied the anemone *Heteractis magnifica* in nearshore zones, while *A. perideraion* occupied *H. magnifica* in offshore zones. But the small anemone fish species (*A. sandaracinos* and *A. leucokranos*) were able to cohabit the same anemone with larger species. Such niche complementarity giving rise to niche differentiation overall is clearly consistent with what would be expected of communities moulded by competition.

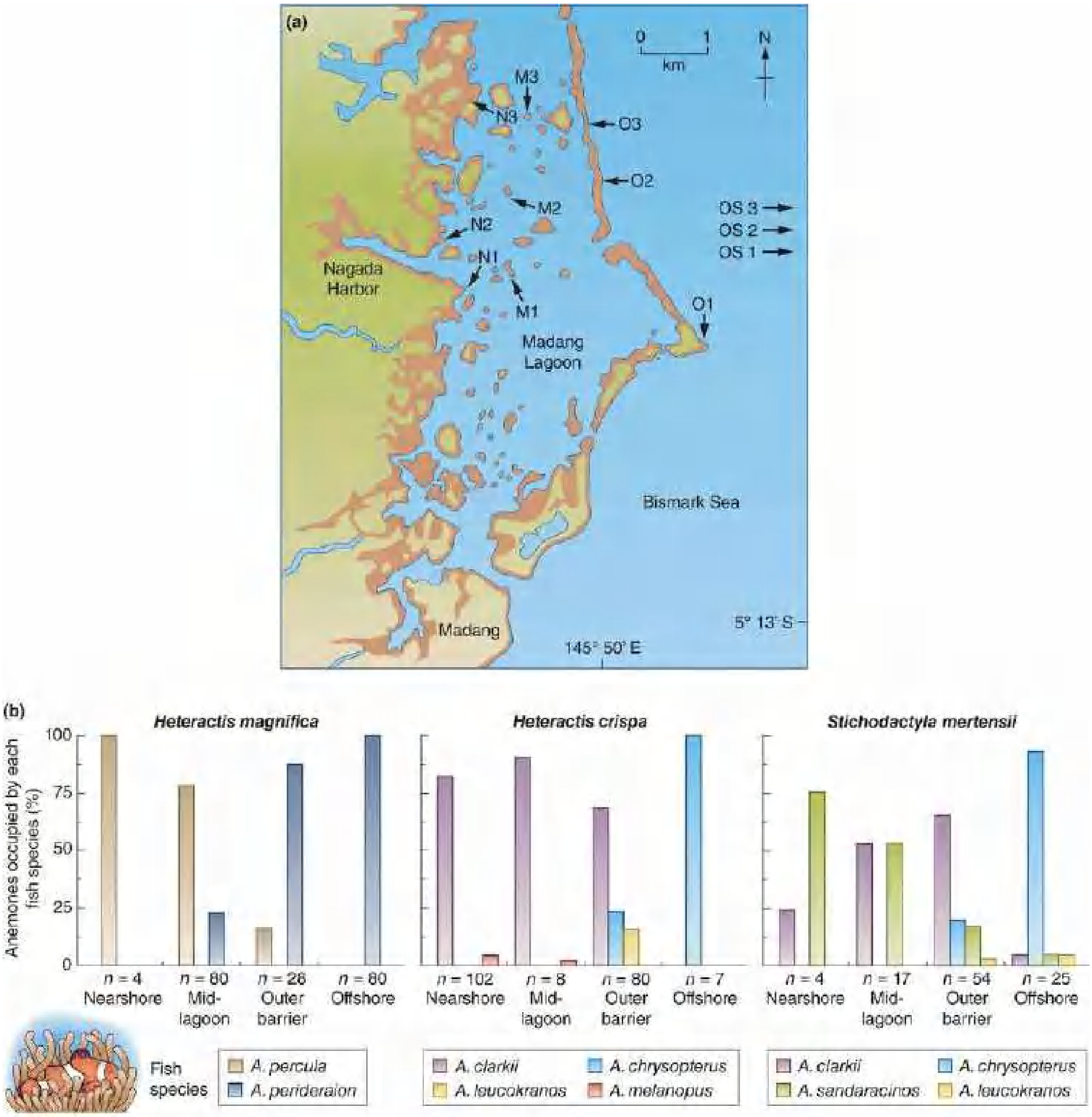


Figure 16.3 Niche complementarity in anemone fish is apparent both in terms of reef zone occupied and fish size. (a) Map showing the location of three replicate study sites in each of four zones within and outside Madang Lagoon (N, nearshore; M, mid-lagoon; O, outer barrier reef; OS, offshore reef). The blue areas indicate water, orange represents coral reef and green tint represents land. (b) The percentage of three common species of anemone (*Heteractis magnifica*, *H. crispa* and *Stichodactyla mertensii*) occupied by different anemone fish species (*Amphiprion* spp., in key below) in each of four zones. The number of anemones censused in each zone is shown by *n*.

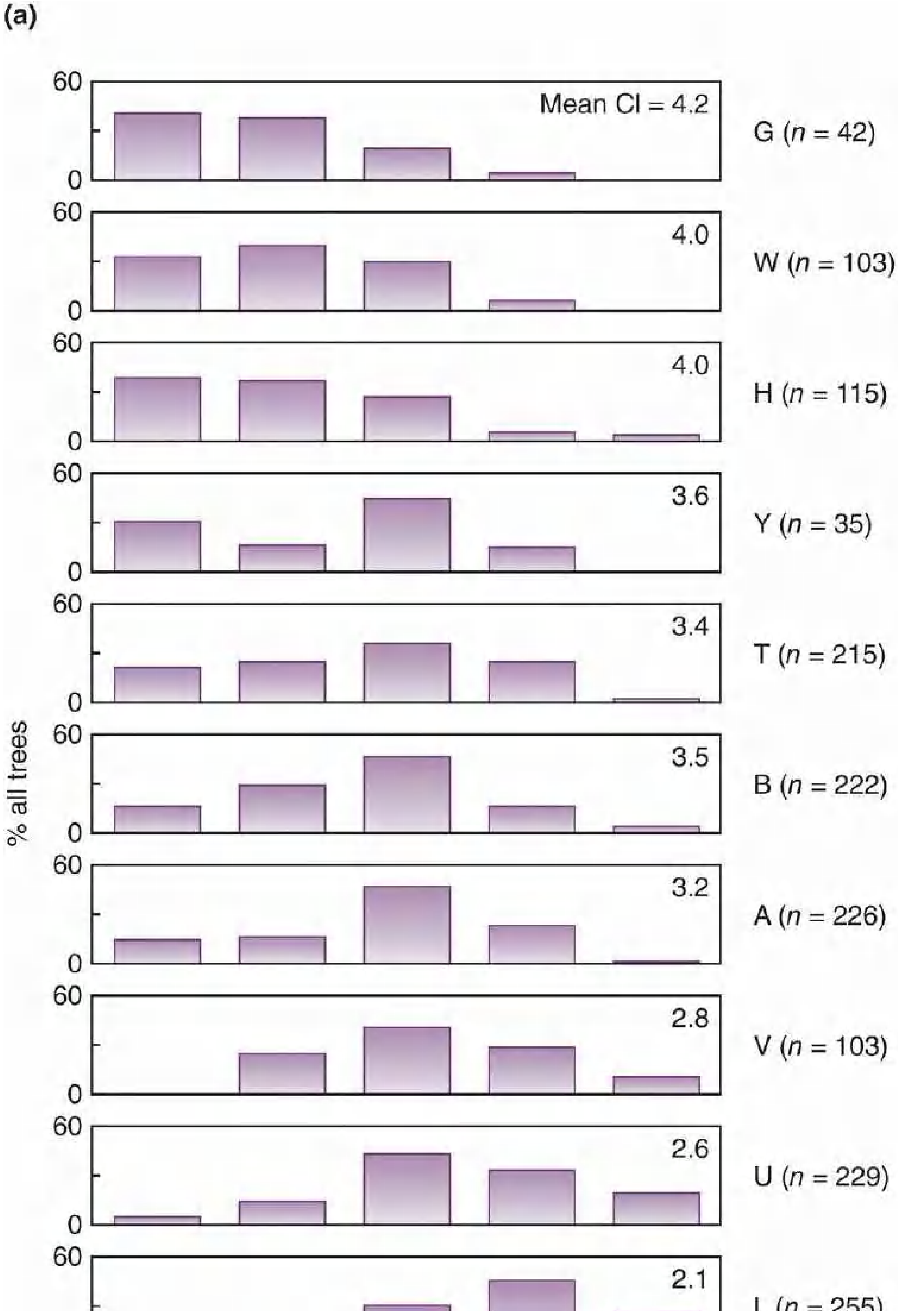
Source: After Elliott & Mariscal (2001).

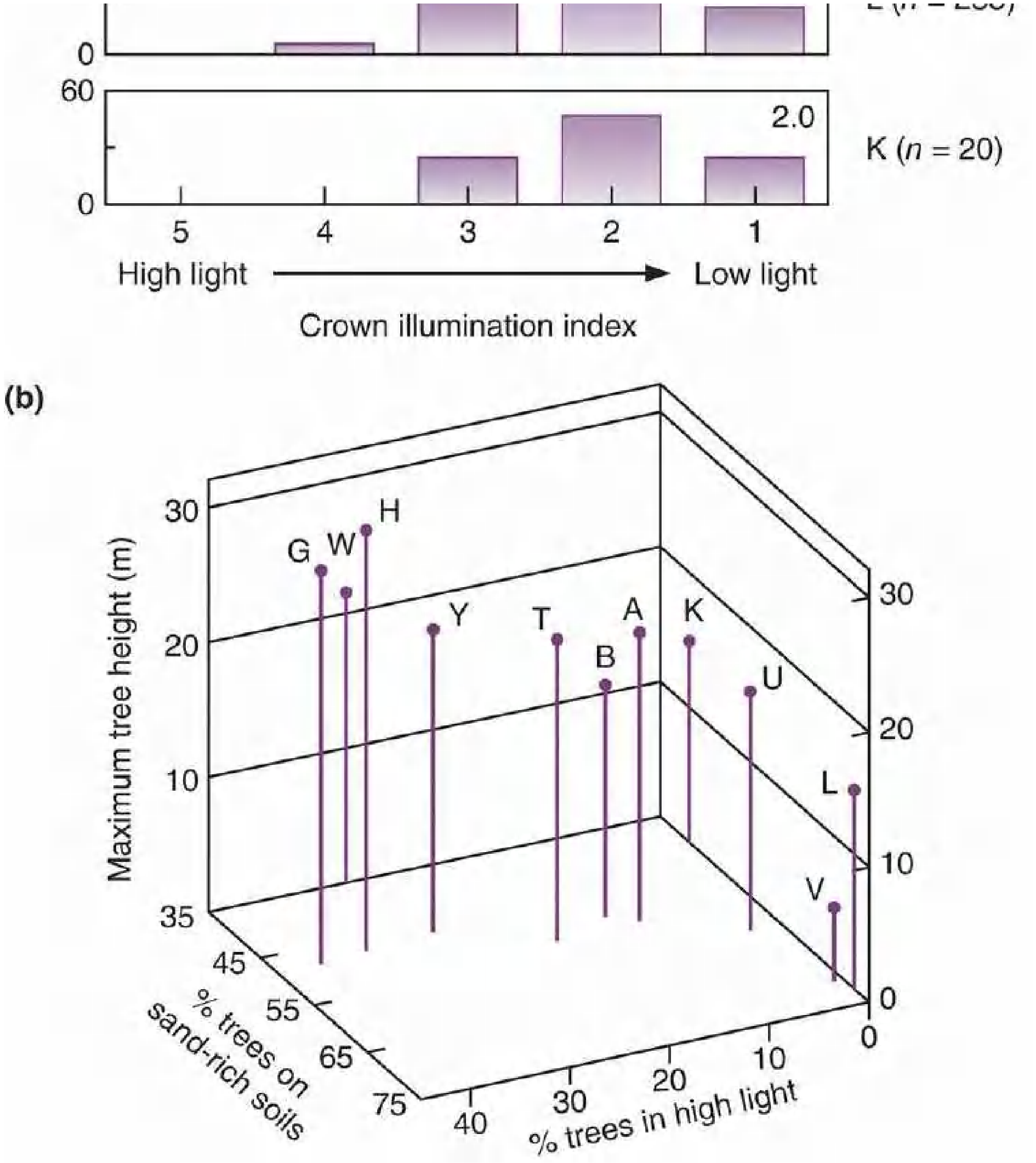
competition and guilds

A further point illustrated by the anemone fish is that they can be considered to be a *guild* – *a* group of species that exploit the same class of environmental resources in a similar way (Root, 1967). If interspecific competition is to occur at all, or if it has occurred in the past, then it will be most likely to occur, or to have occurred, within guilds. But this does *not* mean that guild members necessarily compete or have necessarily competed: the onus is on ecologists to demonstrate that this is the case. Complementary differentiation along several dimensions has also been reported for guilds as diverse as lizards (Schoener, 1974), bumblebees (Pyke, 1982), bats (McKenzie & Rolfe, 1986), rainforest carnivores (Ray & Sunquist, 2001) and tropical trees (Davies *et al*., 1998), as described next.

trees in Borneo differentiated in space

This second example of niche complementarity, then, is provided by a study in Borneo of 11 tree species in the genus *Macaranga* that showed marked differentiation in light requirements, from extremely high light-demanding species such as *M. gigantea* to shade-tolerant species such as *M. kingii* (Figure 16.4a). The average light levels intercepted by the crowns of these trees tended to increase as they grew larger, but the ranking of the species did not change as they grew. The shade-tolerant species were smaller and persisted in the understory, rarely establishing in disturbed microsites (e.g. *M. kingii*), in contrast to some of the larger, high-light species that are pioneers of large forest gaps (e.g. *M. gigantea*). Others were associated with intermediate light levels and can be considered small gap specialists (e.g. *M. trachyphylla*). Also, though, the *Macaranga* species were differentiated along a second niche gradient, with some species being more common on clay-rich soils and others on sand-rich soils (Figure 16.4b). This differentiation may be based on nutrient availability (generally higher in clay soils) and/or soil moisture availability (possibly lower in the clay soils because of thinner root mats and humus layers). In any case, we see evidence again of niche complementarity among the *Macaranga* species, partly related to horizontal heterogeneity in resources (light levels in relation to gap size, distribution of soil types) and partly to vertical heterogeneity (height achieved, depth of root mat). Species with similar light requirements differed in terms of preferred soil textures, especially in the case of the shade-tolerant species.


(a)
% all trees
60
0
Mean CI = 4.2
G (n = 42)
4.0
W (n = 103)
4.0
H (n = 115)
3.6
Y (n = 35)
3.4
T (n = 215)
3.5
B (n = 222)
3.2
A (n = 226)
2.8
V (n = 103)
2.6
U (n = 229)
2.1



[Figure 16.4](#) **Niche complementarity in *Macaranga* trees in Borneo.** (a) Percentage of individuals in each of five crown illumination classes for 11 *Macaranga* species (sample sizes in parentheses). (b) Three-dimensional distribution of the 11 species with respect to maximum height, the proportion of stems in high light levels (class 5 in (a)) and proportion of stems in sand-rich soils. Each species of *Macaranga* is denoted by a single letter: G, *gigantea*; W, *winkleri*; H, *hosei*; Y, *hypoleuca*; T, *triloba*; B, *beccariana*; A, *trachyphylla*; V, *havilandii*; U, *hullettii*; L, *lamellata*; K, *kingii*.

Source: After Davies *et al*. (1998).

tundra plants in Alaska differentiated in time

Intense competition may also be avoided by partitioning resources in time. For example, tundra plants growing in nitrogen-limited conditions in Alaska were differentiated in their timing of nitrogen uptake, as well as in the soil depth from which nitrogen was extracted and its chemical form. To trace how the species differed in their uptake of different nitrogen sources, McKane *et al.* (2002) injected three chemical forms labelled with the rare isotope ^{15}N (inorganic ammonium, nitrate and organic glycine) at two soil depths (3 and 8 cm) on two occasions (24 June and 7 August). Concentration of the ^{15}N tracer was measured in each of five common tundra plants seven days after application. The five plants proved to be well differentiated in their use of nitrogen sources (Figure 16.5). Cottongrass (*Eriophorum vaginatum*) and the cranberry bush (*Vaccinium vitis-idaea*) both relied on a combination of glycine and ammonium, but cranberry obtained more of these forms early in the growing season and at a shallower depth than cottongrass. The evergreen shrub *Ledum palustre* and the dwarf birch (*Betula nana*) used mainly ammonium but *L. palustre* obtained more of this form early in the season while the birch exploited it later. Finally, the grass *Carex bigelowii* was the only species to use mainly nitrate. Apparent niche complementarity can be seen along three niche dimensions.

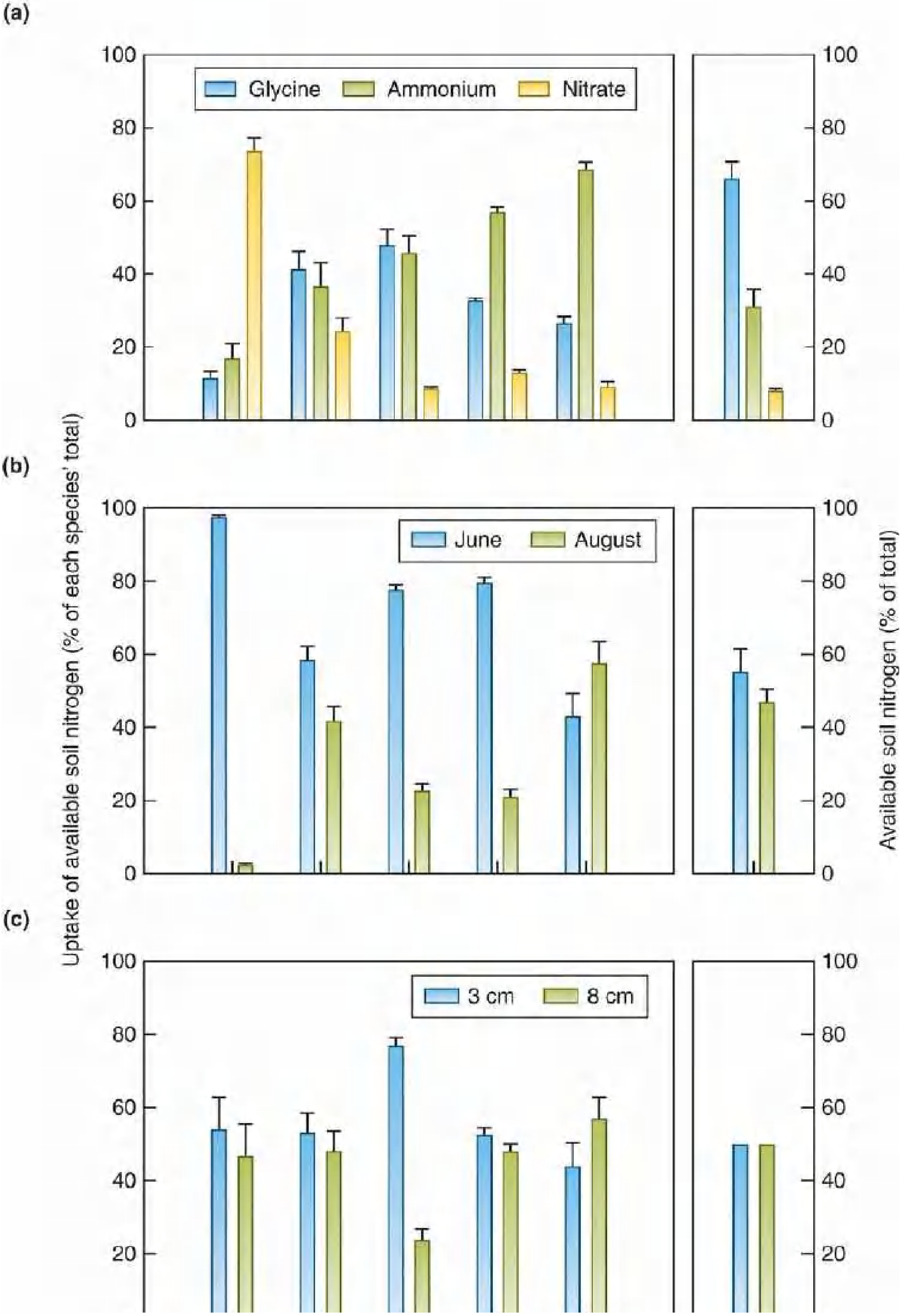

(a)
Glycine
Ammonium
Nitrate
(b)
June
August
(c)
3 cm
8 cm
Uptake of available soil nitrogen (% of each species' total)
Available soil nitrogen (% of total)
100
80
60
40
20
0

Figure 16.5 **Niche complementarity in tundra plants.** Mean uptake of available soil nitrogen (± SE) in terms of (a) chemical form, (b) timing of uptake and (c) depth of uptake by the five most common species in tussock tundra in Alaska. Data are expressed as the percentage of each species' total uptake (left panels). Also shown is the percentage in each case of the total pool of nitrogen available in the soil (right panels).

Source: After McKane *et al.* (2002).

16.2.4 Niche differentiation – apparent or real? Null and neutral models

Many cases of apparent resource partitioning such as these have been reported. It is likely, however, that studies failing to detect such differentiation have tended to go unpublished. It is always possible, of course, that these 'unsuccessful' studies are flawed and incomplete, and that they have failed to deal with the relevant niche dimensions, but a number have been sufficiently beyond reproach to raise the possibility that in certain cases, at least, resource partitioning is not an important feature. Remember that we saw in Section 8.6.1 that, in theory at least, two species with complete niche overlap could still coexist indefinitely as long as their competitive abilities were also equal.

null hypotheses are intended to ensure statistical rigour

We can also criticise what might be seen as too ready a tendency to interpret 'mere differences' as confirming the importance of interspecific competition. Are these differences large enough or regular enough to be distinguishable from what might be found simply by chance among a set of species? This problem led to an approach known as *null model analysis* (Connor & Simberloff, 1979 ; Gotelli & Ulrich, 2012). Null models are models of actual communities that retain certain of the characteristics of their real counterparts, but aim to reassemble the components at random, specifically excluding the consequences of biological interactions. Such analyses are thus attempts to follow a much more general approach to scientific investigation, namely the construction and testing of *null hypotheses*. The idea (familiar to most readers in a statistical context) is that the data are rearranged into a form (the null model) representing what the data would look like in the absence of the phenomenon under investigation (in this case species interactions, particularly interspecific competition). Then, if the actual data show a significant statistical difference from the null hypothesis, the null hypothesis is rejected and the action of the phenomenon under investigation is strongly inferred.

a null model of food resource use in lizard communities ...

We can see the null model approach in action by going back to an early example, less sophisticated perhaps than some later studies, but showing the general principles and advantages clearly. Lawlor (1980) looked at 10 North American lizard communities, consisting of four to nine species, for which he had estimates of the amounts of each of 20 food categories consumed by each species in each community (data from Pianka, 1973). A number of null models of these communities were created, which were then compared with their real counterparts in terms of their patterns of overlap in resource use. If competition is or has been a significant force in determining community structure, the niches should be spaced out, and overlap in resource use in the real communities should be less than predicted by the null models.

Lawlor’s analysis was based on the ‘electivities’ of the consumer species, where the electivity of species *i* for resource *k* was the proportion of the diet of species *i* which consisted of resource *k*. Electivities therefore ranged from 0 to 1. These electivities were in turn used to calculate, for each pair of species in a community, an index of resource-use overlap, which itself varied between 0 (no overlap) and 1 (complete overlap). Finally, each community was characterised by a single value: the mean resource overlap for all pairs of species present.

The null models were of four types, generated by four ‘reorganisation algorithms’ (RA1–RA4, Figure 16.6). Each retained a different aspect of the structure of the original community whilst randomising the remaining aspects of resource use.

- RA1 retained the minimum amount of original community structure. Only the original number of species and the original number of resource categories were retained. Observed electivities (including zeros) were replaced in every case by random values between 0 and 1. This meant that there were far fewer zeros than in the original community. The niche breadth of each species was therefore increased.
- RA2 replaced all electivities, *except zeros*, with random values. Thus, the qualitative degree of specialisation of each consumer was retained (i.e. the number of resources consumed to any extent by each species was correct).
- RA3 retained not only the original qualitative degree of specialisation but also the original consumer niche breadths. No randomly generated electivities were used. Instead, the original sets of values were rearranged. In other words, for each consumer, all electivities, both zeros and non-zeros, were randomly reassigned to the different resource types.
- RA4 reassigned only the non-zero electivities. Of all the algorithms, this one retained most of the original community structure.

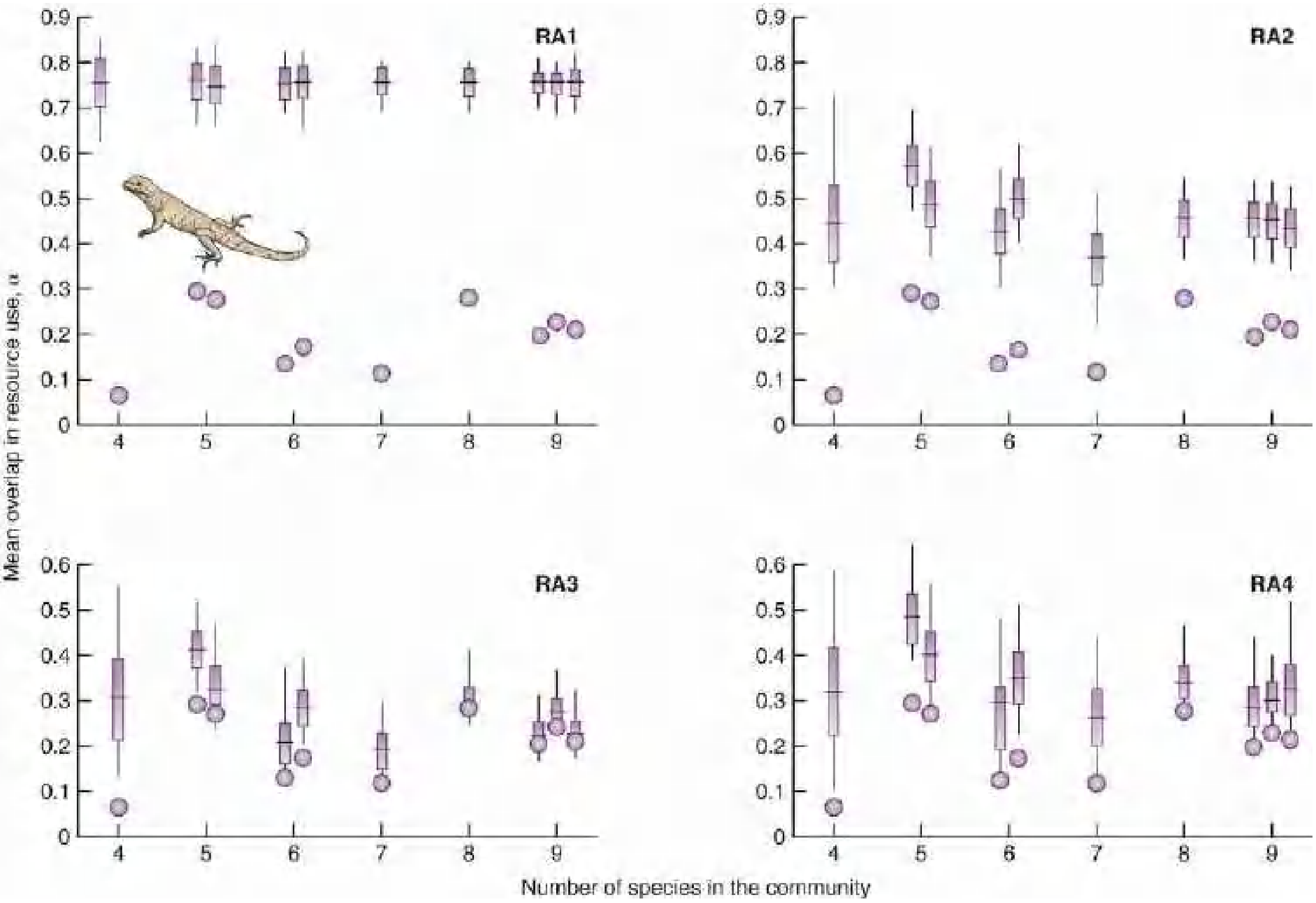


Figure 16.6 Null modelling supports a role for competition in structuring lizard guilds because actual overlap is generally less than would be expected by chance. The mean indices of resource-use overlap for each of Pianka's (1973) 10 North American lizard communities are shown as solid circles. These can be compared, in each case, with the mean (horizontal line), standard deviation (vertical rectangle) and range (vertical line) of mean overlap values for the corresponding set of 100 randomly constructed communities. The analysis was performed using four different reorganisation algorithms (RAs), as described in the text.

Source: After Lawlor (1980).

Each of the four algorithms was applied to each of the 10 communities. In every one of these 40 cases, 100 'null model' communities were generated and the corresponding 100 mean values of resource overlap were calculated. If competition was important in the real community, these mean overlaps should exceed the real community value. The real community was therefore considered to have a *significantly* lower mean overlap than the null model ($P < 0.05$) if five or fewer of the 100 simulations gave mean overlaps less than the real value.

... in which the lizards appear to pass the test

The results are shown in Figure 16.6. Increasing the niche breadths of all consumers (RA1) resulted in the highest mean overlaps (significantly higher than the real communities). Rearranging the observed non-zero electivities (RA2 and RA4) also always resulted in mean overlaps that were significantly higher than those actually observed. With RA3, on the other hand, where all electivities were reassigned, the differences were not always significant. But in all communities, the algorithm mean was higher than the observed mean. In the case of these lizard communities, therefore, the observed low overlaps in resource use suggest that the niches are segregated (not just different by chance), and that interspecific competition plays an important role in community structure.

Similar results have been found in other guilds, for example for the diets of mammalian herbivores in the West African savanna, where niche overlap was lower than those generated by any of the RAs in both the wet and the dry season (Djagoun *et al.*, 2016), and for the temporal patterns of activity throughout the day for rodents in Mexico, which also had less overlap in activity periods than expected by chance (Castro-Arellano & Lacher, 2009).

Hubbell's Unified Neutral Theory

The null model approach was taken a stage further by Hubbell's (2001) Unified Neutral Theory of Biodiversity and Biogeography. He imagined a local community of focal interest and a metacommunity (a community of communities) that provides a well-mixed source of potential immigrants to that and all other local communities. (This, and other metacommunity approaches, are discussed further in Section 18.7.) The local community contains a fixed number of individuals, but at each time step, one of these individuals dies and is replaced either by an individual in the local community giving birth or by an immigrant from the metacommunity – immigration occurring with probability m. Both occur 'at random' in the sense that they depend only on the relative abundances of the different species and not on any species-specific characteristics such as their competitive ability. At the metacommunity level, species turnover is driven by a process of 'random' death and birth like that in the local community, except that immigration is replaced by speciation, occurring with probability v. This sequence of death, birth and immigration/speciation is then continued until the local community has reached equilibrium, at which point its characteristics can be taken as a reflection of this purely 'neutral' process and not of competitive or any other biological interactions.

The original neutral theory has been further developed and some of its unrealistic assumptions eliminated (see, for example, Rosindell *et al.*, 2011), but its core assumption of an absence of species interactions driven by species differences has remained, and while unrealistic, that assumption is, of course, there by design. It allows us to see how the world might look if this were the case, and invites us to interpret deviations from that prediction as potential evidence for those species interactions that have been left out. Conversely, if the predictions of neutral theory do not differ significantly from observed patterns, or are just as consistent with them as are those of more complex models including such species interactions, then this may call into question the need for invoking species differences in understanding community structure. As with null models in general, its proponents do not believe that the world is neutral. They simply propose, much more modestly, that 'Neutral theory, an ensemble of different neutral models of community assembly, is useful in ecological research' (Rosindell *et al.*, 2012).

In fact, most comparisons of the predictions of neutral theory with real datasets have involved *species–abundance distributions*, that is, plots that trace, for a community, the changing frequency of species abundances when abundance classes are arranged in rank order from most to least abundant (but see Rosindell *et al.* (2012) for other possible tests). This, and related *rank–abundance diagrams* – where the abundances of individual species are plotted, ranked from most to least abundant, containing exactly the same information – are discussed further in Section 18.2.2. For now, though, we can note that a variety of models, some purely statistical, some incorporating interspecific competition, can be used to generate theoretical species–abundance distributions, and the correspondence of these, and of the distribution generated by neutral theory, can be compared with observed distributions of real communities (McGill *et al.*, 2007). When such comparisons have been made, the result has typically been to find that the theoretical models in general fit data reasonably well and cannot be statistically rejected (which is not to say that they are positively supported), and that the fits of neutral and other models are not statistically different from one another. Some, for example McGill *et al.* (2006), have concluded that such tests therefore fail to support neutral theory, which is arguably unfair, since its proponents have neither sought nor expected 'support', as such. In other cases – see, for example,

Figure 16.7 – neutral models have been better than niche-based models in accounting for patterns in field data, which has been used to argue in favour of a role for neutral processes in structuring communities (Chisholm & Pacala, 2010). This, though, goes arguably too far, since such correspondence is more correctly consistent with, rather than supportive of, such a role. Perhaps the best stance is to regard neutral models as null models *not* for there being a role for niche differentiation in the structuring of communities, but rather for the *detectability* of niche differentiation in particular types of empirical dataset (Rosindell *et al.*, 2012). From this standpoint, neutral theory, like other null models, helps us guard against an overeagerness to see the hand of competition in structuring communities, and drives us towards the compilation of datasets that allow the assessment of its role to be less equivocal.

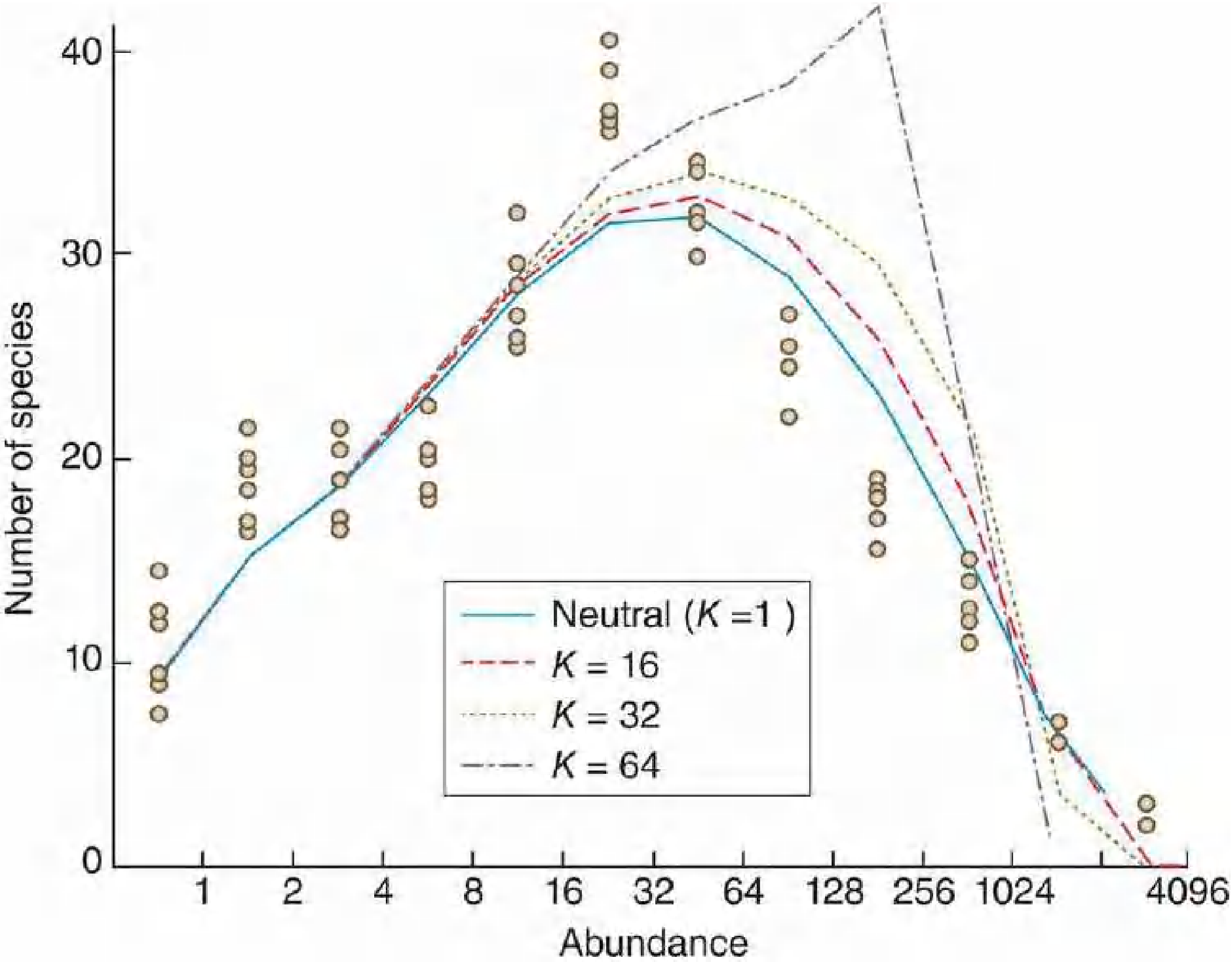


Figure 16.7 Neutral models are better than niche-based models in their ability to describe community patterns in tropical forests, consistent with the hypothesis that competition in this case is not a strong structuring force. Species–abundance relationships (the number of species in ranked abundance categories) for a 50-hectare tropical forest plot on Barro Colorado island in the Panama Canal. The points are data collected during six censuses from 1982 to 2005 with overall means of 232 species and 21 060 individuals. The curves are fitted based on models assuming K equally sized niches, as indicated. $K = 1$ is therefore a neutral model (just one niche that all species share), and increasing values of K indicate an increasing role for niche differentiation. The fit to the data when $K = 1$ is good, and those for $K < 16$ too similar to this to be shown as distinct lines. But the fit to the data gets worse as K increases and is poor by $K = 64$.

Source: After Chisholm & Pacala (2010).

16.2.5 Evidence from morphological patterns – community-wide character displacement

Hutchinson’s ‘1.3 rule’ and its demise

Where niche differentiation is manifested as morphological differentiation, perhaps because different species eat prey of different sizes and hence have mouthparts of different sizes, the spacing out of niches within a guild can be expected to have its counterpart in the spacing out of morphological differences – an example of the ‘limiting similarity’ amongst competitors that we discussed in Section 8.6.2, and something that has been described as *community-wide character displacement*, though the differences may be the result of species exclusion rather than species evolution, as would be the case with true (ecological) character displacement.

Hutchinson (1959) was the first to catalogue examples of sequences of weights or lengths of potential competitors. He even suggested a ‘rule’, in which not only was there a minimum difference between the sizes of adjacent species compatible with coexistence (limiting similarity), but also a tendency for adjacent species to have weight ratios of approximately 2.0 or length ratios of approximately 1.3 (the cube root of 2.0). This rule was adopted enthusiastically for a time. However, a null model reconsideration of a range of apparently supportive examples (Simberloff & Boecklen, 1981) reached a conclusion typical of such analyses, namely that competition (as evidenced by limiting similarity) is important but not of overriding or universal importance: just seven of 18 claims of a minimum size difference that they examined were sustained, and just seven of 21 claims of ratio constancy. The latter, especially, should not surprise us, since, as we saw in Section 8.6, models of competition do not predict specific values for niche separation (and hence for size ratios) that might apply across a range of organisms and environments.

Both ecological character displacement (between two species) and community-wide character displacement have recovered credibility following difficult periods when they were challenged by null model analyses, with rigorous statistical testing now an integral part of most studies (Dayan & Simberloff, 2005 ; Stuart & Losos, 2013). Solid examples can certainly be found (see, for example, Figure 16.8), but many studies aimed at establishing community-wide character displacement (limiting similarity) fail to find it. A meta-analysis of plant studies, for example, found that in only 12% of 1966 tests were traits significantly more distinct than predicted by null models (Götzenberger *et al*., 2012). We must also remember, on the other hand, that failure to establish the existence of community-wide character displacement may mean only an absence of sufficient statistical power, or that competition is important but does not express its influence in this way.

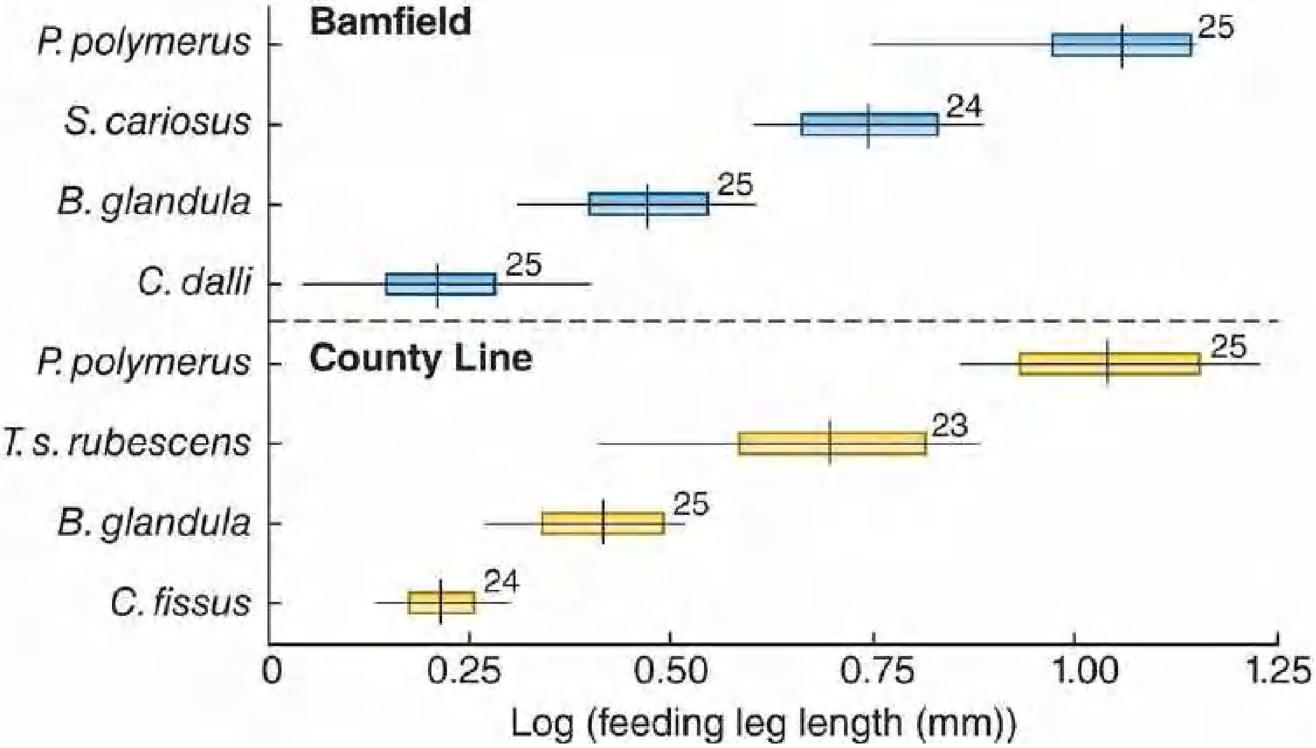


Figure 16.8 Community-wide character displacement in barnacles: **leg length distributions were more evenly dispersed than expected by chance.** Feeding leg lengths are shown for species collected at two distant sites in western North America: Bamfield, British Columbia, Canada and County Line, California, USA. The species are *Pollicipes polymerus*, *Semibalanus cariosus*, *Balanus glandula*, *Chthamalus dalli*, *Tetraclita squamosa rubescens* and *Chthamalus fissus*. In each case the means (vertical lines), standard deviations (bars), ranges (horizontal lines) and sample sizes for the field data are shown. These data were compared with null models run 10 000 times (Bamfield $P = 0.0001$, County Line $P = 0.0064$). Note that species composition differed at the two sites.

Source: After Marchinko *et al.* (2004).

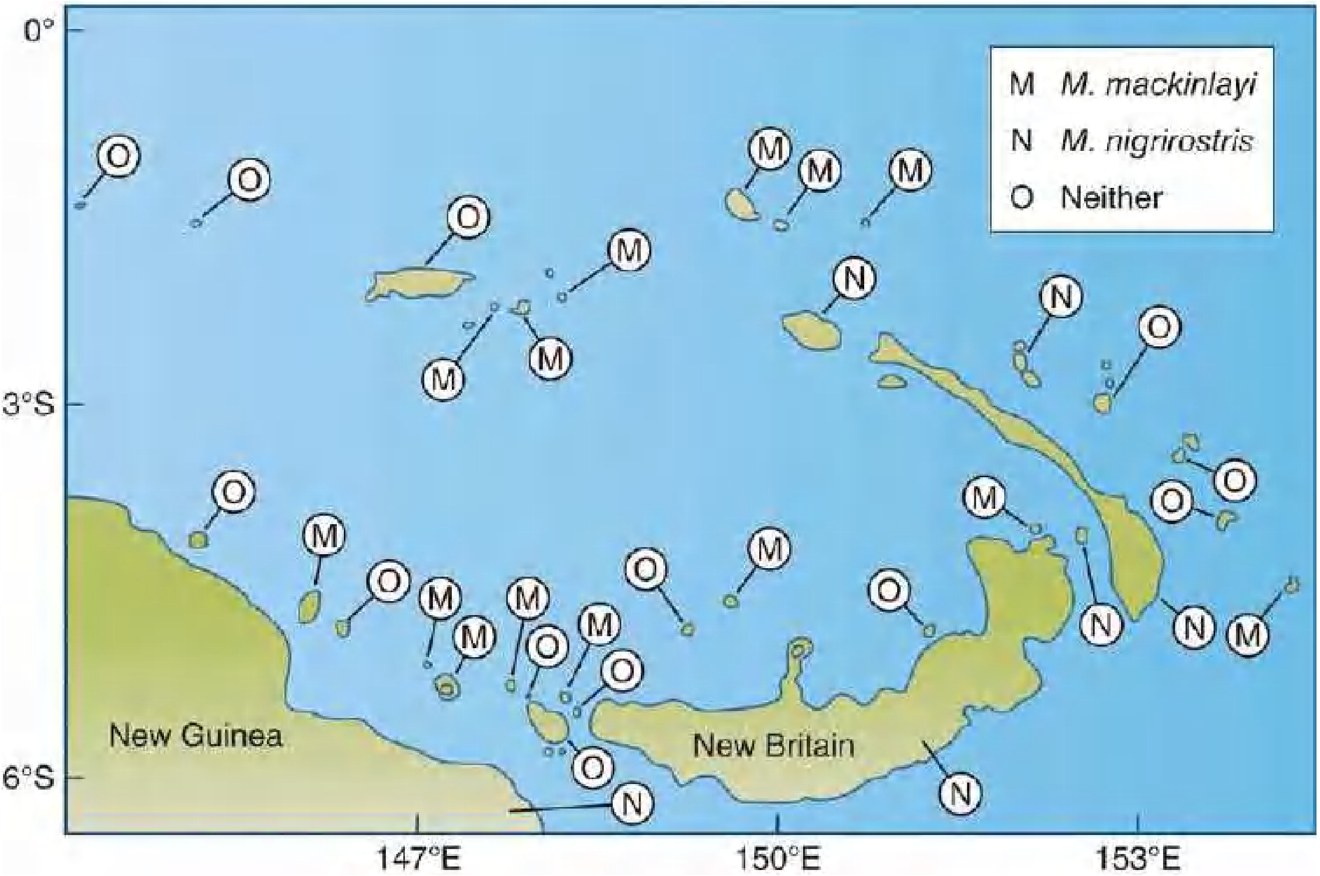


Figure 16.9 **Checkerboard distribution of two small *Macropygia* cuckoo-dove species in the Bismarck region.** Islands whose pigeon faunas are known are designated as M (*M. mackinlayi* resident), N (*M. nigrirostris* resident) or O (neither species resident). Note that most islands have one of these species, no island has both and some islands have neither.

Source: After Diamond (1975).

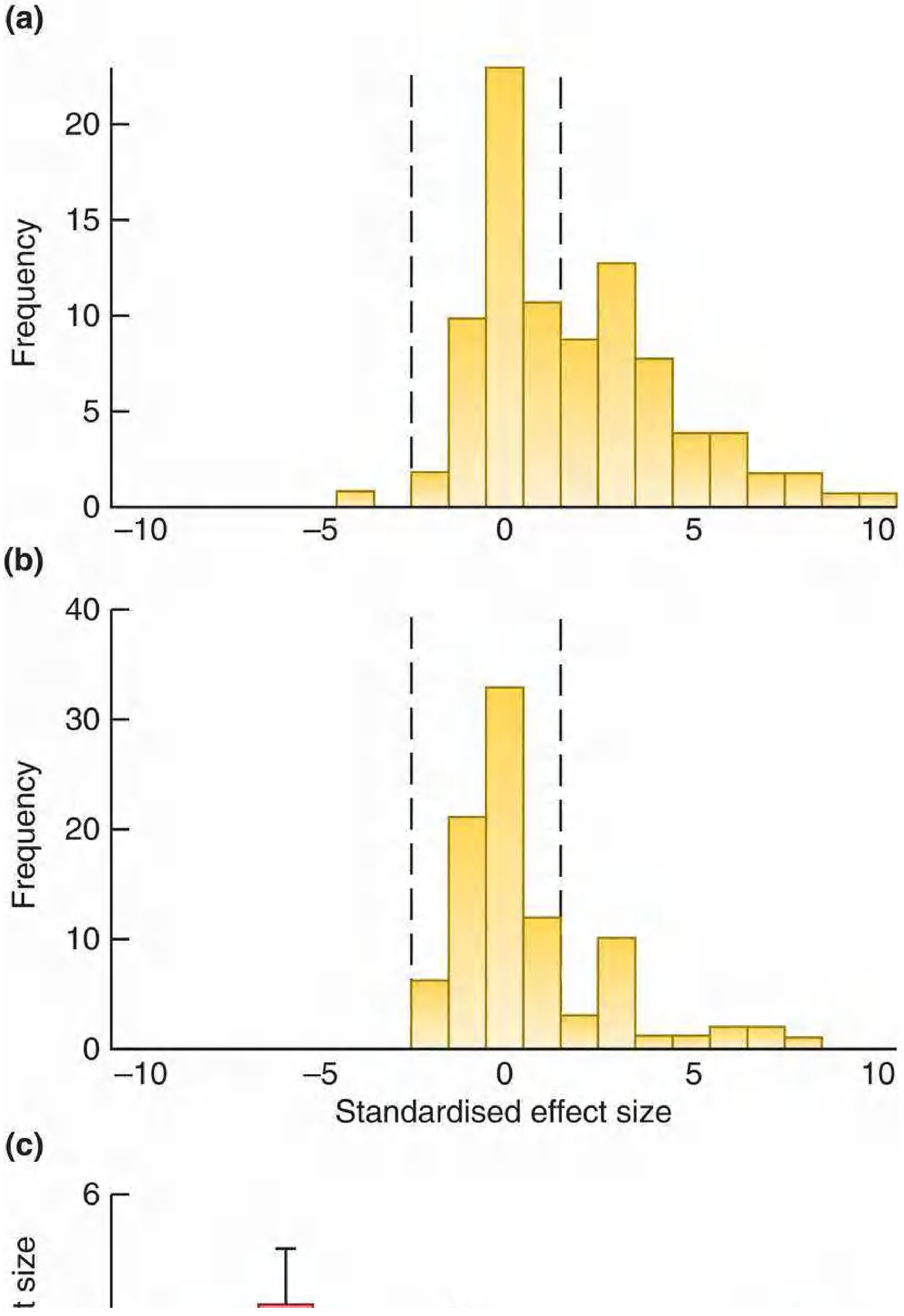
16.2.6 Evidence from negatively associated distributions

‘checkerboard’ distributions in island birds

Patterns in distribution have often been used as evidence for the importance of interspecific competition. The tradition is usually traced back to Diamond’s (1975) survey of the land birds living on the islands of the Bismarck Archipelago off the coast of New Guinea. His most striking evidence came from distributions that Diamond referred to as ‘checkerboards’. In these, two or more ecologically similar species (i.e. members of the same guild) have mutually exclusive but interdigitating distributions such that any one island supports only one of the species (or none at all). Figure 16.19 shows this for two small, ecologically similar cuckoo-dove species: *Macropygia mackinlayi* and *M. nigrirostris*.

Diamond’s conclusions were very quickly challenged by a null model critique (Connor & Simberloff, 1979), and as we have seen elsewhere, null models have subsequently become a standard tool. A null model in this case involves comparing the pattern of species co-occurrences at a suite of locations with what would be expected if each species were distributed at random, with an excess of negative associations then being consistent with a role for competition in determining community structure.

Gotelli and McCabe (2002) carried out a meta-analysis of null-model-based studies of various taxonomic groups in 96 datasets that reported the distribution of species assemblages across sets of replicated sites. For every real dataset, 1000 randomised versions were simulated, and standardised effect sizes (SESs) were computed for two indices. An SES is given by $(I_{\text{obs}} - I_{\text{exp}})/\sigma_{\text{exp}}$, where I_{obs} is the observed value of the index for a community, I_{exp} is the mean of the index from the 1000 simulations and σ_{exp} is their standard deviation. The indices were, first, the number of perfect checkerboard distributions in the community, and second the average of the C score for the whole community (Stone & Roberts, 1990). The C score for each species pair is given by $C = (R_i - S)(R_j - S)$, where R_i and R_j are then numbers of sites occupied by the ith and jth species, respectively, and S is the number of sites occupied by both species. C therefore measures the degree to which species co-occur, taking its maximum value for a checkerboard distribution ($S = 0$; $C = R_i .R_j$) and a minimum value of zero ($R_i = R_j = S$). The null hypothesis in each case was that the mean SES should be zero (real communities not different from simulated communities) and that 95% of the values should lie between −2.0 and +2.0. This was rejected in both cases (Figure 16.10a, b). Furthermore, plants and homeothermic vertebrates tend to have higher SESs for the C score, indicating stronger tendencies towards negative species associations than the poikilotherms (invertebrates, fish and reptiles) have, with the exception of ants (Figure 16.10c).


(a)
Frequency
0
5
10
15
20
−10
−5
0
5
10
(b)
Frequency
0
10
20
30
40
−10
−5
0
5
10
Standardised effect size
(c)
6
t size

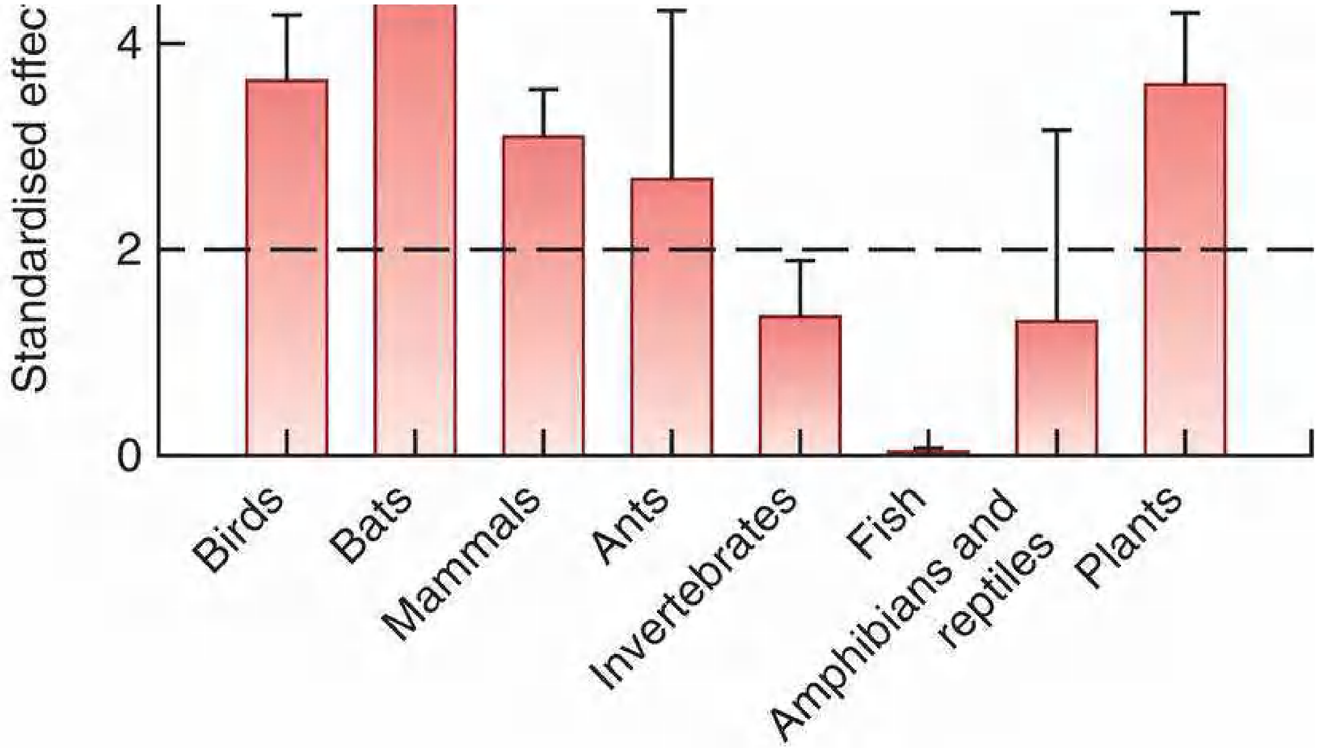


Figure 16.10 The distributions of species pairs are often negatively associated (consistent with a role for competition), especially in some taxonomic groups. Frequency histograms for standardised effect sizes measured for 96 presence–absence matrices taken from the literature in the case of (a) the *C* score and (b) the number of species pairs forming perfect checkerboard distributions. (c) Standardised effect sizes for the *C* score for different taxonomic groups. The dashed lines indicate effects sizes of 2.0, which is the approximate 5% significance level.

Source: After Gotelli & McCabe (2002).

As Gotelli and McCabe (2002) acknowledge, while these results support predictions that emerge if we assume a powerful role for competition, they do not (and cannot) prove it. Some, at least, of the patterns may represent 'habitat checkerboards' in which different species have affinities to non-overlapping habitats, or 'historical checkerboards', where species co-occur infrequently because of restricted dispersal following their speciation in different places. Indeed, if the habitat differences reflect past evolutionary changes then we could be witnessing effects of 'the ghost of competition past'. We need to find ways of incorporating information on habitat characteristics and dispersal limitations that can help distinguish alternative explanations of these segregated distributions (D'Amen *et al*., 2018). Nonetheless, these results, and others like them, add further weight to a widespread role for competition in structuring communities.

16.2.7 Intransitive competition

A seemingly natural, default assumption is that interspecific completion is 'hierarchical'. If species 1 is a stronger competitor than species 2, which is a stronger competitor than species 3, then surely species 1 will outcompete species 3? However, it is apparent from the appropriate community module (see Figure 16.1c) that, for example, consumer 1 could outcompete species 2 by being a specialist on resource 1, and consumer 2 could outcompete species 3 by being a specialist on resource 2, but that species 3 could also interfere directly with species 1, perhaps by producing allelochemicals, and so outcompete it. Such patterns (1 beats 2 beats 3 beats 1, or 'rock–paper–scissors' after the children's game) are characteristic of *intransitive competition*, and it is clear that since no single species can outcompete all others, intransitivity tends to enhance species coexistence and thus species richness within a community. Other routes to intransitivity are discussed by Soliveres and Allan (2018). Soliveres *et al*. (2018) analysed competition experiments carried out on vascular plants, mosses, saprotrophic fungi, aquatic protists and soil bacteria, and calculated the proportion of competitive interactions that were intransitive, either by examining the whole communities studied in the experiments, or by focusing on all the possible triplets in each case. This latter metric is perhaps easiest to understand and relate to the community module discussed earlier (see Figure 16.1c). They found that 38% of triplets were intransitive in the mosses, 19% in the vascular plants, 16% in the protists and around 5% in the bacteria and fungi. Clearly, among some groups at least, intransitivity is not uncommon.

appraisal of the role of competition

The prevalence of interspecific competition in nature can be judged only imperfectly from the inevitably limited number of studies that have been undertaken. The communities chosen may not be typical. The ecologists observing them have usually been specifically interested in competition, and they may have selected appropriate, ‘interesting’ systems. Nonetheless, taken overall, the evidence suggests that interspecific competition is a possible and indeed a plausible explanation for many aspects of the organisation of many communities – but it is not so often a proven explanation. Null models are valuable in helping us guard against the temptation to see competition in a community simply because we are looking for it, but they may often leave us unable to reject the possibility that the patterns we see have some alternative explanation: often pure chance – ecological drift – rather than biological interactions between the species.

16.3 The influence of predation on community structure

In turning to the influence of predation on community structure, it is useful to return to our gallery of community modules (see Figure 16.1). Two are particularly relevant. The first is the diamond-shaped module with a predator preying on two prey species that also share (and compete for) a common resource (see Figure 16.1d). This emphasises first that the range of possible influences is the same whether the ‘predator’ is a grazer, a true predator or a parasite – in all cases the signs of the links in the module are the same. But if we now relax the assumption that the *strengths* of those links are also all the same, we begin to see how the role of predators may vary (Figure 16.11). For example, the effects of a predator are likely to depend on whether it is selective in its predation (the selected prey species may be eliminated) or is a non-selective generalist (the abundance of both may be reduced, promoting their coexistence) (Figure 16.11a, b). The effects will depend, too, on whether a preferred prey is itself a strong competitor or not (Figure 16.11a, c). Indeed, there are far too many variations on the diamond-shaped theme for them all to be included in Figure 16.11, but recognising this range of variations on the common theme will help make sense of the series of examples to which we turn next. The second relevant community module here is omnivory (see Figure 16.1g), where a predator preys on more than one of the lower trophic levels, which may themselves be linked by predator–prey interactions.

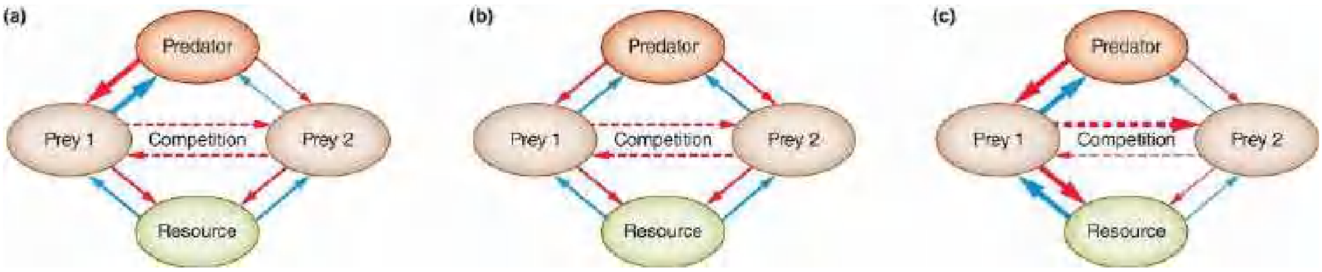


Figure 16.11 Shared-predator community modules with varying interaction strengths. (a) The predator shows a strong preference for prey type 1, and so may eliminate it. (b) The predator is a generalist with no strong preference and so may promote the coexistence of the prey. (c) The predator shows a strong preference for prey type 1, but prey type 1 is also the stronger competitor. Hence predation may, again, promote coexistence among the prey.

parasites may drive vulnerable host species extinct

The most obvious and straightforward effect predation (whether by true predators, grazers or parasites) can have on community structure is the removal of their prey population. Thus, taking parasites as an example, the extinction of nearly 50% of the endemic bird fauna of the Hawaiian Islands has been attributed in part to the introduction of bird pathogens such as malaria and bird pox (van Riper *et al*., 1986); and probably the largest single change wrought in the structure of communities by a parasite has been the destruction of the chestnut (*Castanea dentata*) in North American forests, where it had been a dominant tree over large areas until the introduction of the fungal pathogen *Endothia parasitica*, probably from China.

APPLICATION 16.1 A parasitic threat to Darwin's finches

A further potential threat of host loss from a community due to a parasite comes from the introduction of the parasitic nest fly, *Philornis downsi*, to the Galápagos Islands, probably in the mid-1990s. The Galápagos are iconic for the influence Charles Darwin's visit there had on the development of his ideas, and now for the associated focus they have become for worldwide conservation. By virtue of their isolation, the unique biota of the Galápagos is especially vulnerable to invaders, and *P. downsi* has been implicated in declines in the abundance of several of Darwin's finches (see Section 1.3.2), especially the mangrove finch, *Camarhynchus heliobates*, and the medium tree finch, *C. pauper* (Koop *et al*., 2016). To assess future risks, and what might be done to mitigate those risks, a population viability model was constructed for another of Darwin's finches, the medium ground finch, *Geospiza fortis*, where sufficient parameter estimates were available. Some of the results of this study are shown in Figure 16.12.

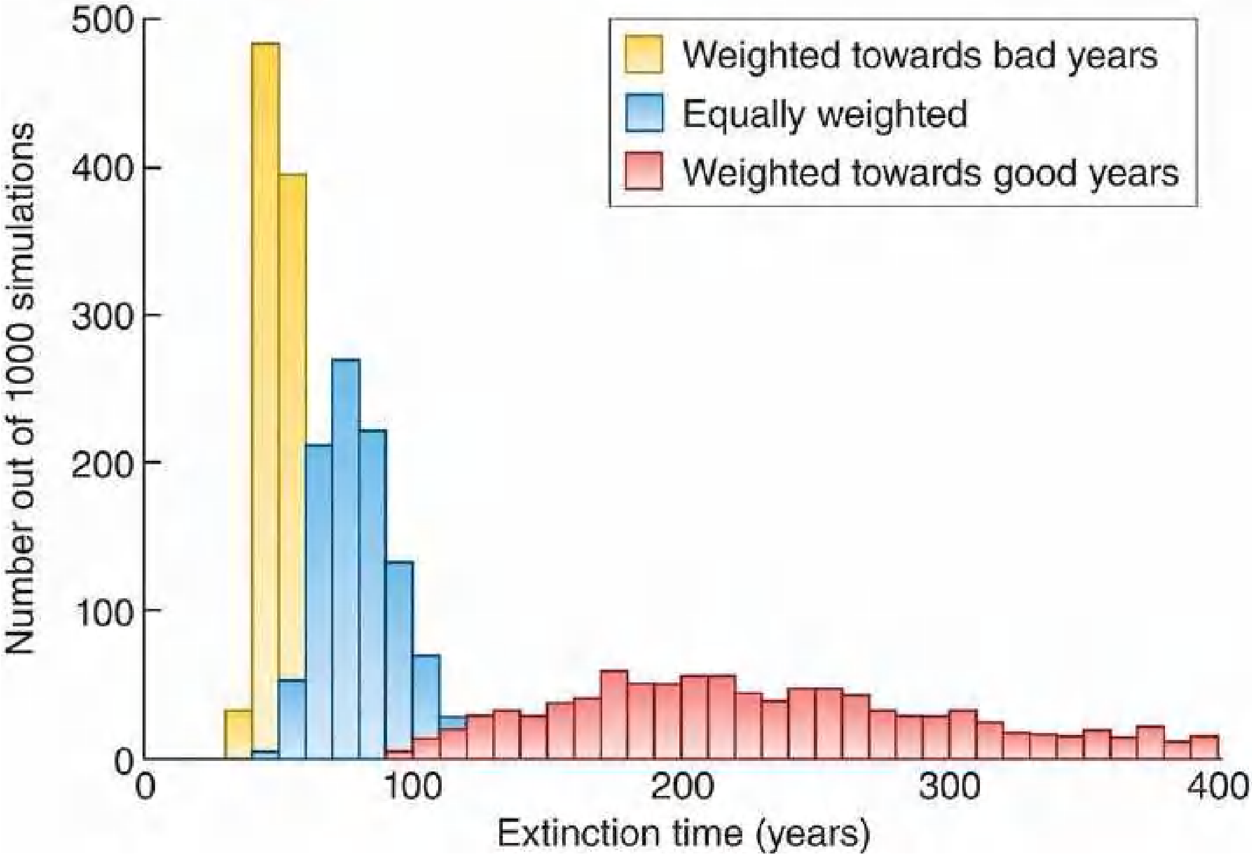


Figure 16.12 Predicted extinction times for the medium ground finch on Santa Cruz island, Galápagos. These times are dangerously high unless models are weighted towards 'good' years in the dataset. For each of the three scenarios indicated, results are shown for 1000 simulations in which the population growth rate and expected time to extinction were modelled for the finch when its fledging success was being reduced by the parasitic nest fly, *Philornis downsi*.

Source: After Koop *et al*. (2016).

Data were collected from five years on the island of Santa Cruz on the fledgling success of the birds from nests infested with the fly and from nests treated for the fly. Breeding success is also profoundly influenced by the weather, which varied over the five years, leading in turn to wide variation in the success of treated birds (0.29 fledglings per nest in 2009 compared with 3.08 in 2012). The population growth rate of the finch was therefore calculated in 1000 simulations of its annual cycle of survival and reproduction under each of three scenarios: data from all five years were given equal weight, or most weight was given to the bad years for fledging (especially 2009 and 2010), or most weight given to the good years (especially 2012 and 2013). These growth rates allowed the estimation in turn of the expected time of the population to extinction. Only when most weight was given to the good years was there a reasonable expectation of long-term persistence (Figure 16.12). However, with equal weighting, and especially when most weight was given to the bad years, the model predicted a serious risk of extinction within the next century. This is bad news for the medium ground finch, but arguably even worse news for the mangrove finch, for example, whose numbers are already much lower. On the other hand, a sensitivity analysis applied to the model found that these outcomes could be improved greatly by even a quite modest reduction in levels of fly infestation. Achieving such reductions should clearly be a conservation priority.

predator-mediated coexistence on a rocky shore

Predation, however, can also increase species richness. Probably the most seminal study of the role of predation on community structure was that carried out by Paine (1966) in the rocky intertidal zone of sheltered shores on the Pacific coast of North America. There, the starfish *Pisaster ochraceus* preys on sessile filter-feeding barnacles and mussels, and also on browsing limpets and chitons and a small carnivorous whelk. Ignoring the rarer species, the community is completed by a sponge and four macroscopic algae. Paine removed all the starfish from a typical piece of shoreline about 8 m long and 2 m deep, and continued to exclude them for several years. At irregular intervals, the density of invertebrates and the cover of benthic algae were then assessed in the experimental area and in an adjacent control site. The control site remained unchanged during the study. Removal of *P. ochraceus*, however, had dramatic consequences. Within a few months, the barnacle *Balanus glandula* settled successfully. Later, barnacles were crowded out by mussels (*Mytilus californianus*), and eventually the site became dominated by these. All but one of the species of algae disappeared, apparently through lack of space, and the browsers tended to move away, partly because space was limited and partly due to lack of suitable food. Overall, the removal of starfish therefore led to a reduction in the number of species from 15 to eight – or, we can say, the presence of the starfish, normally, leads to an increase in the number of species from eight to 15. The main influence of the starfish *Pisaster* appears to be to make space available for competitively subordinate species. It cuts a swathe free of barnacles and, most importantly, free of the dominant mussels that would otherwise outcompete other invertebrates and algae for space. *Pisaster* is therefore said to be responsible for *predator-* or *exploiter-mediated coexistence*: coexistence of a group of species, amongst which, in the absence of the predator, only the most competitive would prevail.

grazing can increase plant species richness ...

The idea goes back at least to Darwin (1859), who seems to have been the first to notice that the mowing of a lawn could maintain a higher richness of species than occurred in its absence. He wrote that:

> If turf which has long been mown, and the case would be the same with turf closely browsed by quadrupeds, be let to grow, the most vigorous plants gradually kill the less vigorous, though fully grown plants; thus out of 20 species growing on a little plot of mown turf (3 feet by 4 feet) nine species perished from the other species being allowed to grow up freely.

Grazing animals are usually choosier than lawn-mowers. To take a simple example: plants that occur in the neighbourhood of rabbit burrows are ones that are unacceptable as food to the rabbits, including the poisonous deadly nightshade, *Atropa belladonna*, and the stinging nettle, *Urtica dioica*. Nevertheless, many grazers seem to have a similar general effect to lawn-mowers. In one experiment, grazing by oxen (*Bos taurus*) and zebu cows (*Bos taurus indicus*) in natural pasture in the Ethiopian highlands was manipulated to provide a no-grazing control and four grazing intensity treatments (Figure 16.13). Significantly, there were more species at intermediate levels of grazing than where there was no grazing or heavier grazing ($P < 0.05$). In the ungrazed plots, several highly competitive plant species, including the grass *Bothriochloa insculpta*, accounted for 75–90% of ground cover. At intermediate levels of grazing, however, the cattle apparently kept the aggressive, competitively dominant grasses in check and allowed a greater number of plant species to persist – another case of exploiter-mediated coexistence. But at very high intensities of grazing, species numbers were reduced again, as the cattle were forced to turn from preferred to less preferred species, driving some to extinction. Where grazing pressure was particularly intense, grazing-tolerant species such as *Cynodon dactylon* became dominant.

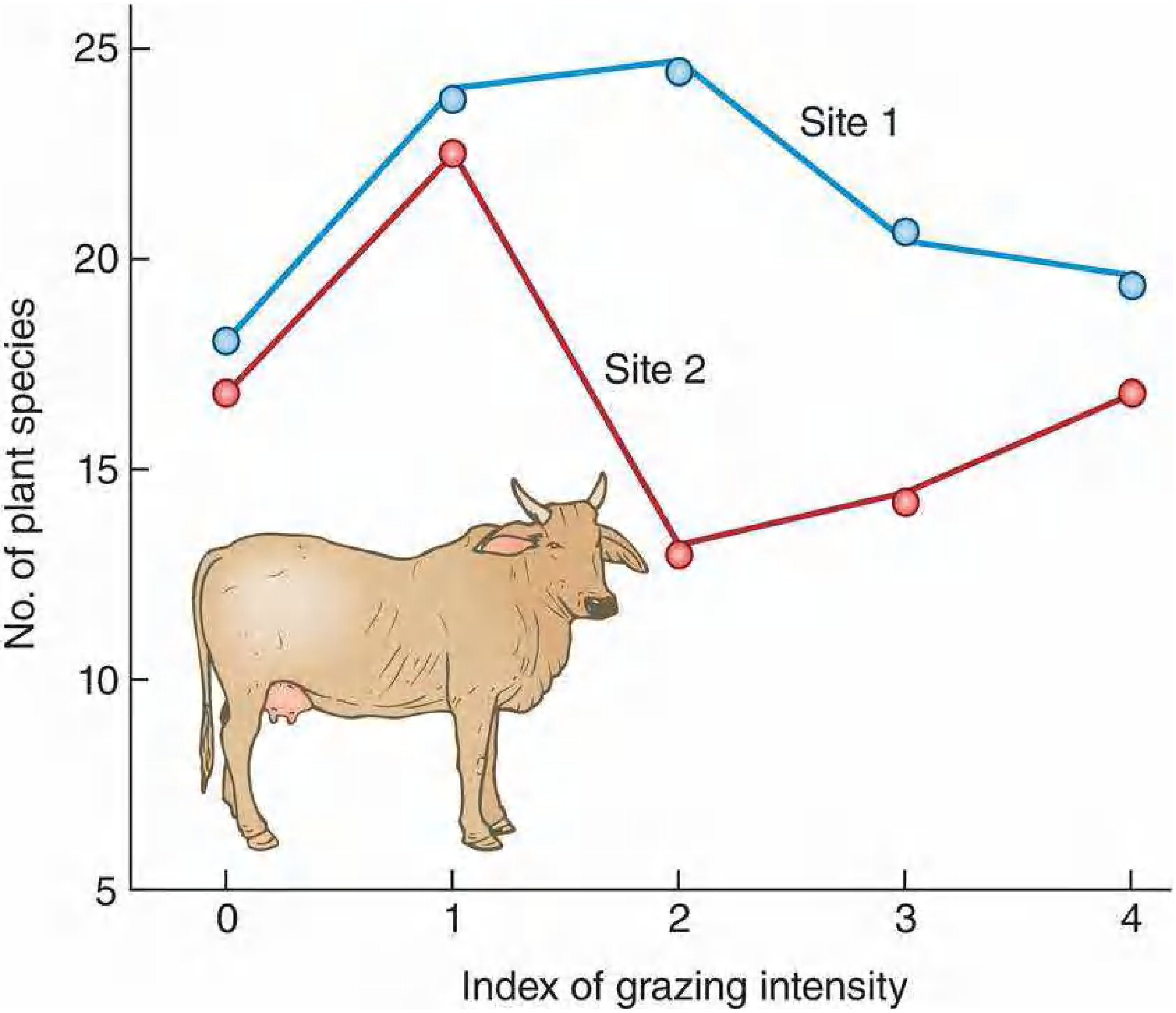


Figure 16.13 Plant species richness is highest at intermediate levels of grazing. Mean species richness of pasture vegetation in plots subjected to different levels of cattle grazing in two sites in the Ethiopian highlands. 0, no grazing; 1, light grazing; 2, moderate grazing; 3, heavy grazing; 4, very heavy grazing (estimated according to cattle stocking rates).

Source: After Mwendera *et al*. (1997).

... but not always

Patterns in which species richness is greatest at intermediate levels of predation also remind us that predation can both decrease species numbers simply by eliminating species and increase them through exploiter-mediated coexistence. It is not surprising, therefore, that exploiter-mediated coexistence is far from universal. Proulx and Mazumder (1998) carried out a meta-analysis of 44 reports of the effects of grazing on plant species richness from lake, stream, marine, grassland and forest ecosystems. The outcome was strongly related to whether the studies had been performed in nutrient-rich or nutrient-poor situations. In all 19 studies from non-enriched or nutrient-poor ecosystems, grazing significantly reduced species richness (Figure 16.14a–c). By contrast, in 14 of the 25 comparisons from enriched or nutrient-rich ecosystems, grazing significantly increased species richness (indicating grazer-mediated coexistence) (Figure 16.14d–g), and nine of the remaining 11 nutrient-rich studies showed no difference at all with grazing regime. The lack of grazer-mediated coexistence in unproductive situations may reflect the poor growth potential of the less competitive species that, in nutrient-rich circumstances, would be released from competitive domination as a result of grazing.

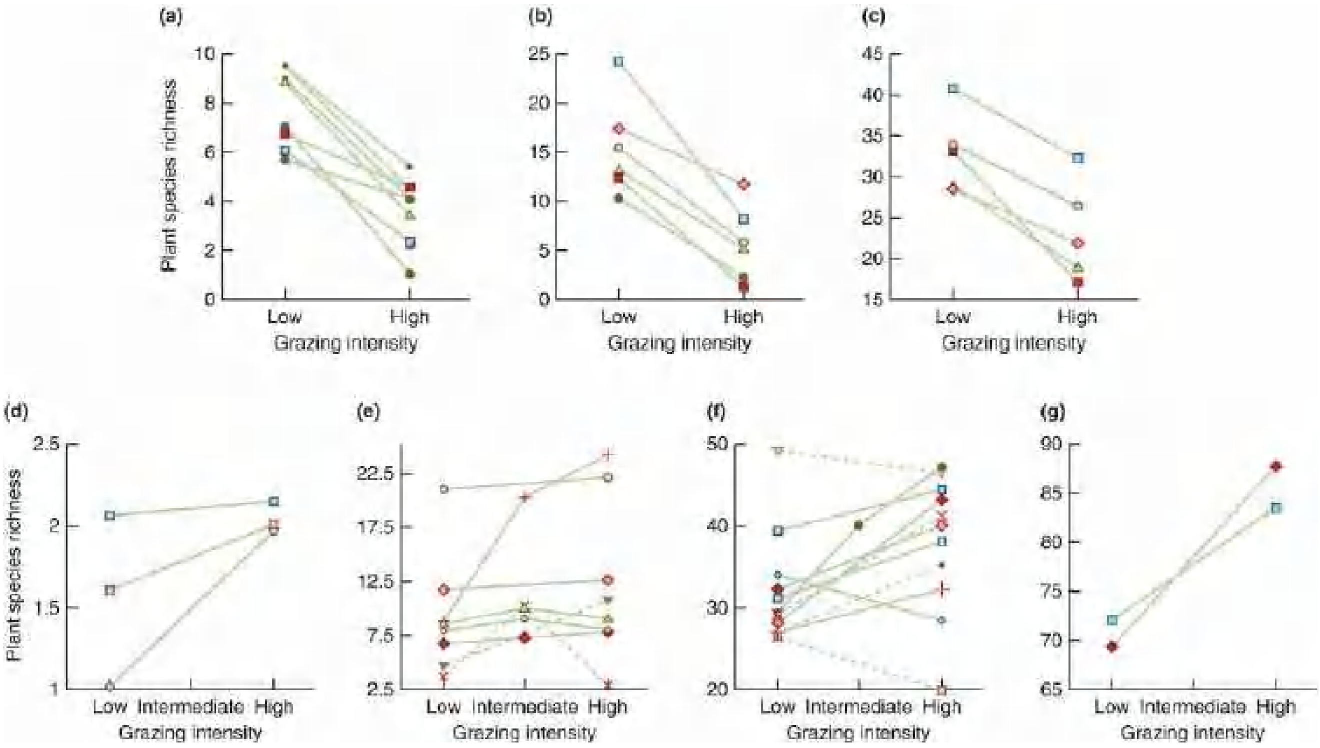


Figure 16.14 Grazing reduces species richness in nutrient-poor ecosystems but often increases richness in nutrient-rich ecosystems. (a–c) Species richness under contrasting grazing pressure (low or high) in non-enriched or nutrient-poor ecosystems. The different lines show the results of different aquatic or terrestrial studies and are presented in separate panels simply for clarity. (d–g) Species richness under contrasting grazing pressure (low, intermediate or high) in various enriched or nutrient-rich ecosystems.

Source: After Proulx & Mazumder (1998).

community responses to grazing may depend on productivity and disturbance

Other studies, too, have found a link between exploiter-mediated coexistence and productivity. Osem *et al.* (2002) recorded the response of communities of annual herbaceous plants in Israel to protection from sheep grazing in four neighbouring topographic situations – south-facing slopes, north-facing slopes, hilltops and wadis (dry streams) (Figure 16.15). Annual above-ground primary productivity was measured at the peak season in four fenced subplots per site and was found to be typical of semiarid ecosystems (10–200 g dry matter m^{-2}) except in wadis (up to 700 g dry matter m^{-2}). These can be taken to represent the 'potential' productivity in adjacent grazed subplots. Grazing only increased plant species richness in the most productive site (wadis) (Figure 16.15d). In the other, less productive sites, species richness was unaffected or declined with grazing. This is likely to reflect variation in the intensity of interspecific competition in the different sites, and hence in the ability of species to make the most of the opportunities afforded by grazing to exploit resources no longer being used by other species. In a similar way, exploiter-mediated coexistence may be less effective in more disturbed habitats, because those disturbances themselves undermine a competitive dominance that a predator might overturn. Thus, whereas, as we saw earlier, in sheltered coastal sites, predation appears to be a dominant force shaping community structure (Paine, 1966), in exposed rocky tidal communities where there is direct wave action threatening survival, predators seem to have a negligible influence on community structure (Menge & Sutherland, 1976 ; Menge *et al*., 1986).

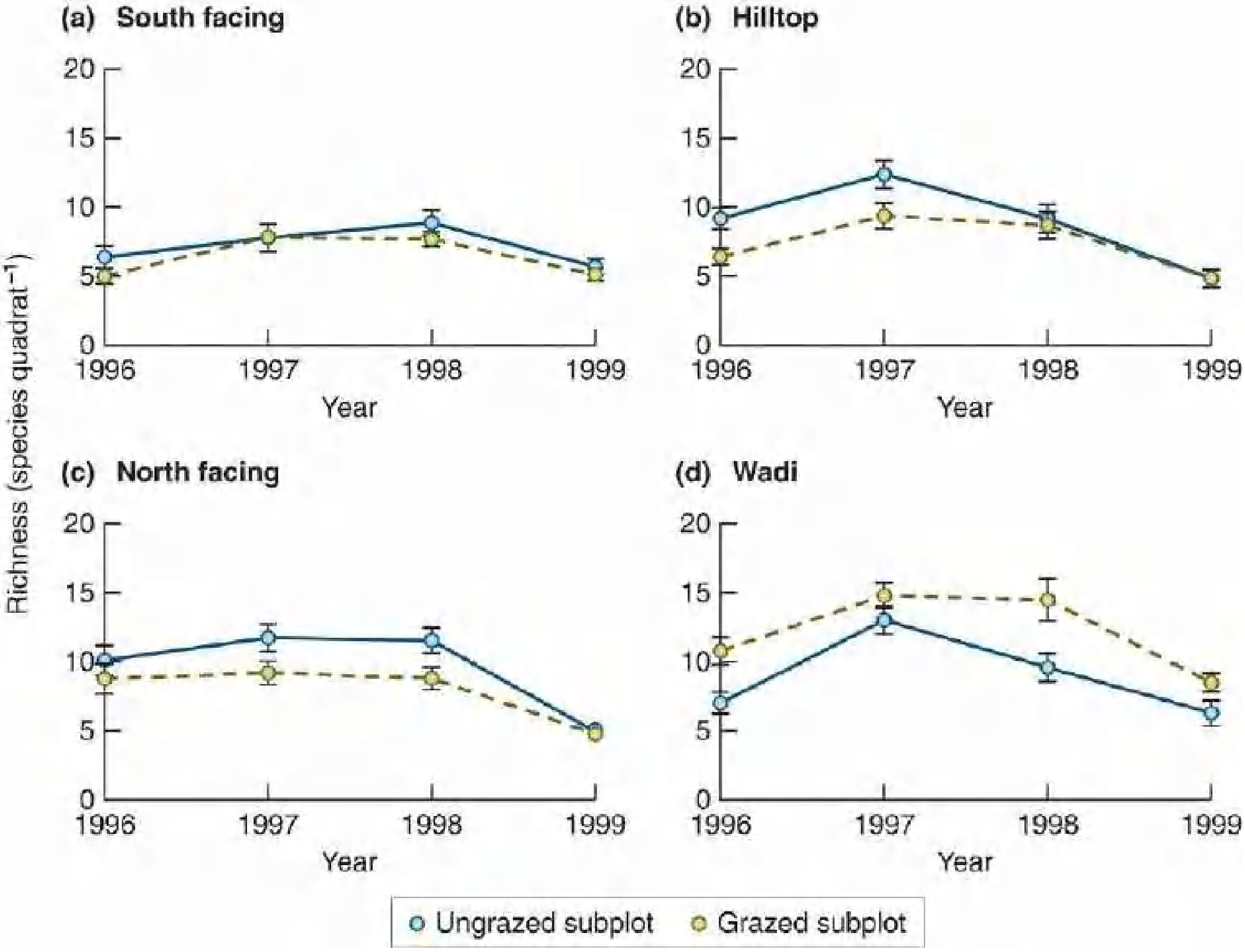


Figure 16.15 Grazing increases species richness in high-productivity sites but not in low-productivity sites in Israel. Species richness (per 20 × 20 cm quadrats) in four topographic sites in Israel in April: (a) south-facing slopes, (b) hilltops, (c) north-facing slopes and (d) wadis.

Source: After Osem *et al.* (2002).

frequency-dependent selection may sometimes enhance diversity: the Janzen–Connell hypothesis …

The response of species richness to predation varies, too, with the species-preferences of the predator (see Chapter 9). Many species take a mixed diet. Others switch sharply from one type of prey to another, taking disproportionately more of the most common types of prey. Such behaviour could lead to the coexistence of a large number of relatively rare species – a frequency-dependent form of exploiter-mediated coexistence. Certainly, there is evidence that predation on the seeds of tropical trees is often more intense where the seeds are more dense (beneath and near the adult that produced them) (Connell, 1971 , 1979); the freshwater zooplanktivorous fish *Rutilus rutilus* switches from large planktonic waterfleas, its preferred prey, to small sediment-dwelling waterfleas when the density of the former falls below about 40 per litre (Townsend *et al.,* 1986); and piscivorous coral reef fish (*Cephalopholis boenak* and *Pseudochromis fuscus*) concentrate on highly abundant cardinal fish (mainly *Apogon fragilis*) when these are present, leaving recruits of many other fish species relatively unmolested (Webster & Almany, 2002).

Indeed, Connell's observation of more intense predation of seeds close to a parent plant has formed the basis of a hypothesis, proposed independently too by Janzen (1970), that such patterns play a key role in increasing the diversity of especially tropical trees, and thus in increasing tropical diversity generally – often called the *Janzen–Connell hypothesis*. By 2013, almost 2000 studies had addressed the hypothesis and more than 1000 of these were experimental tests, but in carrying out a meta-analysis of these studies, Comita *et al.* (2014) applied strict criteria (experimental manipulations were carried out correctly, plants were examined in their natural habits and were accessible by the full range of natural enemies, etc.) to arrive at a dataset containing just 63 articles describing 154 separate tests. The results are shown in Figure 16.16.

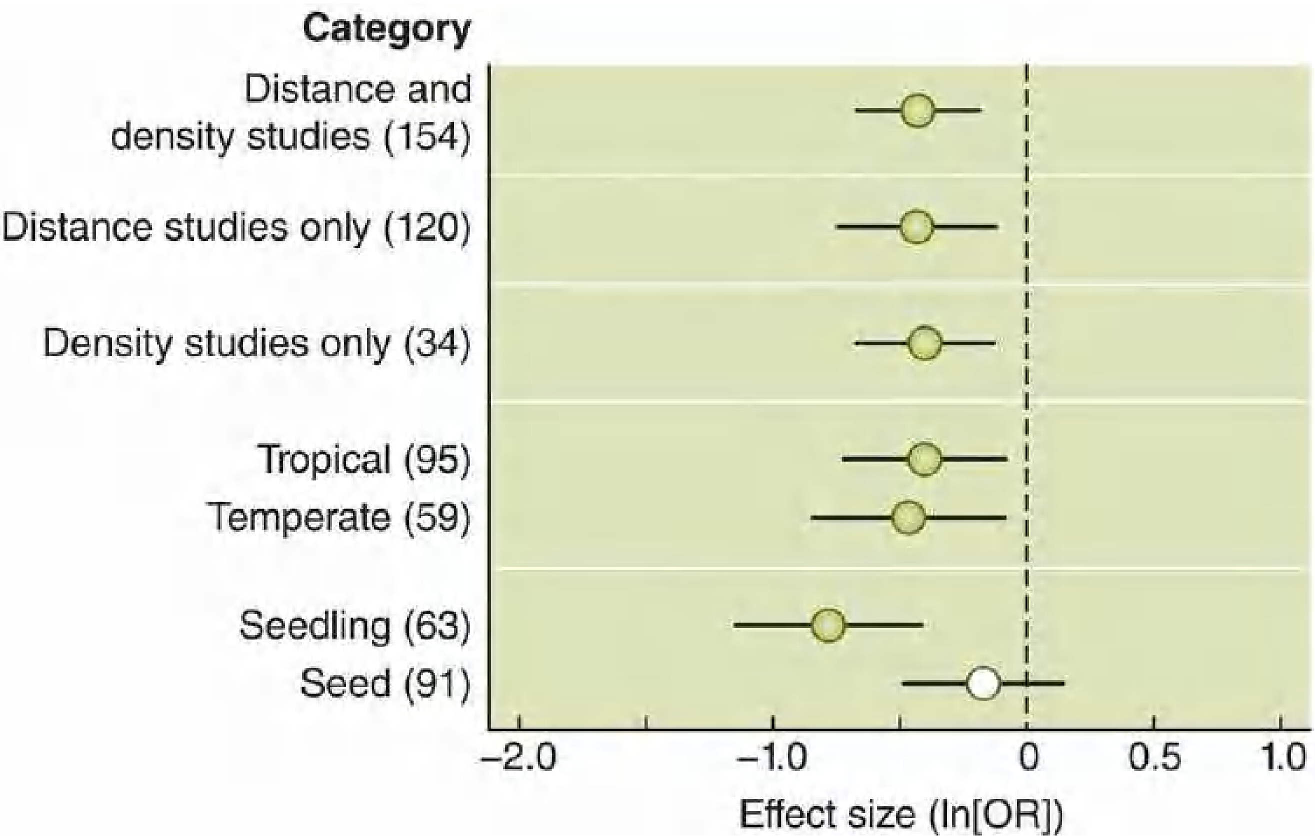


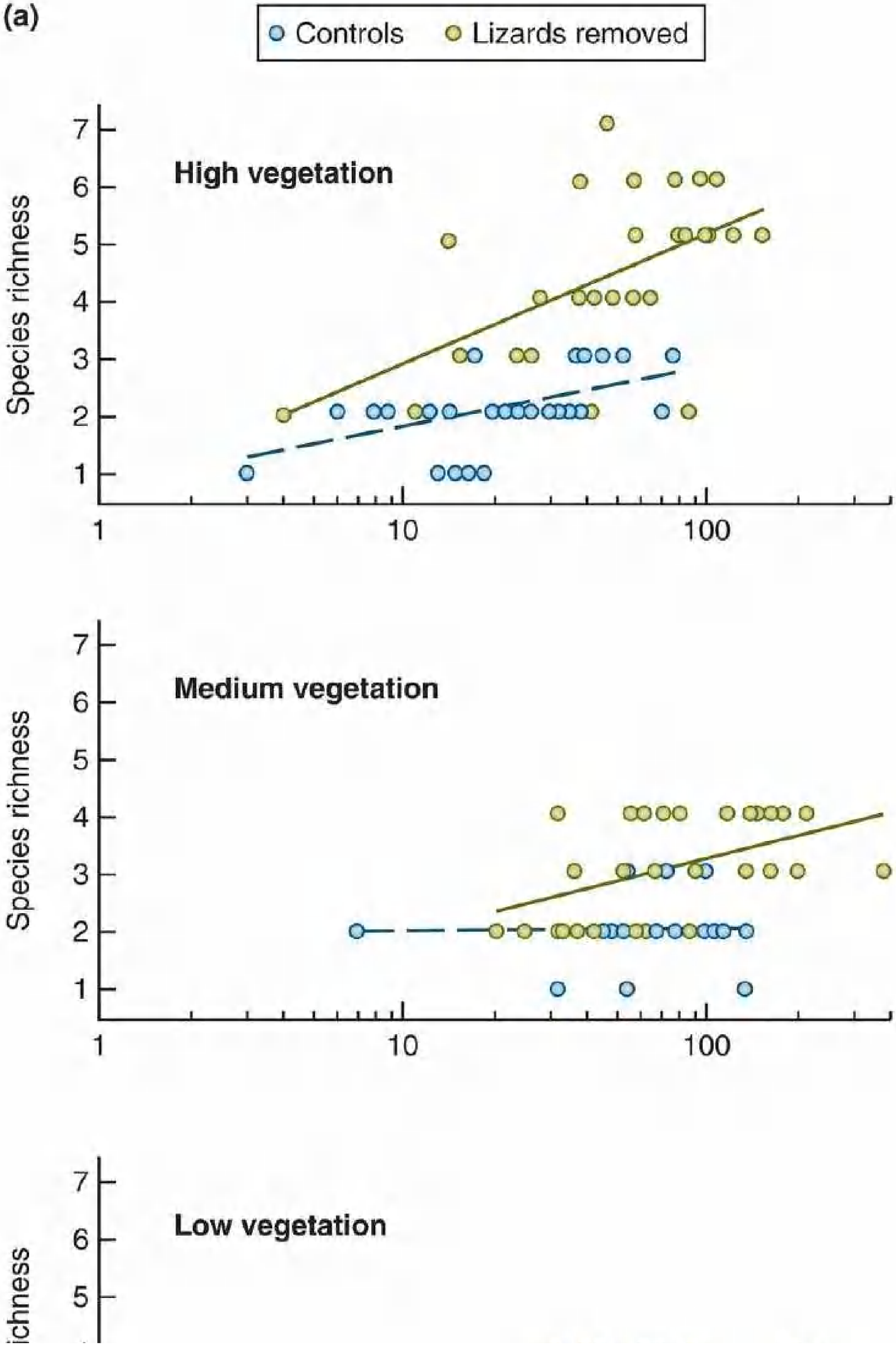
Figure 16.16 A meta-analysis of studies testing the Janzen–Connell hypothesis. The overall effect sizes (log odds-ratios) from 154 studies of the effects of density and proximity to the parent plant on seed and seedling survival. The negative effect sizes therefore indicate that the survival of seeds or seedlings is lower at higher densities or closer to the parent plant. Results are shown for all studies combined and split into categories, as indicated. Bars are 95% CIs. Green dots are significantly different from zero. The open dot is non-significant.

Source: After Comita *et al*. (2014).

The hypothesis itself can be framed as either proposing that predation is most intense closer to the parent plant (distance dependence), or at higher prey densities, which tend to be found closer to the parent plant (density dependence). In natural settings, the two will often be indistinguishable, but in experimental manipulations it is possible to focus on one rather than the other. In fact, though, support for the hypothesis was strong whether all studies were combined or they were split into those testing distance or density dependence (Figure 16.16). What is more, support was strong, too, at both tropical and temperate latitudes, despite the hypothesis having been proposed very much with tropical forests in mind. When the studies were split into those on seeds and those on seedlings, significant support for the hypothesis was found only for those on seedlings. However, looking more closely at the interaction between this and the density–distance distinction, distance had a significant effect on seedling survival but not that of seeds, whereas density had a significant effect on seed survival but not that of seedlings. Of course, these patterns cannot be taken as proof of the proposed mechanism – that the predators and other natural enemies are density- or distance-dependent in their attacks. The strong effect of distance on seedling survival, for example, might reflect an effect of competition from the parent plant itself rather than natural enemies. Nonetheless, of the 59% of studies that explicitly tested for a density- or distance-dependent natural enemy effect, 85% found one.

… but no frequency-dependent predation among communities of insects or spiders

However, such density- and frequency-dependent selection is by no means a general rule. For one thing, some species are so highly specialised that switching is not an option – giant pandas are specialists on bamboo shoots and specialisation in diet is equally extreme amongst many phytophagous insects. And some predators focus on rare not common species. Spiller and Schoener (1998) censused spider populations in the Bahamas at 2-month intervals for four and a half years in enclosures (three replicates) containing or lacking lizards. Species richness was dramatically reduced by the lizards (mainly *Anolis sagrei*) at high and medium levels in the vegetation (at low levels, trends were insignificant; Figure 16.17a). This occurred because the lizards preyed preferentially on rare species of spiders (Figure 16.17b), resulting in increased dominance of the already abundant *Metapeira datona*, a species whose relative invulnerability to predation is probably due to its small size and habit of living in a suspended retreat rather than in the middle of the web.


(a)
Controls
Lizards removed
High vegetation
Species richness
7
6
5
4
3
2
1
1
10
100
Medium vegetation
Species richness
7
6
5
4
3
2
1
1
10
100
Low vegetation
ichness
7
6
5

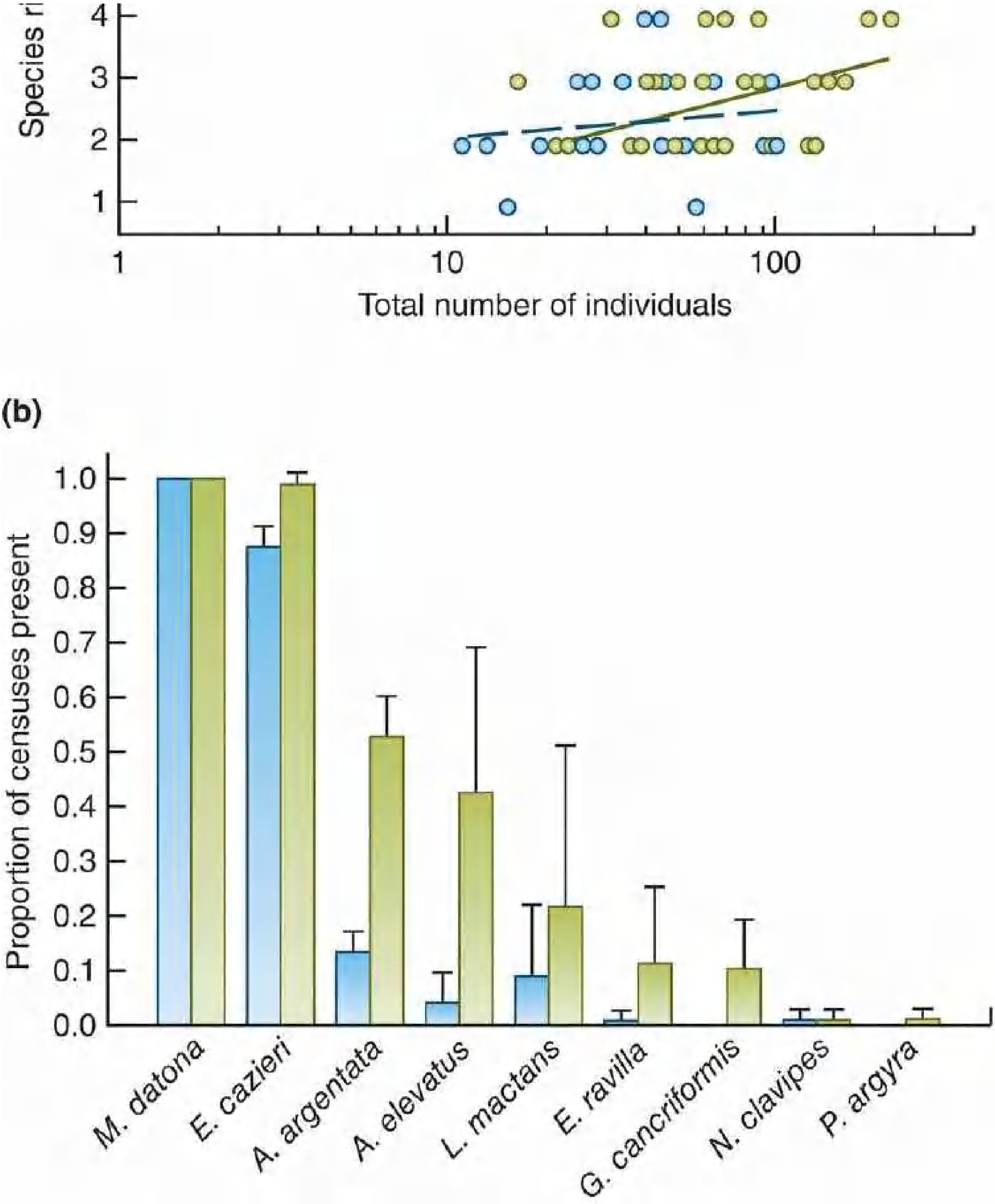


Figure 16.17 Lizards reduce species richness of spiders by preying on rare species. (a) Spider species richness plotted against total number of individuals (all censuses) in the presence and absence of lizards at three heights in the vegetation. For a given number of individuals, enclosures without lizards (green dots) contained a greater number of spider species than enclosures with lizards (blue dots) except low in the vegetation. (b) Mean proportion of censuses in which each web spider was recorded per enclosure in the presence (blue) and absence of lizards (green). Bars are ± SD.

Source: After Spiller & Schoener (1998).

parasite-mediated coexistence in Caribbean lizards but not British squirrels

For parasites, too, there are apparent examples of exploiter-mediated coexistence. For example, the malarial parasite *Plasmodium azurophilum* infects two *Anolis* lizards on the Caribbean island of St Martin. One of the lizards, thought to be the competitive dominant, is widespread throughout the island, while the other is only found in a limited area. However, the superior competitor was much more likely to be infected by the parasite, and the two species only coexist where the parasite is present (Schall, 1992). Once again, though, this is far from a universal pattern. For example, the invading grey squirrel (*Sciurus carolensis*) is displacing the resident red squirrel (*S. vulgaris*) throughout much of its range in Britain. A parapox virus brought in by the invader that has little discernible effect on the grey squirrel itself but a dramatic adverse effect on the health of the native red squirrel is largely responsible (Tompkins *et al*., 2003) (see Section 8.8).

omnivory

Finally, omnivores (see Figure 16.1g) may have particularly far-reaching consequences for a community. For example, omnivorous freshwater crayfish can influence the community composition of plants (which they consume), herbivores and carnivores (which they consume or with which they compete), and even detritivores because their extreme omnivory includes feeding on dead plant and animal material (Usio & Townsend, 2002 , 2004).

appraisal of the effects of predators

Taken overall, therefore, we have seen that the way prey species richness responds to predation depends on the strengths of the links in the shared-predator community module (see Figure 16.11). We have touched on some of the more important variants, but there are undoubtedly other adjustments that could be made to the linkage strengths. Of those we have discussed, exploiter-mediated coexistence, and hence an increase in species richness, is most likely to be the outcome when similar predation pressure is exerted on alternative competing prey, or when the preferred prey is also competitively dominant, especially when competition itself is an important force structuring the community and predation pressure is not too intense. This is most likely to be the case in productive, undisturbed environments. However, if productivity is low, or levels of disturbance are high, competition itself is less likely to be important, and the effects of predation on competitive outcomes is correspondingly slight. Putting these two trends together, it is easy to see why species richness may be greatest at intermediate levels of predation, as we saw for Ethiopian cattle. Conversely, in cases where predators show a preference for rare or uncompetitive prey, or where the alternative prey simply do not compete, predation is most likely to reduce species richness simply by excluding prey species from the community.

16.4 Plurality in the structuring of communities

There have been fashions in ideas about how communities are structured – times when competition was thought to be of over-riding importance, periods when research groups focused on the role of random processes, and so on – but few now would doubt that in most if not all communities, a whole range of forces have a role to play: competition, predation, facilitation and stochastic processes. Indeed, we have seen this in several of the examples discussed above. The competitive interactions amongst anemone fish (see Section 16.2.3) were played out in a landscape of the mutualisms they share with their sea anemone hosts; the many examples of consumer-mediated coexistence are, by their nature, interactions between competitive and predatory processes; and we noted that these were themselves most likely to play an important role when rates of disturbance were low, whereas disturbances may prevent biological interactions from being fully played out. Indeed, we saw in Section 8.7 the several ways in which the outcomes of competitive interactions often had as much to do with responses to disturbance, the colonisation of gaps and so on, as with the niche differentiation that was at the heart of the approaches we discussed in Section 16.2.

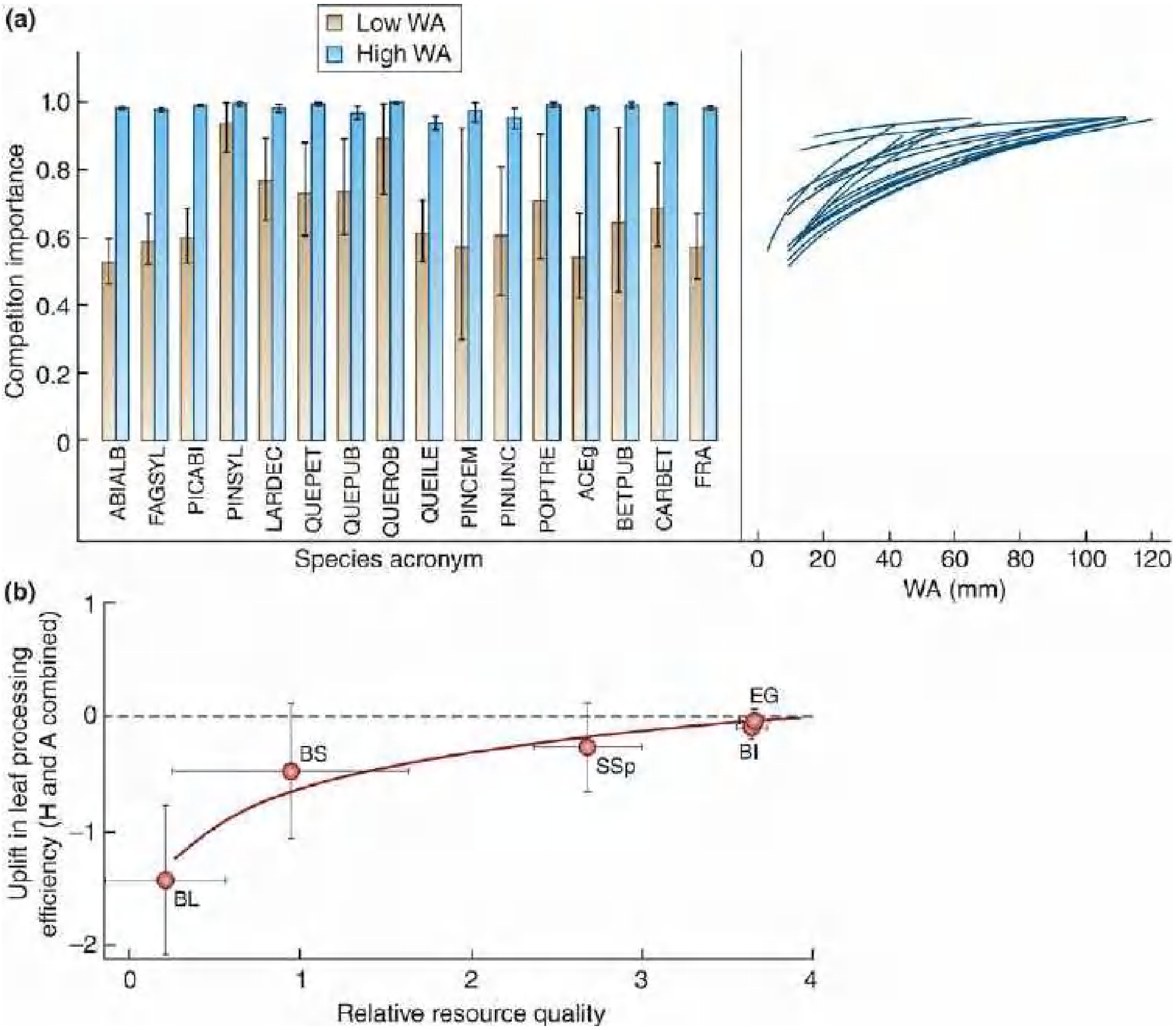
facilitation and the stress-gradient hypothesis

The role of facilitation in structuring communities has been dominated by work on plants, and has been particularly focused on the *stress-gradient hypothesis* (Bertness & Callaway, 1994), which, as we saw in Section 13.2, proposes that the proportion of positive interactions in communities should be greatest in harsher, more stressful environments where species are least likely to survive otherwise. The problem with such proposals, as we discussed in Chapter 2, is that while it may be possible to describe habitats that are stressful for individual species (with conditions or resource levels that are far from the species' optimum) it is much more difficult to define environments as stressful, since some species at least are likely to thrive there, and it is therefore arguably unhelpful even to use the term 'stressful environment'. Nonetheless, given the plurality of factors affecting community structure, it is natural to seek patterns in the relative importance of these factors, especially as, in this case, an increasing proportion of positive, facilitative interactions suggests a decreasing proportion of negative, competitive ones.

Thus, in practice, the stress-gradient hypothesis has generally been taken to predict a higher proportion of positive interactions in unproductive, low-resource environments, especially those lacking water (see Section 13.2). These predictions are themselves related to proposals regarding the importance of competitive interactions in communities, also often focused on plant communities (Craine, 2005), looking at the issue from the perspective of negative not positive interactions. Specifically, Grime (see Grime (2007)) as part of his more general theory of plant strategies (see Section 7.6.3) argued that competitive interactions should be least important, relatively, in stressful (low-resource) environments, which is clearly the other side of the stress-gradient hypothesis coin. By contrast, Tilman (see Tilman (2007)) argued that competition should be *most* important in low-resource environments, where the most successful competitors would be those best able to reduce resource concentrations to the lowest levels (see Section 8.5), though the best competitors in this case would clearly differ depending on which resource was in the most limited supply. This clearly runs counter to the more straightforward predictions of Grime's theories and the stress-gradient hypothesis.

One caveat here is that more recent incarnations of the stress-gradient hypothesis have proposed that the proportion of positive interactions should peak at intermediate (not maximum) levels of stress (Holmgren & Scheffer, 2010). The argument, in essence, is that 'nurse' or 'helper' species, that, for example, provide shade for facilitated species at intermediate levels of water availability, themselves compete with those species when water levels are at their lowest.

There are certainly data supporting the stress-gradient hypothesis, as we saw in Section 13.2. As a further example, a study of trees in the Alps and Jura mountains in France found that the relative importance of competition (though not its intensity) was indeed lower under conditions of water shortage (Figure 16.18a), though the whole approach of separating importance and intensity in determining the role of a process in structuring communities has been questioned (Rees *et al.*, 2012). On the other hand, while some meta-analyses have supported the hypothesis (Section 13.2), others have failed to find support for it (Maestre *et al.*, 2005) but have themselves been criticised on methodological grounds (Lortie & Callaway, 2006). A similarly uncertain picture emerges from the relatively few studies with animals. A test using detritivores in a stream in Ecuador found support in the form of a shift from negative to, on balance, neutral interactions as resource levels declined (Figure 16.18b), but a review of similar studies (nine in total) found equal numbers supporting, refuting and equivocating over the hypothesis (Fugère *et al.*, 2012). Similarly, a meta-analysis of interactions among vertebrate herbivores in alpine environments (15 comparisons from nine studies) found a range of responses (mostly non-significant) but no net effect of harshness on positivity overall (Barrio *et al.*, 2013).



[Figure 16.18](#) **Support for the stress-gradient hypothesis.** (a) Left: variation in the importance of competition between high and low water availabilities, WA (2.5% and 97.5% quantiles, respectively, of the whole range experienced), for 16 species of plant, as indicated by their acronyms (see original publication for full names). Competition importance is the effect of neighbouring trees on the growth of a focal tree as a proportion of all effects on tree growth. WA is the average monthly soil water content (in mm) combining precipitation and evaporation. Bars are 95% credible intervals; non-overlap indicates a significant difference. Right: importance over the whole range of water availabilities experienced by each species. (b) Two detritivores in an Ecuadorian stream (*Hyallela* sp. (H) and *Anomalocosmoecus* sp. (A)) were presented, together, with leaf litter from five different species, as indicated, that differed in quality (determined separately). BI, *Buddleja incana*; BL, *Baccharis latifolia*; BS, *Brugmansia sanguinea*; EG, *Eucalyptus globulus*; SSp, *Solanum* sp. The vertical scale measures the difference between the combined leaf processing efficiency (LPE) of the two species and what would be expected from their respective LPEs, when alone. Bars are 95% CIs. The line describes a logarithmic regression ($r^2 = 0.96$; $P = 0.003$).

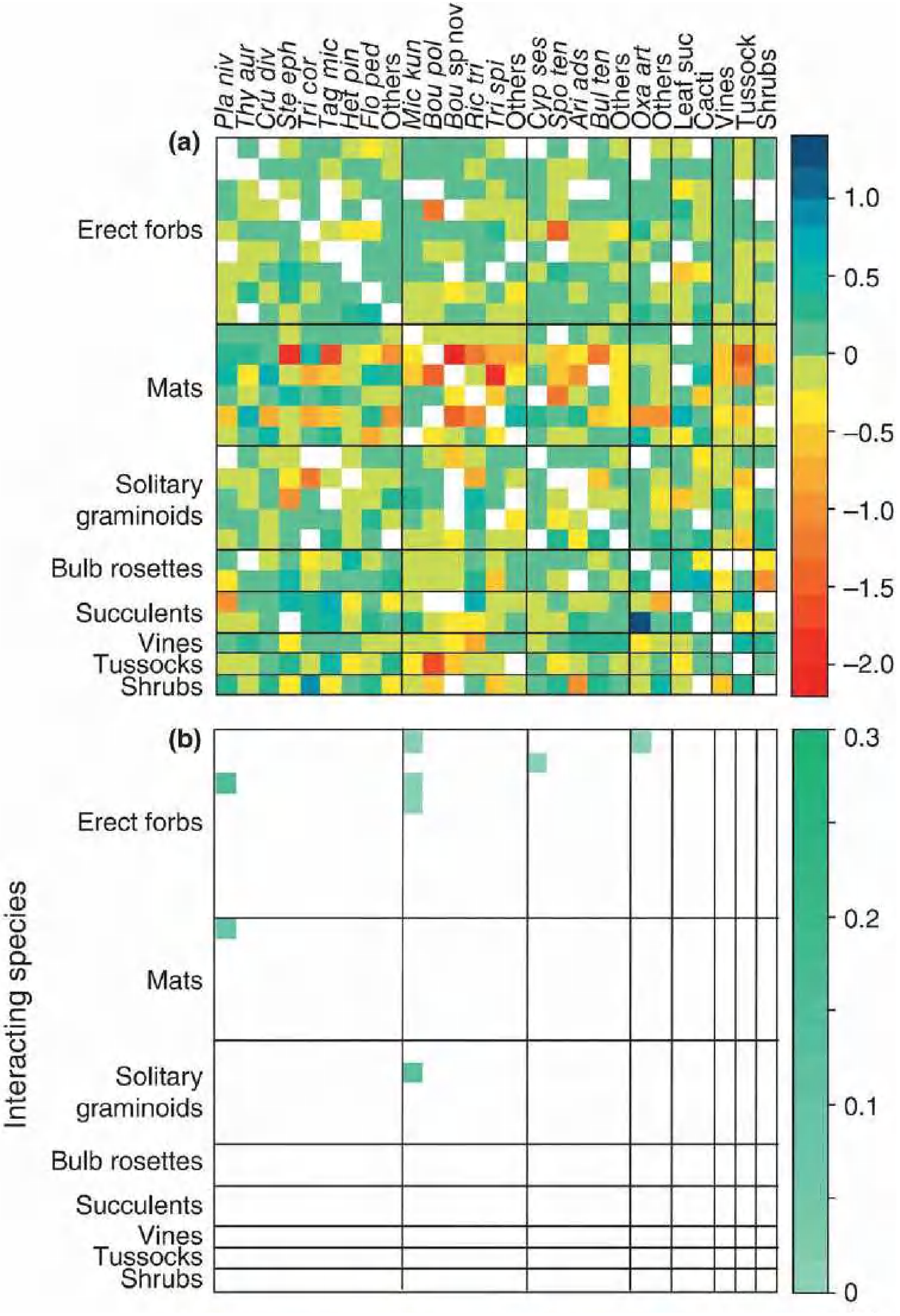
Source: (a) After Kunstler *et al.* (2011). (b) After Fugère *et al.* (2012).

Hence, the available data provide some support both to those advocating and those doubting the stress-gradient hypothesis. With all this, the hypothesis has remained a focus for studies seeking patterns in the relative importance of different processes in structuring communities, but difficulties remain over both its theoretical foundation and the best ways to test it. Further refinement and clarification of the hypothesis is an important challenge (Soliveres *et al.*, 2015).

interspecific interactions and stochastic processes combine to mould structure

An alternative approach to disentangling the roles of different forces in structuring communities is the more intensive study of individual systems. In one example, aimed explicitly at teasing apart the roles of competition, facilitation and stochastic processes, detailed data were analysed from a species-rich plant community in southern Mexico – more than 100 species in the study area, mostly very small (<5 cm in height and diameter) (Martorell & Freckleton, 2014). Hence, this was a study aimed at a community at just one (the lowest) trophic level, with limited opportunity to detect the influence of the predatory interactions discussed in Section 16.3. At each of four 1 ha sites, seven annual transitions (2001–08) were followed by recording the abundance of species in 20 randomly chosen 0.1 × 0.1 m squares in each of 16 1 × 1 m quadrats at each site, generating more than 1 million abundance records overall. These data were then used to parameterise a population model in which the abundance of each species in the community at time $t + 1$ was projected from the abundances of all species at time t in the same square, taking account of both positive and negative interactions between species, dispersal of seeds into the squares, and stochasticity at both annual and intra-annual scales. This was done separately for squares in which a focal species was present at time t – thus reflecting the role of the various factors throughout the life cycle – and squares in which the focal species was absent – thus reflecting the outcome of interactions occurring during colonisation, germination and early growth.

The results are shown in Figure 16.19. For the early, colonisation phase, there was a high proportion of positive interactions between species (Figure 16.19a), and around 15% of these were significantly so, compared with fewer (and fewer significant) negative interactions – around 6%. By contrast there were very few positive interactions detected between pairs of established plants (Figure 16.19b) and relatively few negative, competitive interactions, most of which were intra- rather than interspecific (Figure 16.19c). In the first place, this alerts us to the patchiness of most communities and the importance of distinguishing between the forces driving the colonisation of patches and those moulding species abundances in more established patches. For these established patches, the rarity of interspecific interactions and the importance of intraspecific competition suggest a community in which species coexist only when their niches are differentiated – that is, the results indicate that interspecific competition is important in determining community structure but is not a contemporaneous active force. Although species interactions, both positive and negative, were far more important during the colonisation phase, it is noteworthy that while the population model was good at recreating the dynamics of the community with all factors included, it was still effective at doing so if the interspecific interactions were removed. However, if the stochasticity was also removed, the abundances of virtually all species were consistently underestimated, with a mean reduction of 47%. This supports the importance of the stochastic colonisation of empty patches as another key process in the structuring of the community.


Pla niv
Thy aur
Cru div
Ste eph
Tri cor
Tag mic
Het pin
Flo ped
Others
Mic kun
Bou pol
Bou sp nov
Ric tri
Tri spi
Others
Cyp ses
Spo ten
Ari ads
Bul ten
Others
Oxa art
Others
Leaf suc
Cacti
Vines
Tussock
Shrubs
(a)
Erect forbs
Mats
Solitary graminoids
Bulb rosettes
Succulents
Vines
Tussocks
Shrubs
1.0
0.5
0
−0.5
−1.0
−1.5
−2.0
(b)
Interacting species
Erect forbs
Mats
Solitary graminoids
Bulb rosettes
Succulents
Vines
Tussocks
Shrubs
0.3
0.2
0.1
0

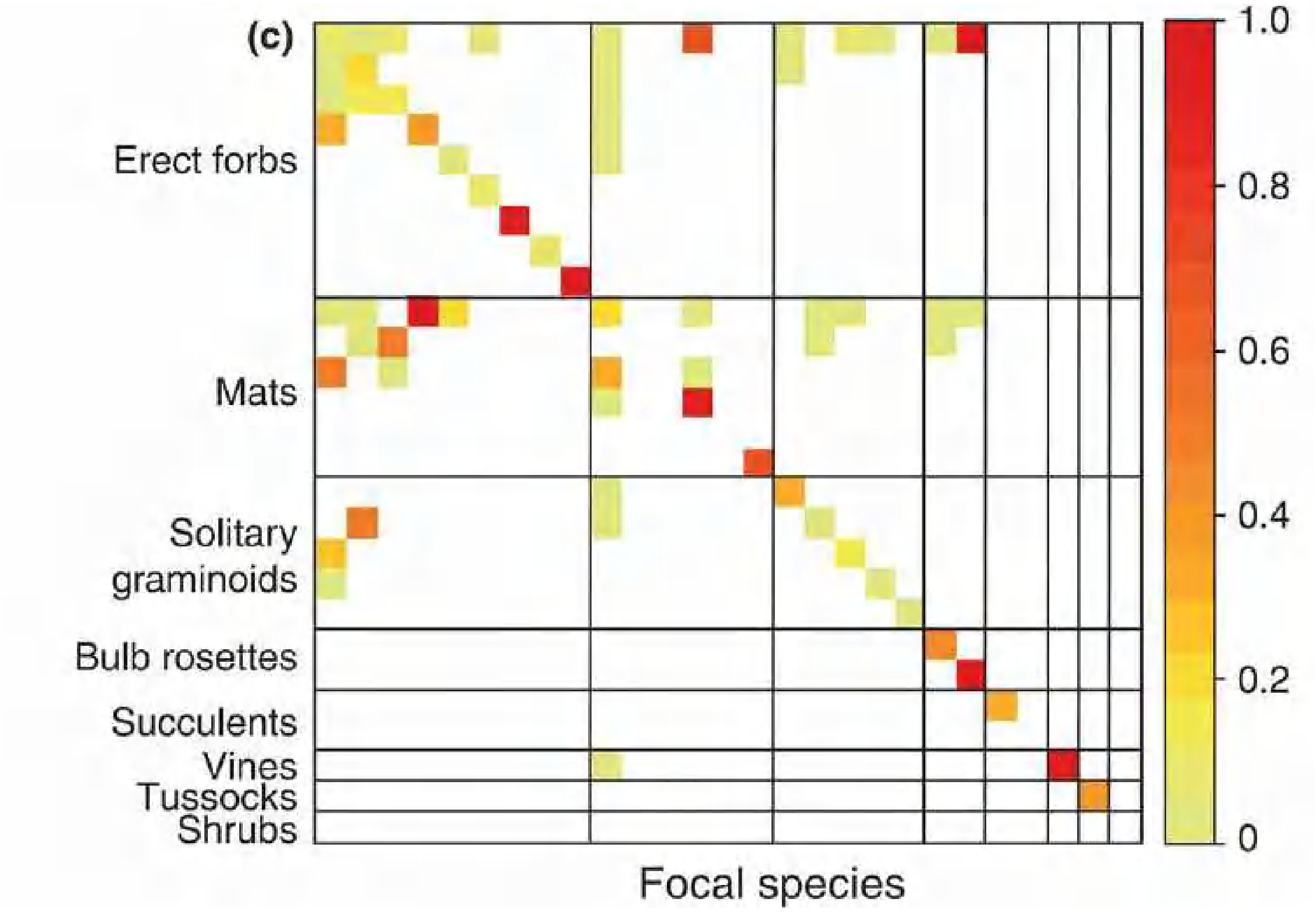


Figure 16.19 The effects of species interactions in structuring a plant community in Mexico suggest facilitation is important during colonisation but not among established plants, where there is also relatively little competition that is not intraspecific. Each panel is a matrix of positive and/or negative per capita interspecific interaction coefficients, with the commoner species named (see original publication for a full description). The black cross lines divide the species into functional types (erect forbs, mats, solitary graminoids, etc.). Rarer species are not named but are grouped together, either as 'others' at the end of the species list within a functional type, or simply into the functional types at the end of the list (vines, tussocks and shrubs). The colour coding is explained in the bar to the right in each case. Open squares signify coefficients that either were negligibly small or had insufficient data for estimation. (a) Interactions affecting the colonisation probability of the focal species (mainly positive – green and blue). (b) Facilitation coefficients affecting the focal species in squares they already occupy (very few are present). (c) Competition coefficients affecting the focal species in squares they already occupy (not many are present other than along the diagonal – intraspecific competition).

Source: After Martorell & Freckleton (2014).

The authors of the study argue for the probable generality of their findings – that strong interspecific interactions and the stochastic colonisation of empty patches are most likely to be important in the early phases of (local) community establishment, but strong self-limitation of species, each in their own niche, may dominate in more established phases. The balance between these phases will itself depend on the size of patches to be colonised and the frequency with which they are created. Further studies are needed to determine if this is so. What is more certain to be general is that in the structuring of any community, especially once we extend from one to multiple trophic levels, will reflect the whole range of interspecific interactions and stochastic processes.

Chapter 17

Food Webs

17.1 Food chains

In the previous chapter we considered how direct interactions between species can shape communities. Our main focus was on interactions between species occupying the same trophic level (interspecific competition and to a lesser extent facilitation) or between members of adjacent trophic levels. In fact, of course, the influence of a species often ramifies further than this. The effects of a carnivore on its herbivorous prey may also be felt by any plant population upon which the herbivore feeds, by other predators and parasites of the herbivore, by other consumers of the plant, by competitors of the herbivore and of the plant, and by the myriad of species linked even more remotely in the food web. In this chapter, we broaden our scope by looking at food chains (see Figure 16.1f) and food webs.

Food chains may be defined as sequences of species running from a basal, autotrophic species to a species that feeds on it, to another species that feeds on the second, and so on up to a top predator (fed on by no other species). This does not imply a belief that communities are organised as linear chains (as opposed to more diffuse webs). Rather, individual chains are identified purely as a means of trying to understand aspects of the structure of food webs – for example the number of links between the base and the apex. We begin by focusing on the interactions themselves – often indirect interactions, linking one species to another, two or more levels below or above it in the food chain.

The removal of a species (experimentally, managerially or naturally) can be a powerful tool in unravelling the workings of a food web. If a predator species is removed, we expect an increase in the density of its prey. If a competitor species is removed, we expect an increase in the success of species with which it competes. There are plenty of examples of such expected results. Sometimes, however, removing a species may lead to a decrease in competitor abundance, or the removal of a predator may lead to a decrease in prey abundance. Such unexpected effects arise when direct effects are less important than the effects that occur through indirect pathways. Thus, the removal of a species might increase the density of one competitor, which in turn causes another competitor to decline. Or the removal of a predator might increase the abundance of a prey species that is competitively superior to another, leading to a decrease in the density of the latter. In a survey of more than 100 experimental studies of predation, more than 90% demonstrated statistically significant results, and of these about one in three showed such indirect effects (Sih *et al.*, 1985).

17.1.1 Trophic cascades

The indirect effect within a food web that has received most attention is the so-called *trophic cascade* (Paine, 1980 ; Ripple *et al*., 2016; see Figure 16.1f). It occurs when a predator reduces the abundance of its prey, and this cascades down to the trophic level below, such that the prey's own resources (typically plants or phytoplankton) increase in abundance. One example of a trophic cascade, but also of the complexity of indirect effects, is provided by a two-year experiment in which bird predation pressure was manipulated in an intertidal community on the north-west coast of the USA, in order to determine the effects of the birds on three limpet species (prey) and their algal food (Wootton, 1992). Glaucous-winged gulls (*Larus glaucescens*) and oystercatchers (*Haematopus bachmani*) were excluded by means of wire cages from large areas (each 10 m^2) in which limpets were common. Overall, limpet biomass was much lower in the presence of birds, and the effects of bird predation cascaded down to the plant trophic level, because grazing pressure on the fleshy algae was reduced. In addition, the birds freed up space for algal colonisation through the removal of barnacles (Figure 17.1).

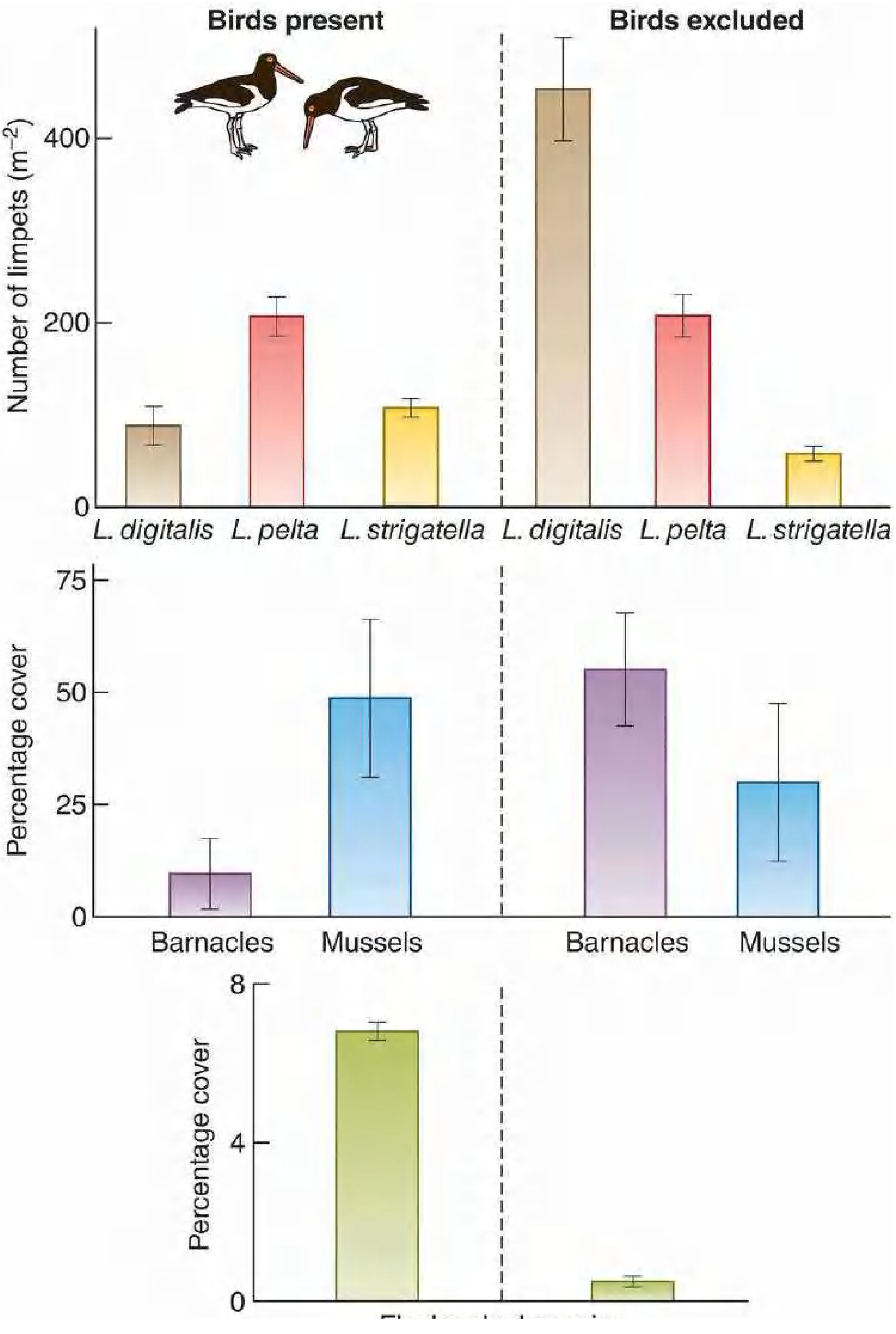

Birds present
Birds excluded
Number of limpets (m^{-2})
400
200
0
L. digitalis
L. pelta
L. strigatella
L. digitalis
L. pelta
L. strigatella
Percentage cover
75
50
25
0
Barnacles
Mussels
Barnacles
Mussels
Percentage cover
8
4
0

Figure 17.1 **A trophic cascade in an intertidal community.** When birds are excluded from the intertidal community, barnacles increase in abundance at the expense of mussels, and three limpet species show marked changes in density, reflecting changes in the availability of cryptic habitat and competitive interactions as well as the easing of direct predation. Algal cover is much reduced in the absence of effects of birds on intertidal animals (means ± SE are shown).

Source: After Wootton (1992).

However, it also became evident that while birds reduced the abundance of one of the limpet species, *Lottia digitalis*, as might have been expected, they increased the abundance of a second limpet species (*L. strigatella*) and had no effect on the third, *L. pelta*. The reasons are complex and go well beyond the direct effects of consumption of limpets. *L. digitalis*, a light-coloured limpet, tends to occur on light-coloured goose barnacles (*Pollicipes polymerus*), whilst dark *L. pelta* occurs primarily on dark Californian mussels (*Mytilus californianus*). Both limpets show strong habitat selection for these cryptic locations. Predation by gulls reduced the area covered by goose barnacles (to the detriment of *L. digitalis*), leading through competitive release to an increase in the area covered by mussels (benefiting *L. pelta*). The third species, *L. strigatella*, is competitively inferior to the others and increased in density because of competitive release.

four trophic levels

In a four-trophic-level system, if it is subject to a trophic cascade, we might expect that as the abundance of a top carnivore increases, the abundances of primary carnivores in the trophic level below decrease, those of the herbivores in the level below that therefore increase, and plant abundance at the lowest level decreases. This is what was found in a study in the tropical lowland forests of Costa Rica. *Tarsobaenus* beetles preyed on *Pheidole* ants that preyed on a variety of herbivores that attacked ant-plants, *Piper cenocladum* (Figure 17.2a). These showed precisely the alternation of abundances expected in a four-level cascade: relatively high abundances of plants and ants associated with low levels of herbivory and beetle abundance at three sites, but low abundances of plants and ants associated with high levels of herbivory and beetle abundance at a fourth site (Figure 17.2b).

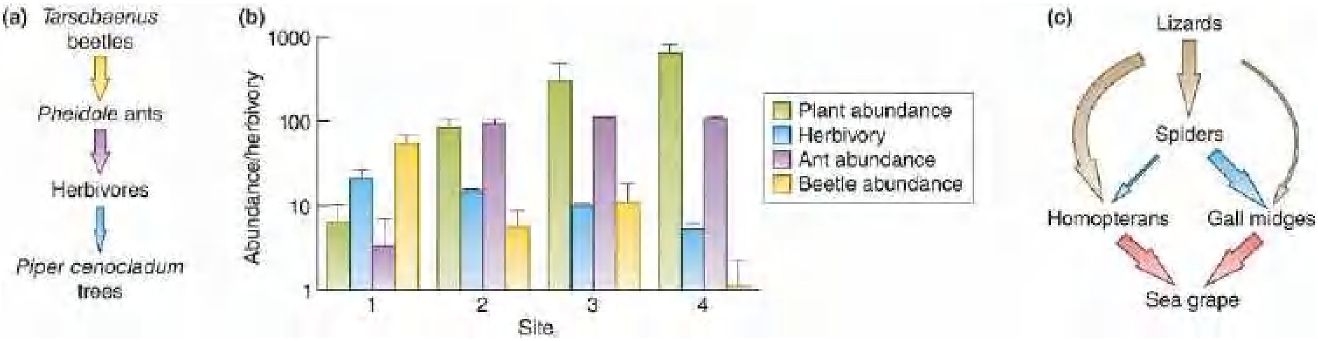


Figure 17.2 Trophic cascades in four-level food chains – which may sometimes function as having only three levels. (a) Schematic representation of a four-level food chain in Costa Rica with full alternation between levels. Arrows denote contributions to mortality in the resource; arrow breadth denotes their relative importance. (b) The relative abundance of ant-plants, strength of herbivory, abundance of ants and abundance of beetles at four sites, with, for example, high beetle abundance at site 1 compared to other sites. The units of measurement are various and given in the original references; means ± SE are shown. (c) Schematic representation of a four-level food chain in the Bahamas, functioning as four levels through the gall midges but only three through the homopterans. Arrows again denote contributions to mortality in the resource; arrow breadth denotes their relative importance.

Source: (b) After Letourneau & Dyer (1998a , 1998b) and Pace *et al.* (1999). (c) After Spiller & Schoener (1990).

four levels can act like three

However, results were more complex in another four-trophic-level community, in the Bahamas (Figure 17.2c), consisting of sea grape shrubs fed on by herbivorous arthropods (homopterans and midges), which were fed on by web spiders (primary carnivores), but in which there were also lizards (top carnivores) that fed on both the spiders and the herbivores – a community we met previously in Figure 16.17. When lizards were removed from experimental plots and results compared with controls, the subtlety arose because the lizards ate large numbers of plant-sucking homopterans but very few gall-forming midges (mostly eaten by the spiders). Therefore, tracing the chain through the midges, the expected four-level alternation was observed. Removing lizards benefited the plants. However, removing the lizards also directly and significantly benefited the homopterans, and so as far as the damage they caused was concerned, this harmed the plants. Through this route from top to bottom, the four-trophic-level community functioned as if it had only three levels.

In rare cases, a cascade can be observed to extend beyond its four levels down to abiotic resources used by the plants. Figure 17.3 shows the effects of the commercial fishing of cod (*Gadus morhua*) and other fish off the coast of Nova Scotia, Canada. As the cod were steadily overfished from the 1960s through the 1990s, the abundance of their own food (small fish and bottom-dwelling invertebrates) increased, the zooplankton on which these feed decreased, and the phytoplankton on which the zooplankton feed increased; it was even possible to monitor a decrease in the concentration of one of the phytoplankton's essential resources, nitrate.

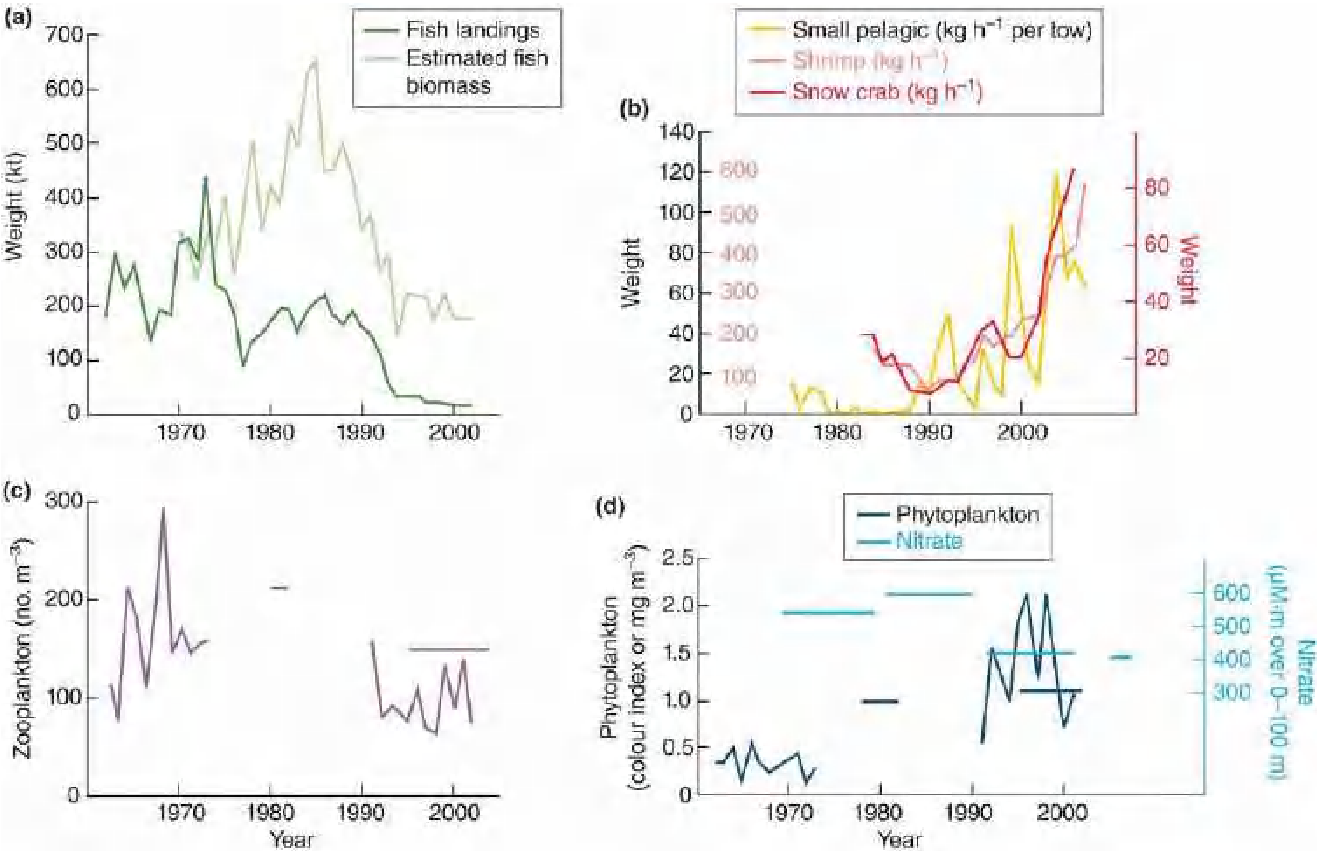


Figure 17.3 **A rare trophic cascade extending beyond four levels to nitrate used by phytoplankton.** The effects of cod fishing (and that of other fish) off Nova Scotia from the 1960s through the 1990s (a) on the cod themselves, (b) on their food, estimated from standard samples from the environment, (c) on the zooplankton (horizontal lines are averages for separate samples from the periods concerned) and (d) on the phytoplankton and on nitrate concentrations.

Source: After Frank *et al.* (2005).

APPLICATION 17.1 Mesopredator release

One example of a trophic cascade not necessarily extending throughout a food chain is what has been called *mesopredator release*, where a predator at the very top of a chain (an ‘apex’ predator) feeds on typically smaller predators (mesopredators), such that the loss or decline of the apex predator allows the abundance of the mesopredator to increase, threatening the abundance of their own prey (Ritchie & Johnson, 2009) – a particular case of the omnivory module (see Figure 16.1g). These prey may themselves be the object of conservation concerns, and there are several good examples of this from Australia. There is a general absence of native apex predators in Australia, these having suffered from extinctions during the Late Quaternary, but a number of apex and mesopredators have subsequently invaded. Notable amongst these is the dingo, *Canis lupus dingo*, a feral dog apex predator, from around 4000 years ago, and red foxes, *Vulpes vulpes*, and cats, *Felis silvestris catus*, introduced much more recently by Europeans, that are preyed upon by dingoes but also themselves prey upon many native small mammals. More recently, persecution by Europeans has led to reductions in the numbers of dingoes. Where dingoes are still present, they tend to keep the mesopredators, the foxes and cats, in check. But in the absence of dingoes, fox and cat abundance may rise to levels that threaten smaller species.

One example is the endangered dusky hopping mouse, *Notomys fuscus*, studied in the Strzelecki Desert in south central Australia, where a dingo barrier fence, 2 m high and more than 5000 km long, has been erected to protect sheep from dingo attack but also allows *N. fuscus* abundance in the presence and virtual absence of dingoes to be compared. While the abundance of the mouse increased from north to south in the study area as whole, numbers were consistently higher in the absence of dingoes (Figure 17.4a). Statistical models seeking to account for variations in mouse abundance in terms of the factors that may influence them showed that they were most abundant in wetter areas, where there was more food for them, but also where the abundance of dingoes was higher and the abundance of red foxes lower (Figure 17.4b), all indicative of dingoes protecting mice by suppressing foxes – or of foxes released from control by dingoes threatening the continued existence of the mice. Dingoes, perhaps because they are relatively recent arrivals in Australia and so are not considered truly ‘native’, are generally regarded as pests and as a species to be controlled. Results such as this, and more generally (Johnson *et al*., 2007), suggest that a reassessment of their status may be required, in which the benefits they bring, as well as the harm they do, are given due consideration.

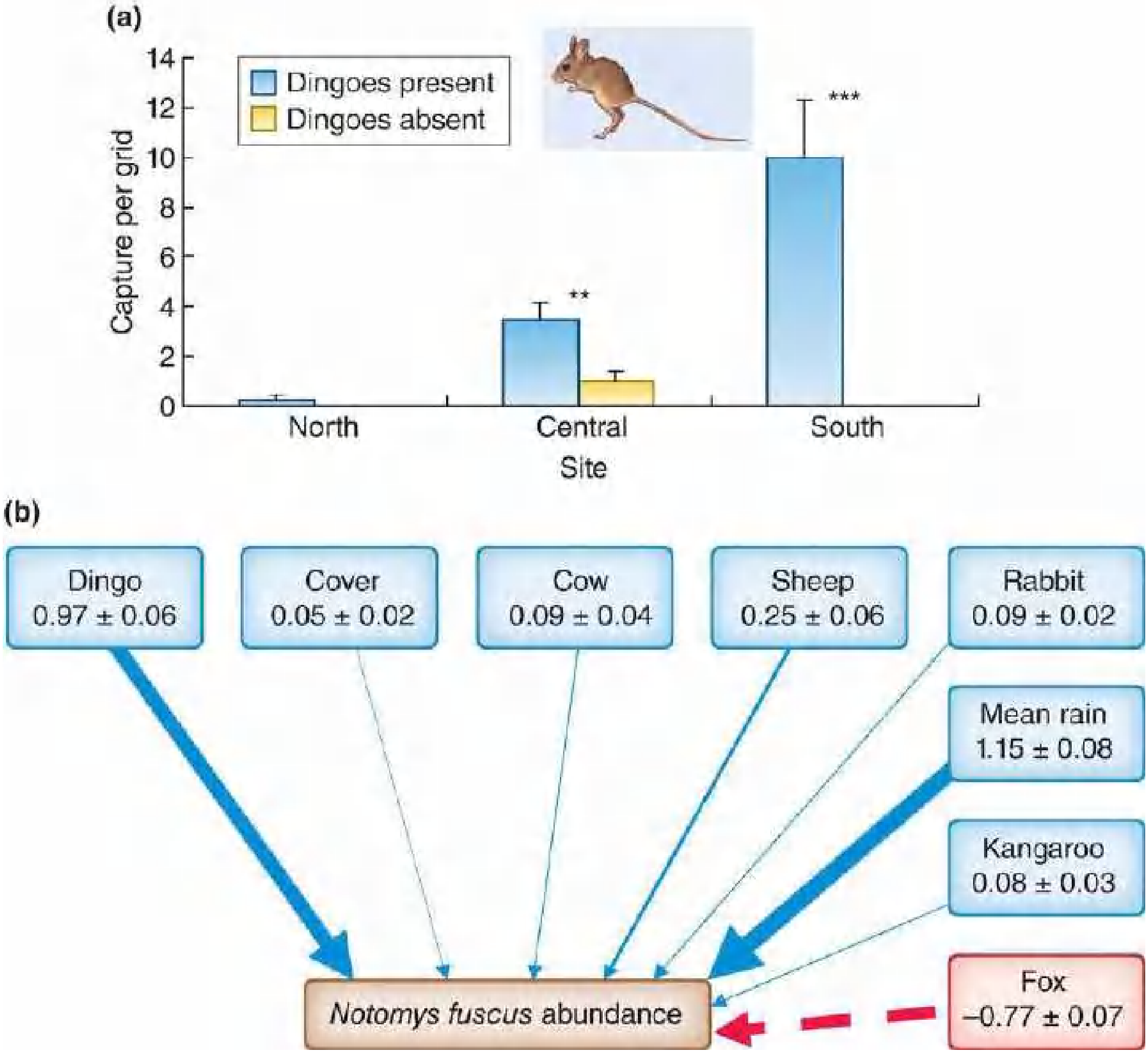


Figure 17.4 Mesopredator release (of foxes and cats by removing dingoes) threatening an endangered marsupial mouse in Australia. (a) Mean numbers of the dusky hopping mouse, *Notomys fuscus*, captured from grids at three sites in the Strzelecki Desert in Australia in the presence and absence of dingoes, *Canis lupus dingo*. Bars are SEs. ** $P < 0.01$, *** $P < 0.001$. (b) Results of a regression model accounting for variations in mouse abundance. Line width reflects the size of the coefficient in each case (with SEs). Solid arrows are positive effects; dashed red arrow indicates negative effects.

Source: After Letnic *et al.* (2009).

17.1.2 Top-down or bottom-up control of food webs?

Trophic cascades are generally viewed ‘from the top’, starting at the highest trophic level. So, in a three-trophic-level community, we think of the predators controlling the abundance of the herbivores – so-called top-down control. Reciprocally, the predators are subject to bottom-up control: abundance determined by their resources. With these three levels, the plants are also subject to bottom-up control, having been released from top-down control by the effects of the predators on the herbivores. Thus, in a trophic cascade, top-down and bottom-up controls alternate as we move from one trophic level to the next, though this is generally referred to as ‘top-down control’ overall. But suppose instead that we start at the other end of the food chain, and assume that the plants are controlled bottom-up by competition for their resources. It is still possible for the herbivores to be limited by competition for plants – their resources – and for the predators to be limited by competition for herbivores. In this scenario, all trophic levels are subject to bottom-up control (also called 'donor control'), because the resource controls the abundance of the consumer but the consumer does not control the abundance of the resource.

The question therefore arises: 'Are food webs – or are particular types of food web – dominated by either top-down or bottom-up control?' We have seen examples of trophic cascades. There are also examples of bottom-up control to set alongside these. In one study in Brazil, for example, the perennial shrub, *Chromolaena pungens*, was subjected to two levels of clipping (around 50% of leaves removed or cut at the base) and allowed to regrow. This is known to give rise to poorly defended, nutrient-rich foliage in the rapidly regrowing plants, which were then compared with control, unclipped plants, in terms of the plants' own traits, and in terms of the herbivores and predators in the two trophic levels above. Partial clipping had relatively little effect, but clipped plants grew faster and had larger leaves and were much less likely to flower (Figure 17.5a). This led to the degree of infestation of flower heads with seed-eaters being higher in the clipped plants, as was the level of leaf consumption (Figure 17.5b, c); and this increased availability of food led in turn to much larger numbers of predators on the clipped plants (Figure 17.5d).

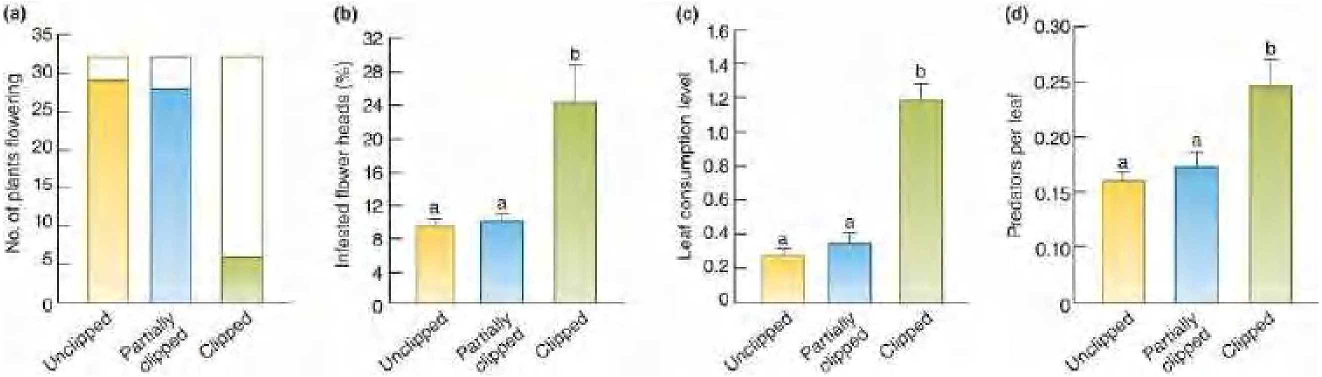


Figure 17.5 Bottom-up control of a food web in Brazil. The responses of plants, herbivores and predators when plants, *Chromolaena pungens*, were either partially or fully clipped or left as controls and allowed to regrow. (a) The number of plants out of 32 in each group that flowered. (b) The percentage of flower heads infested with seed-eaters. (c) Mean leaf consumption level after consumption had been scored to six classes ranging from 0 (0% consumed) to 5 (76–99% consumed). (d) Mean number of predators per leaf. Bars are SEs; bars sharing the same letter are not significantly different.

Source: After Kersch-Becker & Lewinsohn (2012).

top-down cascades or bottom-up control

There are some communities that tend, inevitably, to be dominated by bottom-up control, because consumers have little or no influence on the supply of their food resource. The most obvious group of organisms to which this applies is the detritivores (see Chapter 11), but consumers of nectar and seeds are also likely to come into this category (Odum & Biever, 1984) and few of the multitude of rare phytophagous insects are likely to have any impact upon the abundance of their host plants (Lawton, 1989). More generally, the results of a meta-analysis of a large number of studies of possible top-down or bottom-up control are shown in Figure 17.6. Predator manipulation experiments consistently reveal a significant negative effect of the predators on herbivores, and a somewhat less profound but nonetheless significant positive, top-down, trophic cascade effect of the predators on the plants (Figure 17.6a). However, while fertilisation experiments consistently have a significant bottom-up positive effect on the plants themselves, there is no evidence overall for this effect being transmitted further up the food chain to the herbivores, whether or not there are predators present as well (Figure 17.6b). Also, fertilisation from below does not tend to attenuate the top-down effects of predator manipulation (Figure 17.6a). Hence, while both may be important, it seems that top-down effects are more likely to reverberate throughout the system than bottom-up effects.

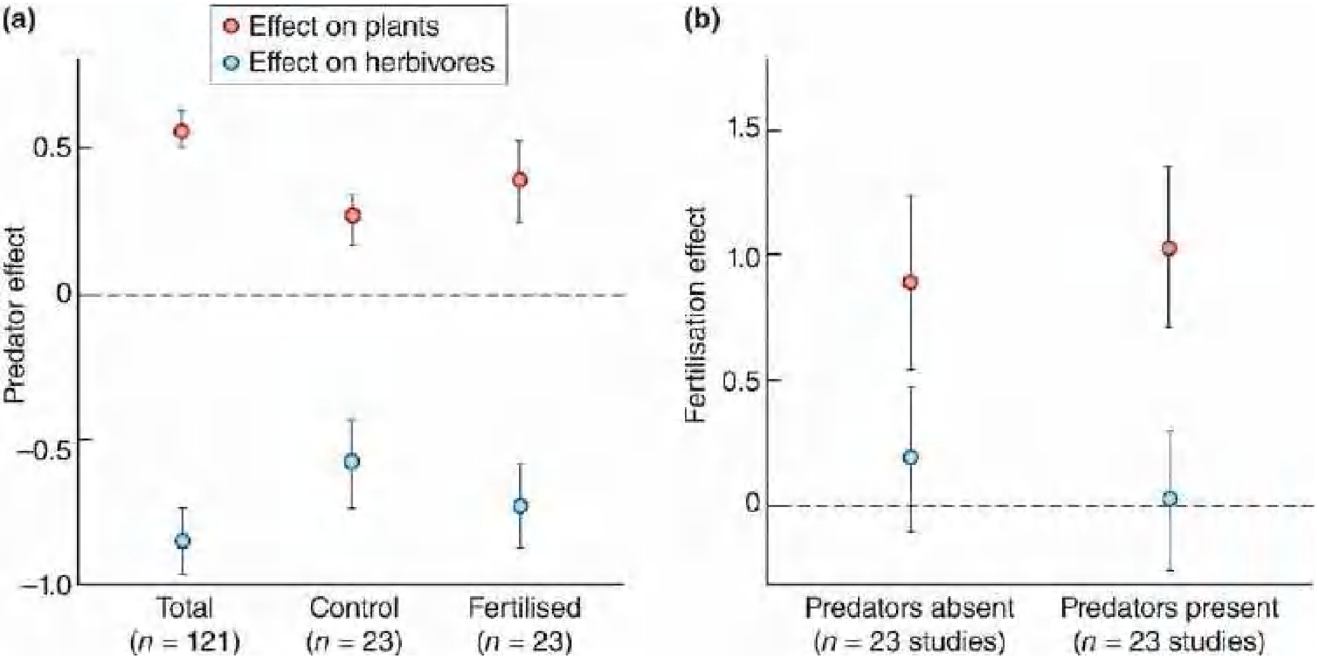


Figure 17.6 Meta-analyses of manipulation studies of top-down and bottom-up community control. (a) Effects of predator manipulation on their herbivore prey and on plants across 121 studies from terrestrial, freshwater and marine systems ('total'), and in the subset of 23 studies that not only manipulated predators but also added nutrients ('Fertilised' compared with 'Control'), generally showing negative effects on herbivores and positive effects on plants. The predator effect is the log ratio of biomass with and without predators. (b) Effects of fertilisation in the presence and absence of predators in the 23 studies in which nutrients were added, generally showing positive effects on plants but not on herbivores. The fertilisation effect is the log ratio of biomass with and without fertilisation. In both cases, zero represents no effect. Bars show SEs.

Source: After Borer *et* al. (2006).

17.1.3 Why is the world green?

community- and species-level cascades

A useful distinction can be made between community- and species-level trophic cascades (Polis *et al.*, 2000). In the former, the predators in a community, as a whole, control the abundance of the herbivores, such that the plants, as a whole, are released from control by the herbivores. But in a species-level cascade, increases in a particular predator give rise to decreases in particular herbivores and increases in particular plants, without this affecting the whole community. Polis *et al.* (2000) proposed that community-level cascades are most likely to be observed in systems where, for example, (i) the habitats are relatively discrete and homogeneous (since in extensive, complex communities cascading effects tend to be diluted out); (ii) the prey population dynamics, including those of the primary producers, are uniformly fast relative to those of their consumers (since in woodlands, for example, effects on trees from higher trophic levels may be slow in appearing, and hence difficult to observe, even if they are occurring); and (iii) the common prey tend to be uniformly edible (since inedible prey are likely only to generate species-level cascades of specialists).

An early preponderance of marine and freshwater studies among the examples of trophic cascades prompted the question ‘Are trophic cascades all wet?’ (Strong, 1992). Schmitz *et al.* (2000) appeared to contradict this by reviewing a total of 41 studies in terrestrial habitats demonstrating trophic cascades, but as Polis *et al.* (2000) pointed out, these were all essentially species-level cascades, with short-term, small-scale effects on plants rather than broader scale responses of the biomass or productivity of a whole community. Subsequently, the number of terrestrial examples has outstripped those from marine and especially freshwater systems. However, it does appear to be the case that community-level cascades are more likely in, though by no means limited to, pelagic communities of lakes and in benthic communities of streams and rocky shores, which are more likely to satisfy the Polis *et al.* criteria, and that terrestrial cascades are more likely to be at the species level (Ripple *et al.*, 2016). Community cascades may also be more likely in more productive systems with longer food chains, as we discuss next.

HSS, EEH and ATCH

The widespread importance of top-down control and community-level trophic cascades was, in fact, foreshadowed in an article by Hairston, Smith and Slobodkin (HSS) in 1960, though they used none of these terms. They asked ‘Why is the world green?’ and answered, in effect, that the world is green because of three-trophic level, community-level cascades: green plant biomass accumulates because predators keep herbivores in check. Others, notably Murdoch (1966), challenged these ideas, emphasising that many plants have evolved physical and chemical defences that make life difficult for herbivores (see Section 9.3), such that herbivores may be competing fiercely for a limited amount of palatable plant material, and their predators may, in turn, compete for scarce herbivores, such that a world controlled from the bottom-up may still look green. Nonetheless, as an overall generality, the HSS view has prevailed.

Oksanen *et al.* (1981), in what is often referred to as the Ecosystem Exploitation Hypothesis (EEH), elaborated on the HSS ideas by linking them, through a model, to levels of primary productivity, and to food chain length, which is likely to increase with productivity (see Section 17.2.8). At the very lowest levels of primary productivity, for example in the polar tundra, there may be effectively only one trophic level, the plants or lichens, especially if we acknowledge the existence of a trophic level only if it is capable of exerting some degree of control on the level below. With somewhat higher (but still low) levels of productivity, there may be insufficient herbivores to support effective populations of predators and hence only two trophic levels. At higher levels still, and this will often be the case, productivity is sufficient for three trophic levels, and the world will indeed be green as argued by HSS. While at the very highest productivities, four trophic levels may be supported, and the plants will once again be limited by their herbivores.

In a further elaboration, Polis and Strong (1996) emphasised that the energetic route from primary producers to top predators need not necessarily go through herbivores. Much primary production dies before it is consumed and enters the detritivory food chain, but this is linked to the herbivory chain, not only because the two typically share top predators, but also because many species at lower trophic levels are omnivores that participate in both, consuming both live and dead material. This generates a new, albeit rather complex community module (Figure 17.7) as the basis for what has been called the Apparent Trophic Cascade Hypothesis (ATCH), so named because the herbivores and detritivores comprise an apparent competition submodule (see Figure 16.1b) with their shared predators.

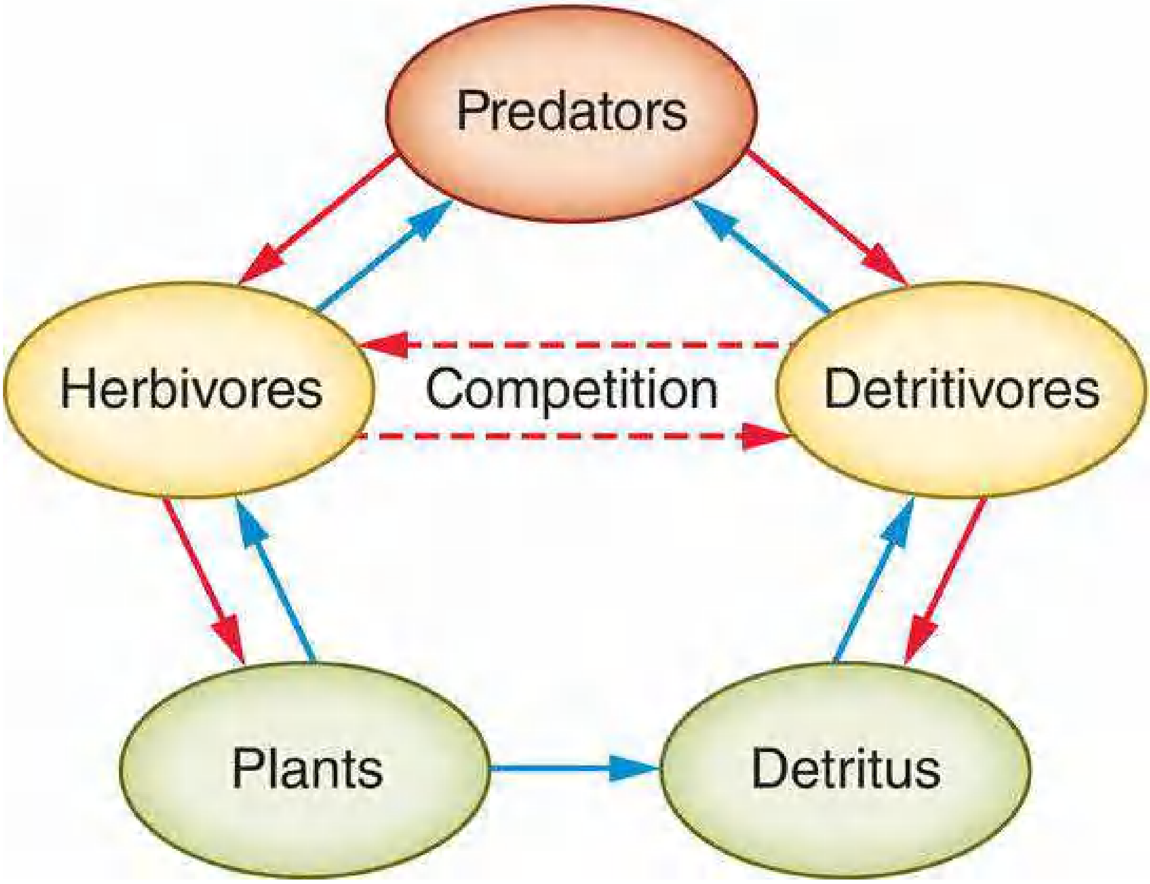
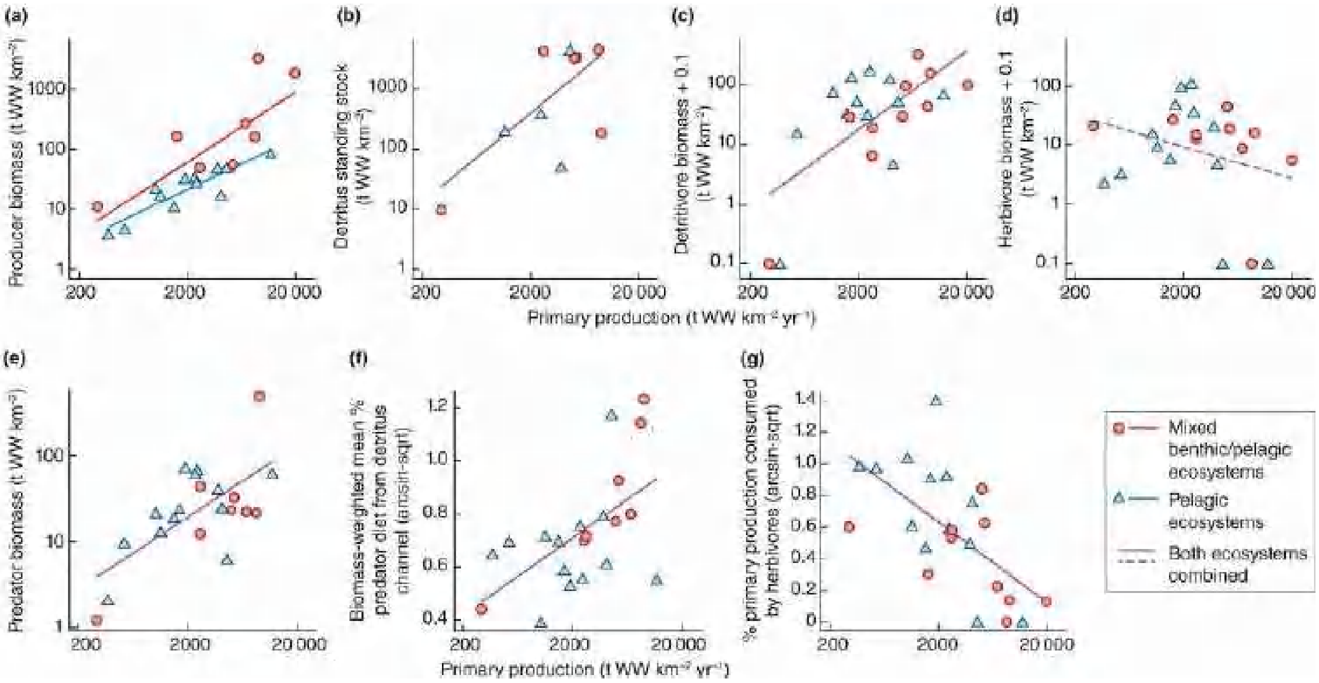


Figure 17.7 The community module underlying the apparent trophic cascade hypothesis. Some plant material becomes detritus before it is consumed by herbivores, and because the herbivores and detritivores share many of the same predators, there is apparent competition between them.

In a rare test of these ideas, Ward *et al.* (2015) examined data from 23 large marine food webs where it was possible to estimate not only the overall biomass for the various trophic categories, but also the proportions of predator biomass, at the tops of the webs, that had been derived via the detritivory and herbivory channels (Figure 17.8). Note first that these data cannot be used to test HSS or EEH directly, because both assume only a single food chain linking primary producers and predators. Indeed, any attempt to do so by ignoring part of the food web in Figure 17.7 may lead to an erroneous conclusion as a consequence – for example, that the predators were responding to herbivores when in fact they were responding to detritivores. The data themselves show that at the bottom of the food web, as productivity increases, the biomasses of both the primary producers and the detritus increase (Figure 17.8a, b). At the next level up, however, while the biomass of the detritivores also increases with productivity (Figure 17.8c) that of the herbivores declines, though not significantly (Figure 17.8d). And when the two chains converge at the top, the biomass of predators also increases with productivity (Figure 17.8e). Crucially, however, as productivity increases, an increasing proportion of the energy making its way from the bottom to the top of the food web passes through the detritus and the detritivores, rather than the herbivores (Figure 17.8f, g), perhaps as a reflection of the tendency for plants to invest more in defence at higher levels of resource availability (Ward *et al.*, 2015; see Section 9.3.5).



[Figure 17.8](#) A test supporting the Apparent Trophic Cascade Hypothesis. (a–e) Analysis of how the flow of biomass (tonnes of wet weight per km^2) varied in 23 marine food webs as the level of primary production (tonnes of wet weight per km^2 per year) varied. (a) Primary producers increase (red line and symbols, mixed benthic/pelagic ecosystems, P = 0.011; blue line and symbols, pelagic ecosystems, P < 0.001). (b) Detritus increases (purple line, both ecosystems combined; P = 0.037). (c) Detritivores increase (P < 0.001). (d) Herbivores decrease (but effect is not significant, dashed purple line; P = 0.18). (e) Predators increase (P < 0.001). (f) As the level of primary production increased, the percentage of predator diet derived from detritus increased (P = 0.007). (g) As the level of primary production increased, the percentage consumed by herbivores decreased (P = 0.001).

Source: After Ward *et al.* (2015).

If we were to consider the herbivory chain in isolation, this would provide support for the EEH – an increasing propensity for community-level cascades as productivity increases: higher abundances of predators giving rise to lower abundances of herbivores allowing higher abundances of primary producers. However, considering the detritivory chain in isolation suggests bottom-up control – increasing productivity giving rise to more primary producers, more detritus, more detritivores and ultimately more predators. But linking the two, as we must, and acknowledging the shift from herbivory to detritivory as productivity increases, provides support for the ATCH – an increasing tendency for the detritivory chain to feed the biomass of predators, helping or perhaps allowing them to exert top-down, trophic cascade control on the trophic levels beneath them (Ward *et al.*, 2015). Indeed, there is positive feedback between the two chains: detritivores feeding predators that suppress herbivores, allowing greater plant survival, more detritus, more detritivores, and so on. To what extent such patterns will be repeated in future studies remains to be seen. But it seems clear that a broader perspective – food webs (see [Figure 17.7](#)) rather than food chains – is going to be required. We now turn to food webs.

17.2 Food web structure, productivity and stability

Any ecological community can be characterised by its structure (number of species, interaction strength within the food web, average length of food chains, etc.), by certain quantities (especially biomass and the rate of production of biomass, which we can summarise as ‘productivity’) and by its temporal stability (Worm & Duffy, [2003](#)). In the remainder of this chapter, we examine some of the interrelationships between these three. Much of the very considerable recent interest in this area has been generated by the understandable concern to know what might be the consequences of the inexorable decline in biodiversity (a key aspect of structure) for the stability and productivity of biological communities.

We will be particularly concerned with the effects of food web structure (food web complexity initially; food chain length and a number of other measures in subsequent subsections) on the stability of the structure itself and the stability of community productivity. We must remember from the outset, however, that progress in our understanding of food webs depends critically on the quality of data that are gathered from natural communities. Several authors have called this into doubt, particularly for earlier studies, pointing out that organisms have often been grouped into taxa extremely unevenly and sometimes at the grossest of levels. For example, even in the same web, different taxa may have been grouped at the level of kingdom (plants), family (Diptera) and species (polar bear). Some of the most thoroughly described food webs have been examined for the effects of such an uneven resolution by progressively lumping web elements into coarser and coarser taxa (Martinez, 1991 ; Hall & Raffaelli, 1993 ; Thompson & Townsend, 2000). The uncomfortable conclusion is that most food web properties seem to be sensitive to the level of taxonomic resolution that is achieved. These limitations should be borne in mind as we explore the evidence for food web patterns in the following sections.

Ecologists are interested in community stability for two reasons. The first is practical – and pressing. The stability of a community measures its sensitivity to disturbance, and natural and agricultural communities are being disturbed at an ever-increasing rate. It is essential to know how communities react to such disturbances and how they are likely to respond in the future. The second reason is less practical but more fundamental. The communities we actually see are, inevitably, those that have persisted. Persistent communities are likely to possess properties conferring stability. The most fundamental question in community ecology is: ‘Why are communities the way they are?’ Part of the answer is therefore likely to be: ‘Because they possess certain stabilising properties that have allowed them to persist’.

17.2.1 What do we mean by ‘stability’?

First, however, it is necessary to distinguish between the various different types of stability and disturbance, though regrettably, agreement on the meaning of terms is not total (see Donohue *et al.*, 2016). We need to consider the two terms together, because it is natural to think of the stability of a community in response to a disturbance. For disturbance, the most important distinction is between a *pulse* – a single event with the potential to alter the state of a community – and a *press* – a persistent change in some factor or factors with the potential to determine the state of the community.

resistance, resilience and robustness

Of the various aspects of stability, an initial distinction can be made between the resilience of a community and its resistance. *Resistance* describes the extent to which the community is altered by a disturbance – a maximally resistant community is not altered at all. *Resilience* describes the speed with which a community returns to its former state after it has been disturbed and displaced from that state – a maximally resilient community bounces back immediately. However, the term resilience has also sometimes been used to describe an ability to maintain function in the face of disturbance (Hodgson *et al.*, 2015): more like resistance. A third term, commonly applied, is *robustness*, which has been used to describe the tendency of a community to suffer subsequent, secondary extinctions following the primary extinction of one of its species. Robust communities are thus those that suffer fewer secondary extinctions. Specifically, robustness has been defined as the proportion of the original number of species in a community that need to suffer a primary extinction in order to generate a total number of extinctions (primary plus secondary) equal to a set percentage (typically 50%) of the species originally present (Dunne &Williams, 2009).

asymptotic stability and variability

A further distinction is between local stability and global stability. *Local stability* describes the tendency of a community to return to its original state (or something close to it) when subjected to a small disturbance. *Global stability* describes this tendency when the community is subjected to a large disturbance. We should also distinguish between, on the one hand, *asymptotic stability* – the tendency of a system to converge on a stable 'point', which is therefore most often used in theoretical, mathematical modelling studies, from where the term is derived – and on the other hand, some measure of *variability* in a key characteristic of a community, which is therefore more often used in empirical studies, where it is natural to describe temporal variability in a set of data. Lastly, it remains for us to specify the aspect of the community on which we will focus. Ecologists have often taken a demographic approach, concentrating on stability in the *structure* of a community, for example the number of species, or variations in their abundance. However, as we shall see, it is also possible to focus on the stability of *aggregate ecosystem properties*, especially productivity.

no simple answers

Having established that there are several alternative valid meanings of 'stability', several aspects of a community that may be more or less stable, and indeed several types of disturbance that may put the stability of a community to the test, it would be wrong to suppose that there is a single, simple answer to questions like 'Are more complex communities more stable?', though many have sought simple answers. In reality, the answer may vary with the type of stability, the aspect of the community in question and on the temporal and spatial scale of observation. Getting several, sometimes apparently contradictory, answers in the sections that follow, therefore, need not indicate a lack of understanding (though there is undoubtedly much that we do not know). Rather, it reminds us that answers to the question should themselves specify the type of stability being considered, the property of the community and so on. They should not be presumed to apply to other types and other properties. The subject matter of ecology is too multifaceted for that (see also Loreau, 2010a).

17.2.2 Strong interactors and keystone species

Whatever variant we use, 'stability' usually means stability in the face of a disturbance, and most disturbances, in practice, lead to the loss of one or more populations from a community. Some species, though, are more intimately woven into the fabric of the food web than others. A species whose removal produces a significant effect (extinction or a large change in density) in at least one other species may be thought of as a strong interactor, and the removal of some strong interactors leads to significant changes spreading throughout the food web. We saw in the previous chapter, for example (see Section 16.3), that removal of the starfish *Pisaster* on a rocky shore led to the extinction of several species, producing a community with a very different species composition and, to our eyes, an obviously different physical appearance. This prompted Paine (1969), who carried out the study, to refer to such species as *keystone species*, since a keystone is the wedge-shaped block at the highest point of an arch that locks the other pieces together, such that its removal causes the arch to collapse.

what is a keystone species?

The term has been widely used ever since, but an ever-widening range of definitions and applications have caused some to question its value altogether. Cottee-Jones and Whittaker (2012) discuss the evolution and taxonomy of its usage. It is now usually accepted that keystone species need not be top predators, like Paine's starfish, but can also occur at other trophic levels (Hunter & Price, 1992). For example, European rabbits (*Oryctolagus cuniculus*), plateau pikas (*Ochotona curzoniae*) and prairie dogs (*Cynomys* spp.), often thought of as pests, play a keystone role in many parts of the world both for their role, as grazers, in maintaining prairie grassland ecosystems, and for the very large number of predatory species they maintain, often of conservation importance (Delibes-Mateos *et al*., 2011). Keystone mutualists (Mills *et al*., 1993) may also exert influence out of proportion to their abundance: examples include a pollinating insect on which an ecologically dominant plant relies, or a nitrogen-fixing bacterium supporting a legume and hence the whole structure of a plant community and the animals reliant on it. However, it is not helpful simply to award keystone status to any species whose removal from a community would have a drastic effect. In particular, 'ecological dominants' at lower trophic levels, where one species may provide the resource on which a whole myriad of other species depend – for example, a coral in a coral reef, or the oak trees in an oak woodland – are arguably better described as 'foundation species' than as keystones (Cottee-Jones & Whittaker, 2012), in the sense, continuing the architectural analogy, that a structure will also, but more obviously, collapse if its foundations are removed.

These considerations have led others to define a keystone more narrowly as a species whose impact is 'disproportionately large relative to its abundance' (Power *et al*., 1996). This certainly captures the architectural keystone concept, but even with this increased focus, and while various indices have been proposed to measure 'keystone-ness' (Piraino *et al*., 2002), agreeing on one index has been problematic, and collecting the data to apply such metrics even more so, whether from observation or experimentation (Cottee-Jones & Whittaker, 2012). Nonetheless, it remains important to acknowledge that while all species no doubt influence the structure of their communities to a degree, some are far more influential than others, often out of all proportion to their own abundance. Including 'keystone species' in the ecological lexicon is therefore valuable as a metaphor, as are attempts to quantify species' influence on community structure. But quibbling about which species do and which do not deserve keystone status – calling in the keystone cops – is unlikely to be productive.

APPLICATION 17.2 Sea otters – keystone initiators of a trophic cascade

Recognising the concept of keystone species may also be important from a practical point of view because keystone species are likely to have a crucial role in conservation: changes in their abundance will, by definition, have significant repercussions for a whole range of other species. For example, the sea otter (*Enhydra lutris*) is a keystone species in coastal habitats along the Pacific coast of North America, including those in California (Estes, 2015). Sea otters there, *E. lutris nereis*, were hunted almost to extinction by fur traders in the 18th and 19th centuries, but having been listed as 'threatened' under the US Endangered Species Act (ESA) in 1977, they are now increasingly widespread and abundant as a result of their protected status. Sea otters consume many species of prey, but in particular they maintain at low abundance the sea urchins that would otherwise themselves consume large amounts of kelp (seaweed). They therefore indirectly promote the abundance of kelp in a classic trophic cascade. They also act as a keystone species in creating a whole kelp forest community that itself provides habitats and resources for many other species, including abalones (gastropods in the family Haliotidae) that are the basis for commercial and recreational fisheries. Thus, sea otters both promote the abundance of abalones in their role as a keystone species and reduce their abundance through predation (Figure 17.9a).

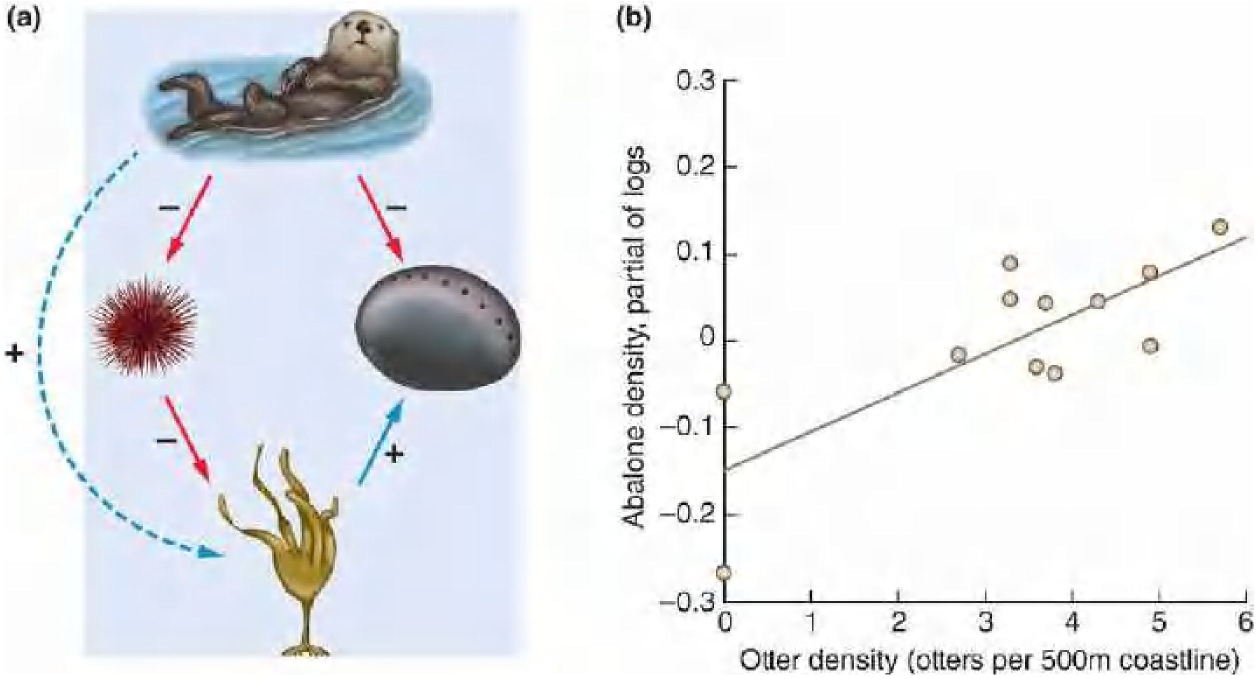


Figure 17.9 The keystone status of sea otters along the Pacific coast of North America. (a) The food web in which sea otters (top) consume both sea urchins (middle left), which consume kelp (below), and abalones (middle right), for which the kelp provides key resources and habitat. Sea otters therefore also have an indirect positive trophic-cascade effect on kelp. (b) The density of black abalone, *Haliotis cracherodii*, increases with sea otter density along the central Californian coast ($P = 0.004$). Abalone density is measured as partial residuals in a multiple regression analysis in which it was related to both habitat quality and sea otter density.

Source: (a) After Estes (2015). (b) After Raimondi *et al.* (2015).

Commercial fishing for black abalone, *Haliotis cracherodii*, started in California in 1968 and peaked in 1973 but has been illegal since 1993 as a result of serious declines in abundance due, in large part at least, to a disease, withering syndrome. In 2009 it was listed as ‘endangered’ under the ESA. Thus, with both the sea otter and the black abalone focuses for conservation, it is important to understand how the indirect positive effect of the otter on the abalone through its keystone trophic cascade is balanced against its direct negative effect as a predator. Might conservation of one be in conflict with conservation of the other? To address this question, 12 sites in California supporting black abalone but with varying densities of sea otters, and with no withering disease, were surveyed. The densities of the two were positively correlated (Figure 17.9b). This suggests that the keystone influence of the otters on black abalone outweighs its effects as a predator, and that conservation of the sea otters should not constitute a threat to the endangered black abalone.

APPLICATION 17.3 Humans as hyper-keystones

It has been argued that the keystone metaphor may also usefully be extended to humans as a way of focusing on the wholly disproportionate effect that our ingenuity and sophistication allows us to have on ecosystems at local and global scales, making us 'hyper-keystones' (Worm & Paine, 2016). Indeed, the prefix 'hyper' is appropriate not only as a reflection of the extremity of the reach of our influence, but also because our actions are often directed at species that are themselves keystones. Figure 17.10 elaborates some examples of this for the Pacific north-west of the USA.

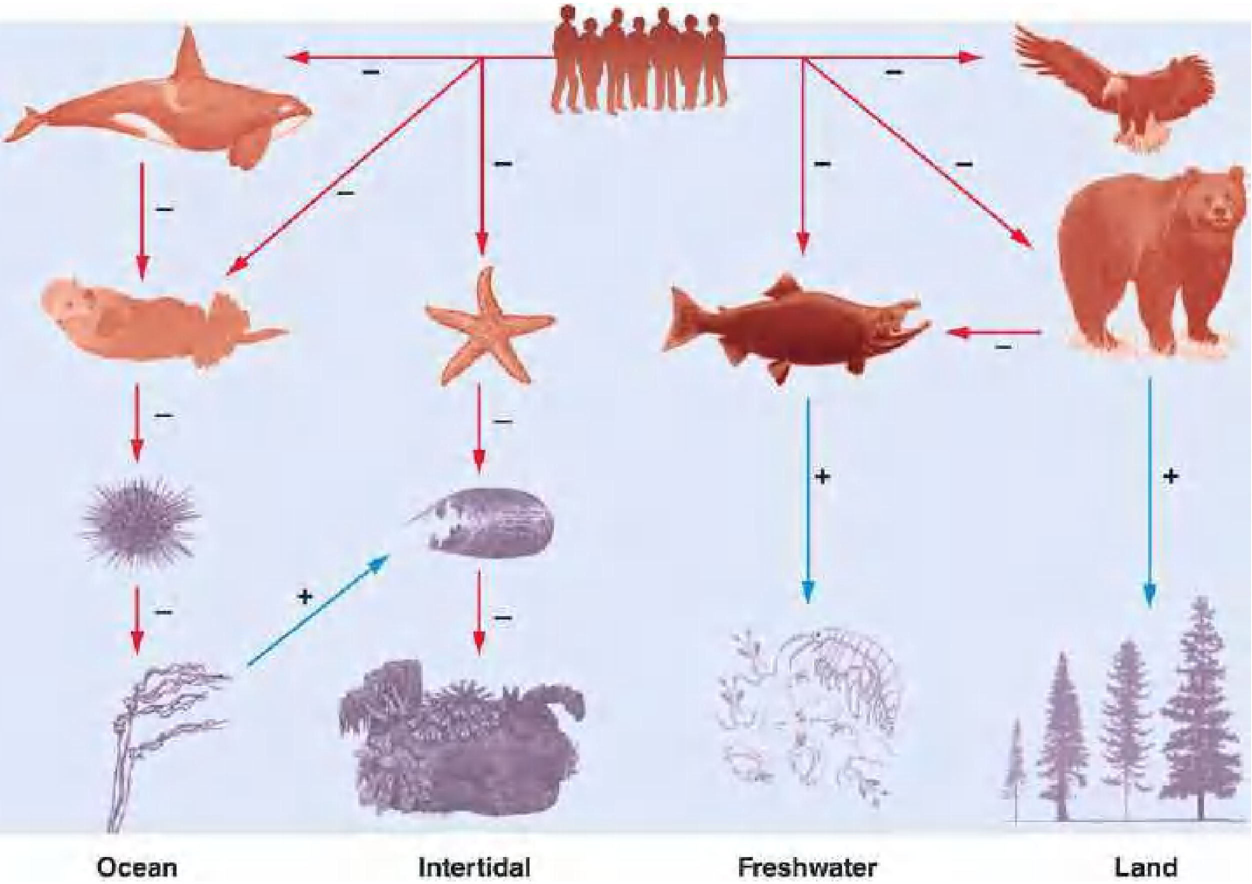


Figure 17.10 Humans are hyper-keystone species in the Pacific north-west of the USA. Across a range of habitats, humans have profound detrimental effects on a number of species that have themselves been identified as keystone species: orcas (whaling), sea otters (hunting), starfish (climate change effects), salmon (hunting), eagles (persistent pollutants) and bears (hunting). The food web effects of these keystones are also indicated in outline.

Source: After Worm & Paine (2016).

17.2.3 Complexity and stability in model communities

older and newer conventional wisdoms

The connections between food web structure and food web stability have preoccupied ecologists for well over half a century. Initially, from the 1950s, the 'conventional wisdom', promoted especially by Elton (1958) and MacArthur (1955), was that increased complexity within a community leads to increased stability; that is, more complex communities are better able to remain structurally the same in the face of a disturbance such as the loss of one or more species. Increased complexity, then as now, was variously taken to mean more species, more interactions between species, greater average strength of interaction, or some combination of all of these things.

If we focus initially on structural aspects of communities, and on the individual populations that comprise a community, the conventional wisdom was undermined by a watershed theoretical study by May (1972). He constructed model food webs comprising a number of species, and examined the way in which the population size of each species changed in the neighbourhood of its equilibrium abundance (i.e. the local stability of individual populations). Each species was influenced by its interaction with all other species, and the term β_{ij} was used to measure the effect of species *j*'s density on species *i*'s rate of increase. The food webs were 'randomly assembled', with all self-regulatory terms (β_{ii}, β_{jj}, etc.) set at –1, but all other β-values distributed at random, including a certain number of zeros. The webs could then be described by three parameters: *S*, the number of species; *C*, the 'connectance' of the web (the fraction of all possible pairs of species that interacted directly, i.e. with β_{ij} non-zero); and $\bar{\beta}$, the average 'interaction strength' (i.e. the average of the non-zero β-values, disregarding sign). May found that increases in the number of species, in connectance and in interaction strength all tend to increase instability. Yet each of these represents an increase in complexity. Thus, in this model, complexity leads to *in*stability, indicating, at least, that there is no necessary, unavoidable connection linking increased complexity with increased stability.

population stability: no single answer

Neither May nor anybody since has suggested that real communities are constructed 'at random'. Nonetheless, the random model has acted as a useful benchmark, allowing many subsequent studies to investigate systems that depart from this random structure – often departures inspired by observations of natural communities – and determine the consequences of these departures for community stability. By no means all generate the same pattern as May's original model. Hence, overall, there is no all-purpose answer to the question 'How does a decrease in the complexity of a model community affect the stability of its structure?' The answer depends on the details of the model, the chosen definition of stability and the nature of the disturbance (Ives & Carpenter, 2007). Nonetheless, one pattern that emerges from a wide variety of models is that robustness, the curtailing of knock-on, secondary extinctions following a primary extinction, tends to increase with both species richness and connectance (Dunne & Williams, 2009).

variance-to-mean relationships

We get a somewhat different perspective on the relationship between complexity and stability if we focus not on the stability of community structure but on the variability (that is, the *in*stability) of the individual population sizes within a community. This perspective also allows us to connect the stability of populations to the stability of aggregate community properties like biomass and productivity, since these reflect the combined abundances of the individual populations. The relationship between species richness and the variability of populations is based on the more general relationship known to exist between the mean (*m*) and variance (s^2) of abundance of individual populations over time (Tilman, 1999), as follows:

$$s^2 = cm^z, \tag{17.1}$$

where *c* is a constant and *z* is the so-called scaling coefficient. Generally, values of *z* lie between 1 and 2 (Murdoch & Stewart-Oaten, 1989; Cottingham *et al*., 2001), and the variability of individual species' abundances therefore tends to increase with their mean abundance. If we now assume, for simplicity, that in a community of N species those m individuals are shared equally among them, such that each has an abundance m/N, then we can use standard statistical arguments to determine the effect of increasing species richness, N, on both the variability of individual population abundances, V_P, and the variability of the overall community abundance, V_C (Tilman, 1999). Skipping these details and moving straight to the conclusions ($\propto$ means proportional to):

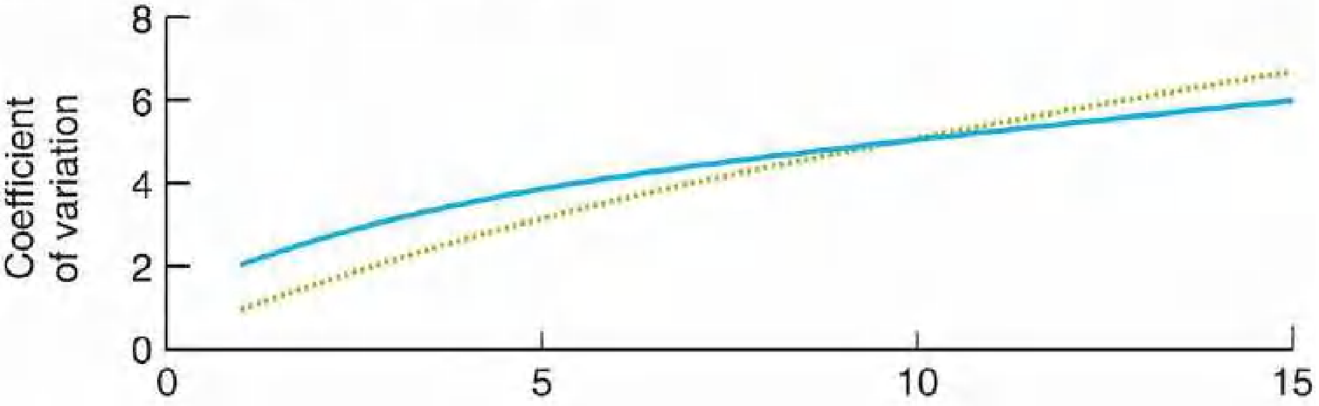
$$V_P \propto N^{(z/2-1)} \tag{17.2}$$

and

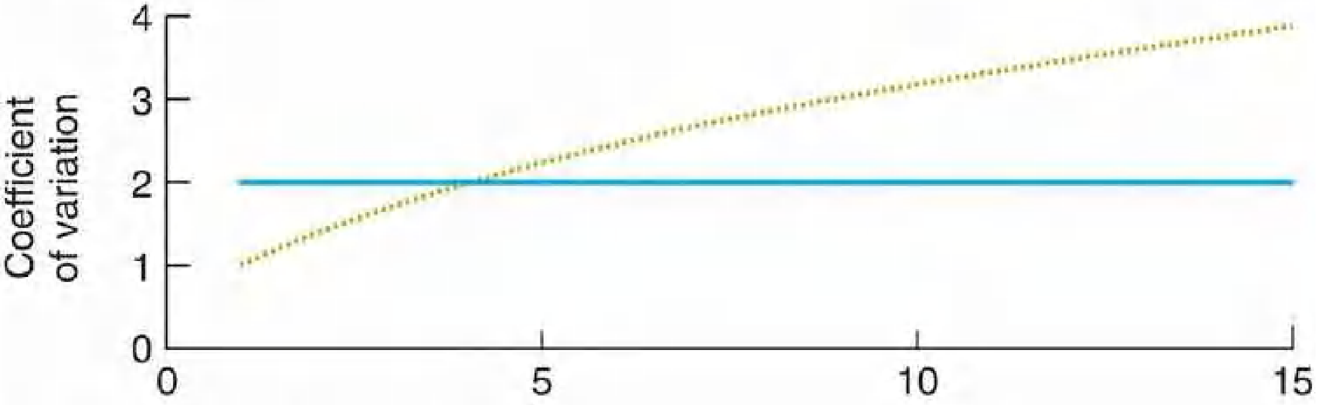
$$V_C \propto N^{(z-1)/2}. \tag{17.3}$$

That is to say: since variability increases if the exponent is positive, the variability of population abundances is predicted to increase with species richness only if $z > 2$, whereas the variability of community biomass is predicted to increase with richness as long as $z > 1$. Within the normal range ($1 < z < 2$), therefore, population stability is predicted to *de*crease with richness, as it did in May's original model, whereas the stability of aggregate properties is predicted to increase with richness (Figure 17.11).

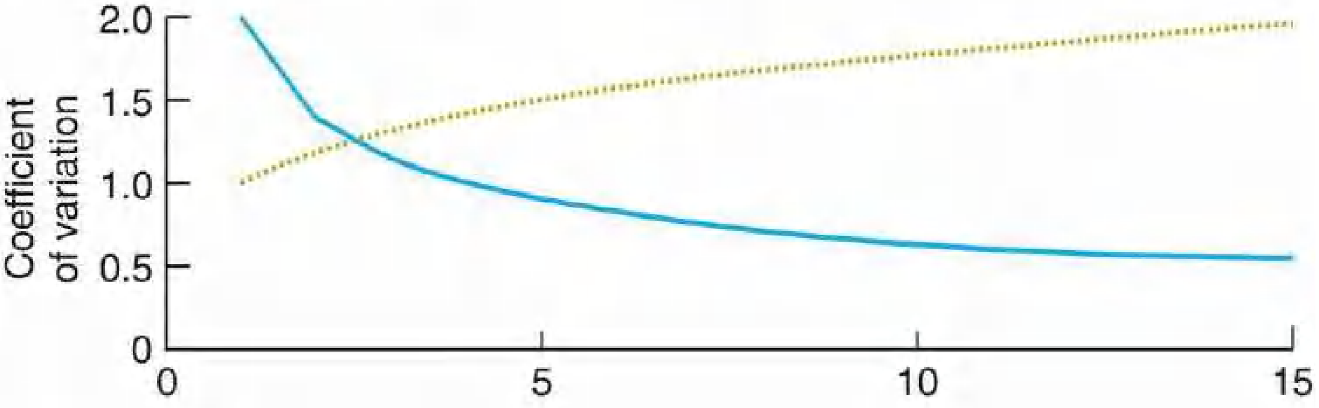
(a) Unusually low z ($z = 0.6$)



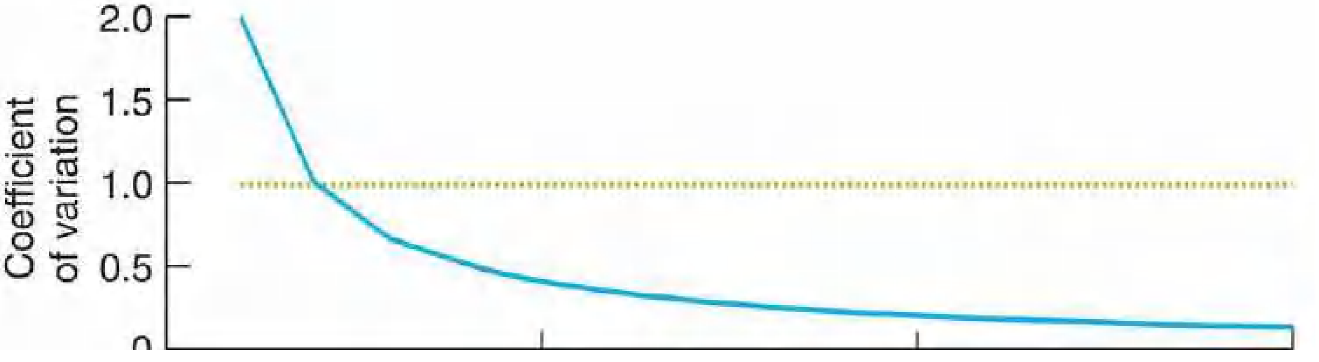
(b) $z = 1.0$



(c) Typical z ($z = 1.5$)



(d) $z = 2.0$



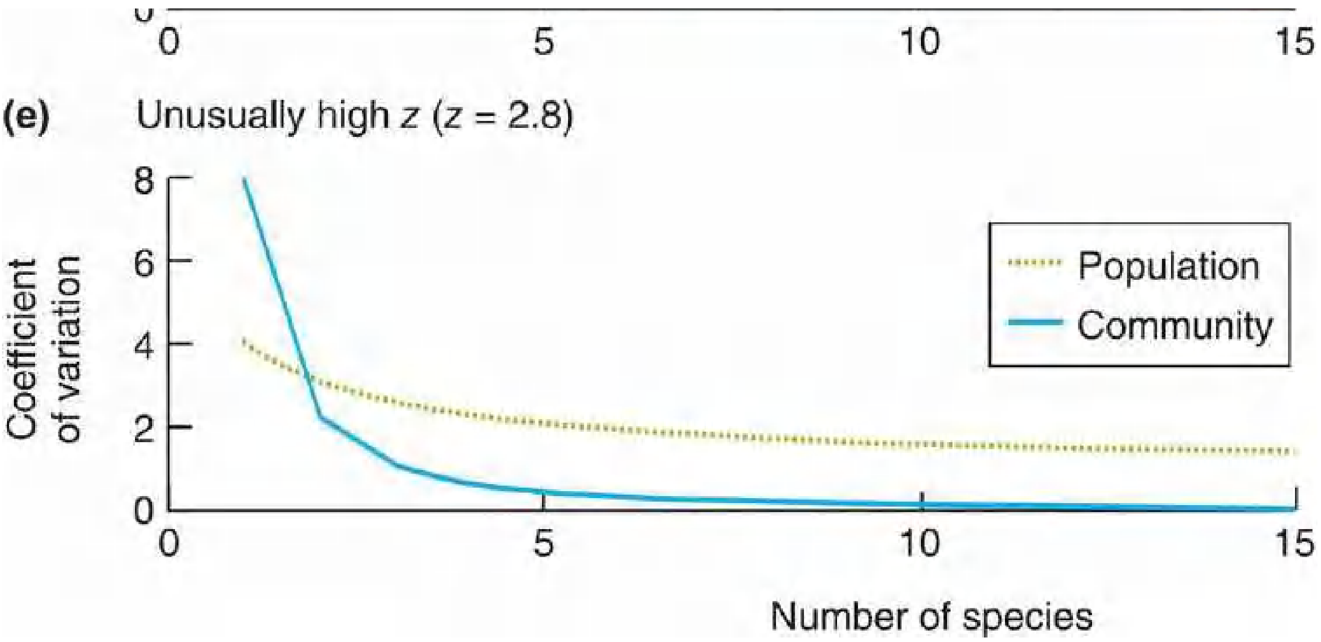


Figure 17.11 Effect of species richness on the temporal variability of population size and aggregate community abundance. Temporal variability was measured as the coefficient of variation, CV, in model communities in which all species were equally abundant and had the same CV, for various values of the scaling coefficient, z, in the relationship between the mean and variance of abundance (Equation 17.1). (a) $z = 0.6$, an unusually low value. (b) $z = 1.0$, the lower end of typical values. (c) $z = 1.5$, a typical value. (d) $z = 2.0$, the upper end of typical values. (e) $z = 2.8$, an unusually high value.

Source: After Cottingham *et al.* (2001).

the stability of aggregate properties increases with richness

Staying now with changes in the stability of aggregate properties, the predictions from theory are much less equivocal and model-specific than they were when we looked at community structure earlier (Cottingham *et al.*, 2001; see also Tilman *et al.*, 2014). To begin with, we have two purely statistical effects leading to an increase in stability with richness. The first is *variance-to-mean scaling* amongst population abundances, described earlier. But this acts alongside a *portfolio effect*, whereby as long as the fluctuations in different populations are asynchronous, there is an inevitable statistical averaging when populations are added together – when one goes up, others will be going down – and this tends to increase in effectiveness as richness increases.

competition, and complementarity and selection effects

These effects may themselves be enhanced by a number of ecological processes. As richness increases, the number of *competitive interactions* tends to increase, leading in turn to an increase in negative covariance between populations (one increasing at the expense of another), further boosting the statistical averaging effects. Competition, too, insofar as it is accompanied by niche differentiation, tends to give rise either to *complementarity effects*, where larger combinations of species are likely to occupy greater proportions of the overall niche space (and there may also be facilitation), or to *selection effects*, where more diverse communities are more likely to contain the most productive species, which, being productive, are likely to come to dominate the communities.

The complementarity and selection effects are effects on productivity. With complementarity effects there is likely to be 'overyielding': biomass production being greater than the average of the constituent species in monoculture, because the species, collectively, are occupying a larger niche space. Moreover, if the complementarity effects are strong, and especially if there are also selection effects, there may be *transgressive overyielding*: overall biomass being greater in the community than even the single most productive species in monoculture. Of course, with selection effects alone, the highest productivity cannot exceed that of the most productive species in the mixture. We return to these effects of species richness on overall productivity in Section 20.3.6. Note, though, that these complementarity and selection effects may also enhance stability (reduce the temporal variance in overall productivity), since both give richer communities a greater capacity to sustain productivity as conditions or the balance of resources change.

Finally, and related to this, there is a greater opportunity in richer communities for different species to respond differently to external disturbances, so that if one is harmed, others may be unaffected or may even benefit, such that overall responses are generally dampened in what has been called an *insurance effect*. The term 'insurance effect' has sometimes also been used to describe the whole collection of biological effects enhancing stability as species richness increases. For a variety of interrelated reasons, therefore, in richer communities, the dynamics of these aggregate properties are predicted to be more stable.

stability cascades in food chains

Note, though, that by leaning so heavily on the effects of competition, these arguments are most applicable at a single trophic level – a point to which we will return when we look at empirical studies later in the chapter. Shanafelt and Loreau (2018), by contrast, developed models with multiple trophic levels to investigate the effects of food chain length on stability. They found cascade effects for stability linked to those we noted for abundance in Section 17.1. The highest trophic level tended to be the most stable (measured as numerical invariability) and the level just beneath the top to be the least stable, with a cascading alternation of stability moving from level to level further down the chain. Their informal explanation for this is that the top predators are constrained only by the availability of food in the trophic level below, whereas variability in that pre-apex level is generated both by the pressures from the unconstrained predators above and from resource shortages below. In their model, food chain length itself tends to increase with the level of resource input to the primary producers at the lowest trophic level. (We discuss this and other determinants of food chain length in Section 17.2.8.) Hence, the disturbance of adding or removing trophic levels, as productivity or other factors vary, need not so much alter the overall stability of a community as shift the balance of stability between the trophic levels, increasing the stability of some but decreasing the stability of others.

17.2.4 Relating theory to data: aggregate properties

In parallel with the consistent prediction from theory of an increase in the stability of aggregate community properties with species richness, there has been widespread confirmation of this association in natural communities (Tilman *et al.*, 2014). Tilman (1996) found a strong positive effect of richness on the stability of community biomass in natural Minnesota (USA) grasslands, and subsequently obtained a similar pattern in experimental plots planted with between one and 16 species (Tilman *et al.*, 2006). McGrady-Steed *et al.* (1997) manipulated richness in aquatic microbial communities (producers, herbivores, bacterivores and predators) and found that variation in another ecosystem measure, carbon dioxide flux (a measure of community respiration), also declined with increasing richness.

In a compilation of eight coordinated grassland studies from across Europe, Hector *et al.* (2010) went further in not only establishing positive correlations between temporal stability and species richness (Figure 17.12a) but also associating temporal stability, in turn, with increases in overyielding (Figure 17.12b) and temporal asynchrony (Figure 17.12c). At each of the eight sites, experimental plots were established with between one and 32 species, and the biomass of each was measured over each of three years. Overyielding was calculated as the deviation of the yield of mixtures from the null expectation of the weighted mean of the monoculture yields of the constituent species. Temporal asynchrony was calculated as a correlation of biomasses (0 is uncorrelated, −1 is 'perfectly' asynchronised, etc.), but only for species-pairs, since as more species are added, the overall correlation inevitably tends towards zero. Hence, in this case both the pattern – the association between species richness and stability – and two underlying processes were supported.

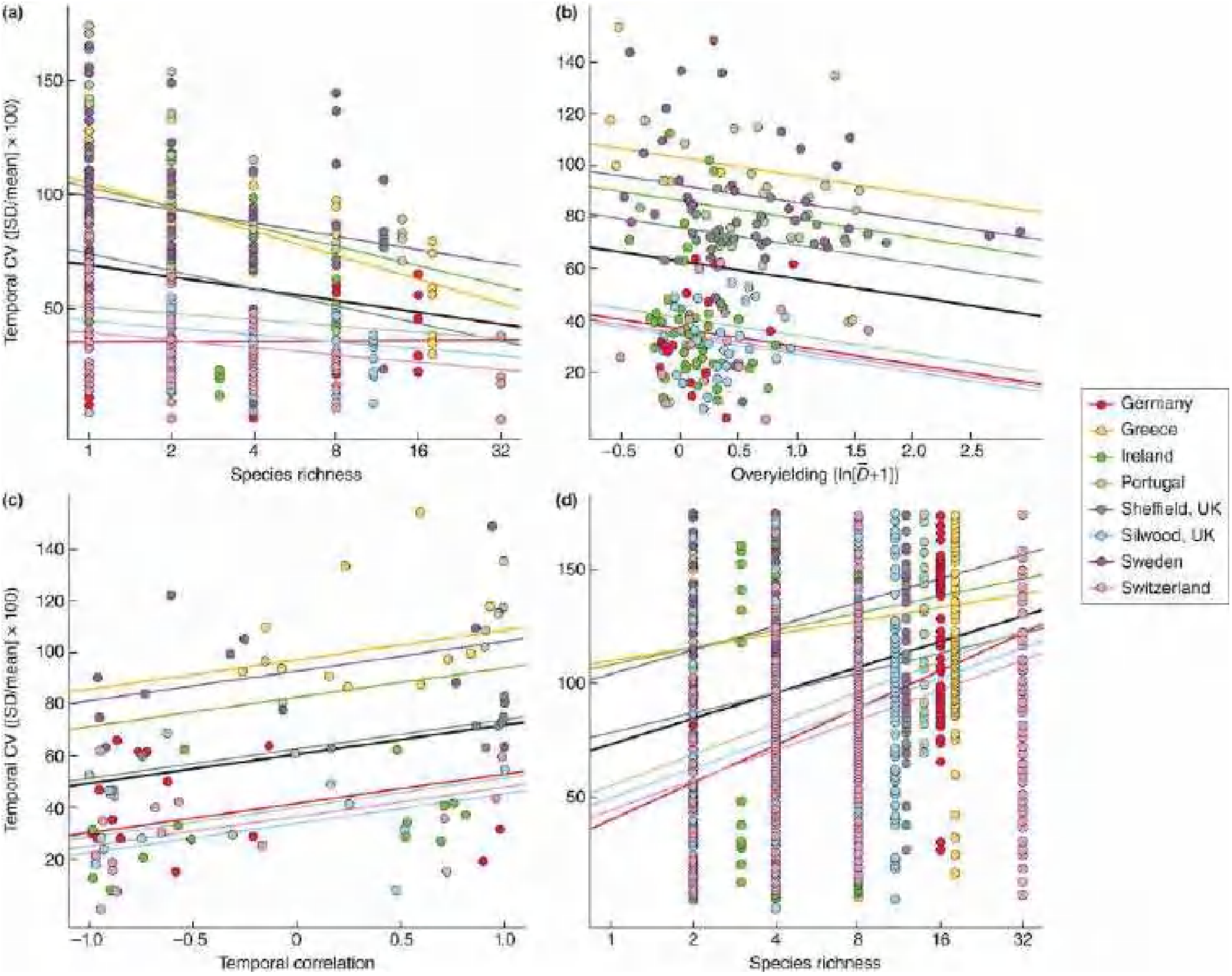


Figure 17.12 Species richness stabilises grassland productivity through overyielding and population asynchrony but destabilises population variability. (a) Temporal coefficients of variation (CVs) of community biomass measured at replicate plots containing between one and 32 grassland species at eight sites, as indicated. The black line shows the overall effect, which is significant ($P < 0.05$), as are those for all individual sites except Germany. Note that as temporal CVs decrease, stability increases. (b) Equivalent relationships between temporal CVs of community biomass and overyielding (the difference between mixture and summed monoculture yields). All relationships are significant ($P < 0.05$). (c) Equivalent relationships between temporal CVs of community biomass and population asynchrony (the correlations between the biomasses of species-pairs). All relationships are significant ($P < 0.05$). (d) Equivalent relationships between temporal CVs for population biomass and species richness. All relationships are significant ($P < 0.05$).

Source: After Hector *et al.* (2010).

APPLICATION 17.4 Complementarity and food security

Food security requires more than there being enough food for people to eat. Access to that food and a stability of supply are also crucial. Lack of food security is one of the many problems facing indigenous people around the world, including those relying on salmon fisheries in western North America. A study of the First Nations salmon fisheries within the Fraser River catchment of British Columbia, Canada, therefore, sought to relate the stability of supply to a cline of declining diversity from south to north in the catchment. Some of the results are shown in Figure 17.13. Stability was measured by calculating the portfolio effect of statistical averaging among the asynchronous variations in the catches of different salmon species. This was the difference between the 'expected' coefficient of variation (CV) in catch biomass from year to year (the average of the CVs of individual species, weighted by their relative biomasses) and the observed CV in total biomass. Thus, the greater the portfolio effect, the greater the stability in supply.

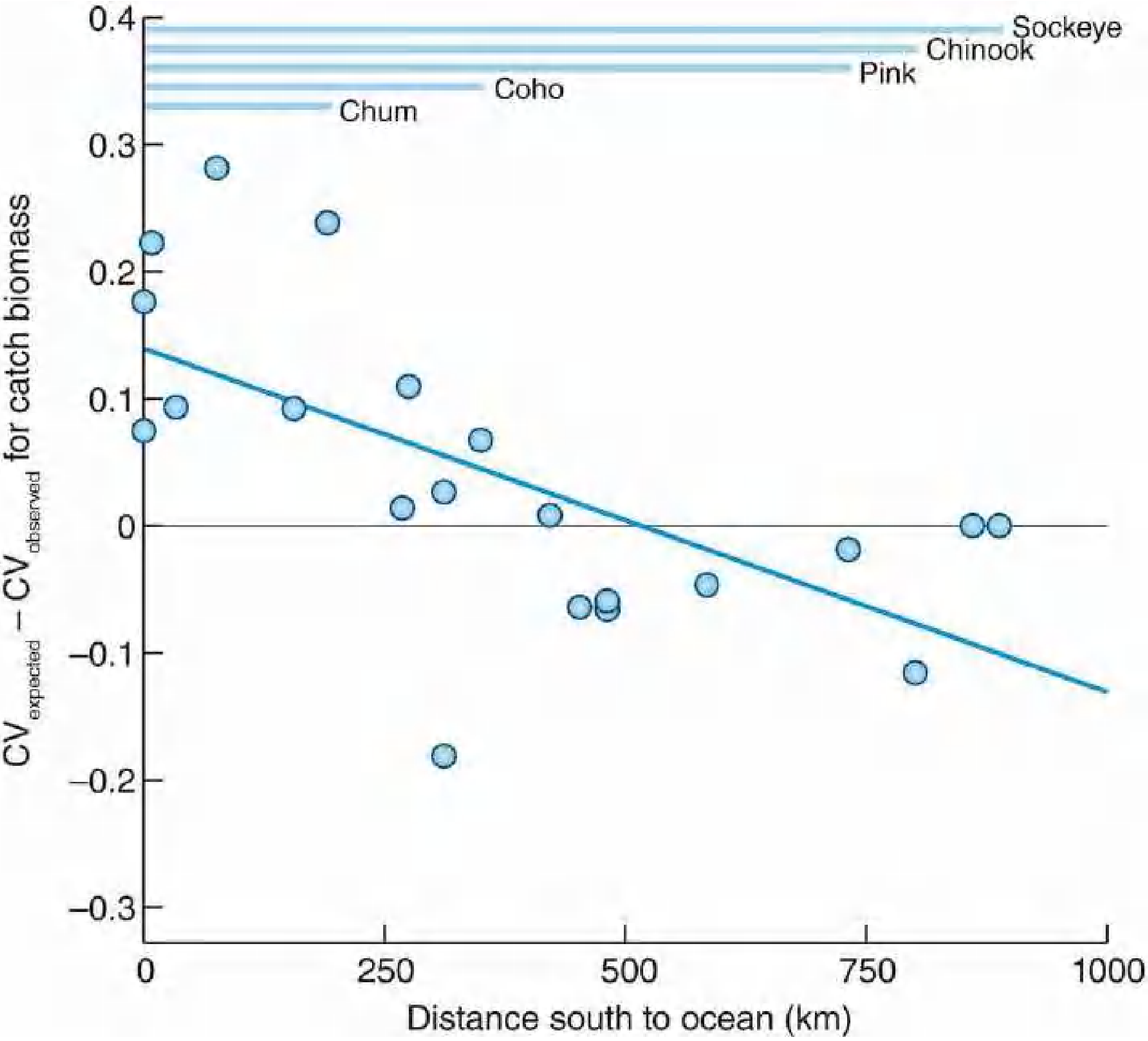


Figure 17.13 The stability of salmon catch biomass in Canada is greater when species richness is higher, closer to the sea. The portfolio effect on salmon biomass declines with distance from the sea (coefficient 0.15, SE = 0.04, r^2 = 0.39). The distributions of five species of salmon are shown above. Species richness was higher, closer to the sea.

Source: After Nesbitt & Moore (2016).

Although some of the portfolio effects were negative, the overall pattern was for stability of supply to increase with species richness. In British Columbia, 41% of indigenous people lack food security and 91% say they would prefer to eat more traditional food, such as salmon. This is a clear case where the conservation of biodiversity, apart from being good in its own right, provides a tangible and undeniable ‘ecosystem service’ (see Section 15.4.1).

17.2.5 Relating theory to data: community structure

As we noted above, in contrast to aggregate properties, there are fewer clear-cut predictions from theory regarding the effects of community complexity on the stability of community structure. Rather than seeking universal patterns when none exist, therefore, there has been an increasing trend towards asking the question: ‘Are there particular structures, or aspects of structure, that confer stability, and that we are therefore more likely to observe for that reason?’

robustness increases with connectance

Nonetheless, some of those fewer, clear-cut predictions have received strong support. For example, Dunne *et al.* (2002) investigated the association of community robustness with connectance and species richness, using an approach somewhere between pure theory and the analysis of field data. They took 16 published food webs (and thus structures that were real) and subjected them to the simulated sequential removal of species according to one of four criteria: (i) removing the most connected species first; (ii) randomly removing species; (iii) removing the most connected species first excluding basal species (those having predators but no prey); and (iv) removing the least connected species first. The stability of the webs was then judged by their robustness, extinctions occurring when species were left with no prey (and so basal species were subject to primary but not secondary extinction). Most fundamentally, community robustness increased with connectance (Figure 17.14). On the other hand, there was no association between robustness and species richness (though other studies have found such an association). It was also clear, however, that secondary extinctions followed most rapidly when the most connected species were removed, and least rapidly when the least connected species were removed, with random removals lying between the two. There were also some interesting exceptions when, for example, the removal of a least connected species led to a rapid cascade of secondary extinctions because it was a basal species with a single consumer, which was itself preyed upon by a wide variety of species. This alerts us immediately to the importance of *which* species are lost from a community, not simply how many, reminiscent of the keystone concept described above (see Section 17.2.2).

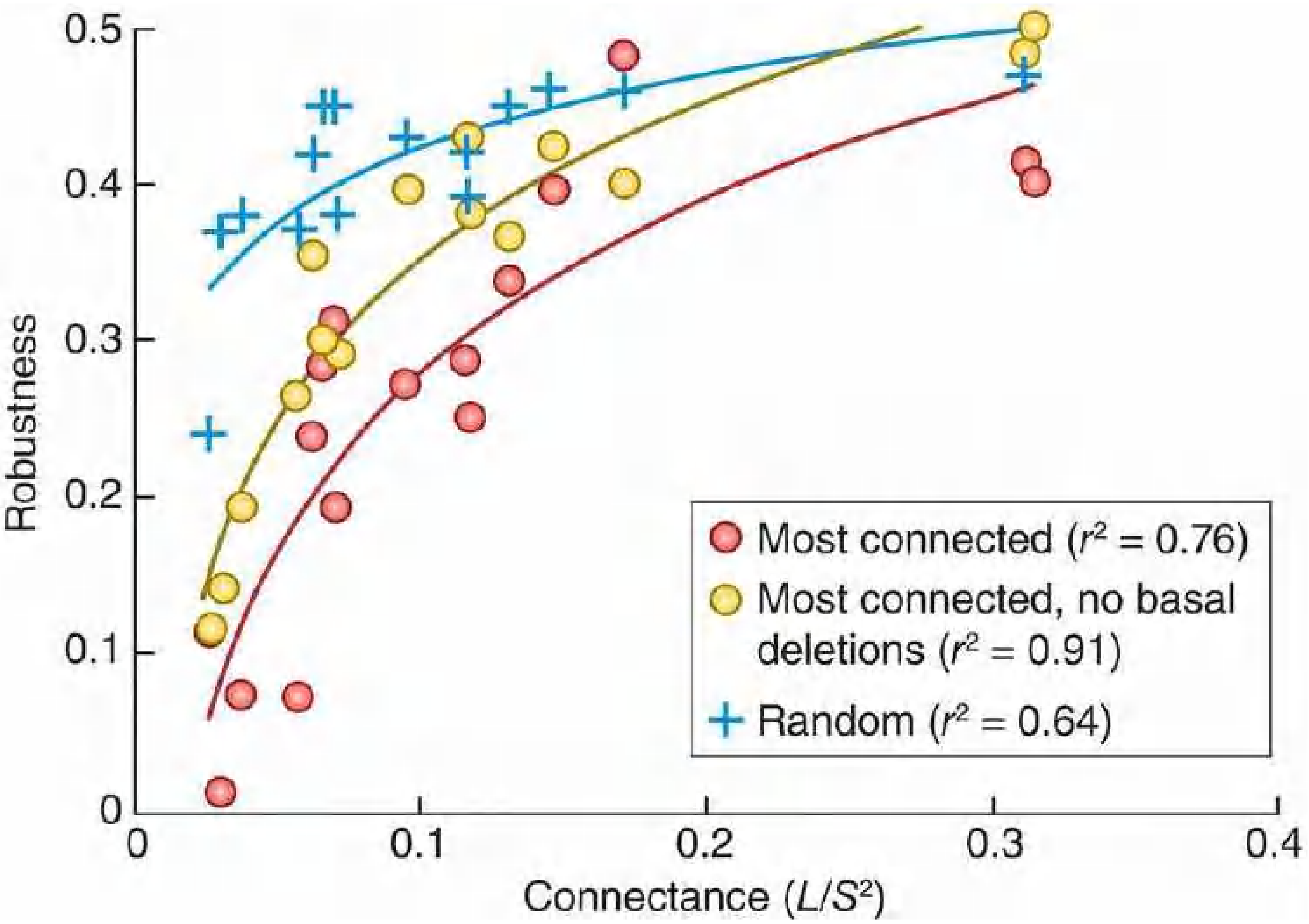


Figure 17.14 Robustness increases with connectance following simulated extinctions in natural communities. For 16 communities, robustness was measured as the proportion of species that needed to be removed for ≥50% of species to be lost ultimately; species were removed at random, the most connected were removed at each stage, or the most connected were removed but no basal species were removed, as indicated. There was no significant association with connectance if the least connected species were removed first (not shown). Connectance was measured as L/S^2, where L is the number of realised links and S the number of species – the fraction of all possible links that are realised.

Source: After Dunne *et al*. (2002).

inconsistent associations between population variability and richness

The prediction that population variability, and in this sense community instability, should increase as species richness increases has also received support, though just like the theoretical work in this area (see Figure 17.11), the evidence is equivocal. In Hector *et al*.'s (2010) study of European grasslands, for example, the positive correlation between the stability of community biomass and species richness was accompanied by a negative correlation between population biomass and richness (see Figure 17.12d). A similar pattern was apparent in Tilman's natural and experimental grasslands (Tilman, 1996 ; Tilman *et al*., 2006). On the other hand, when we turn from single to multiple trophic levels, the picture may be rather different. When experimental aquatic microcosms containing primary producers (single-celled algae), decomposers (bacteria), bacterivores (protozoa), omnivorous herbivore/bacterivores (protozoa and rotifers) and omnivorous top predators (further protozoa) were set up with either one, two or four species per trophic group, there was no discernible relationship between population variability and overall species richness at either of two productivity levels (Figure 17.15a). And when multitrophic microcosms were set up at a range of species richnesses by diluting the water from natural rock pool communities in Jamaica, population variability declined with species richness (Figure 17.15b). These data therefore add further support to the idea that there is no all-encompassing link between community complexity and the stability of community structure.

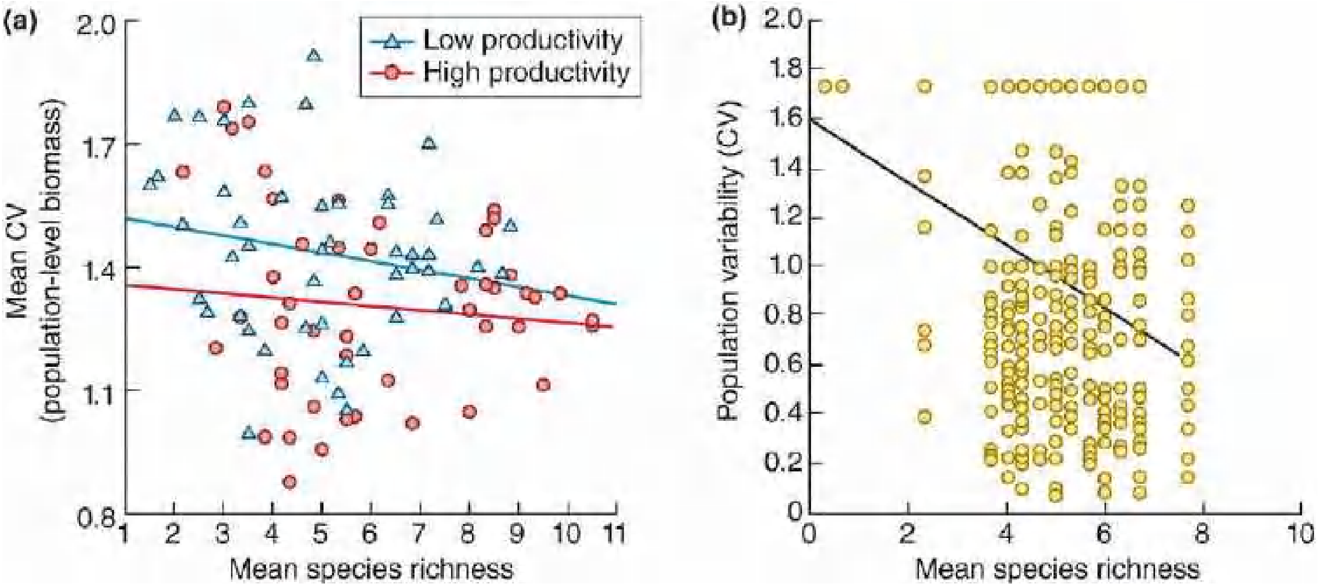


Figure 17.15 Population stability shows no or a positive correlation with species richness in multitrophic communities (a) No significant relationship between the mean coefficients of variation (CVs) of population biomass and species richness at the time of sampling for experimental aquatic microcosms at high and low productivity levels, as indicated ($P > 0.14$). (b) CVs of the abundances of individual populations in microcosms based on natural rock pool communities in Jamaica, plotted against the species richnesses of the microcosms at the time of sampling ($r^2 = 0.11$, $P < 0.01$). Variability declines and so stability increases.

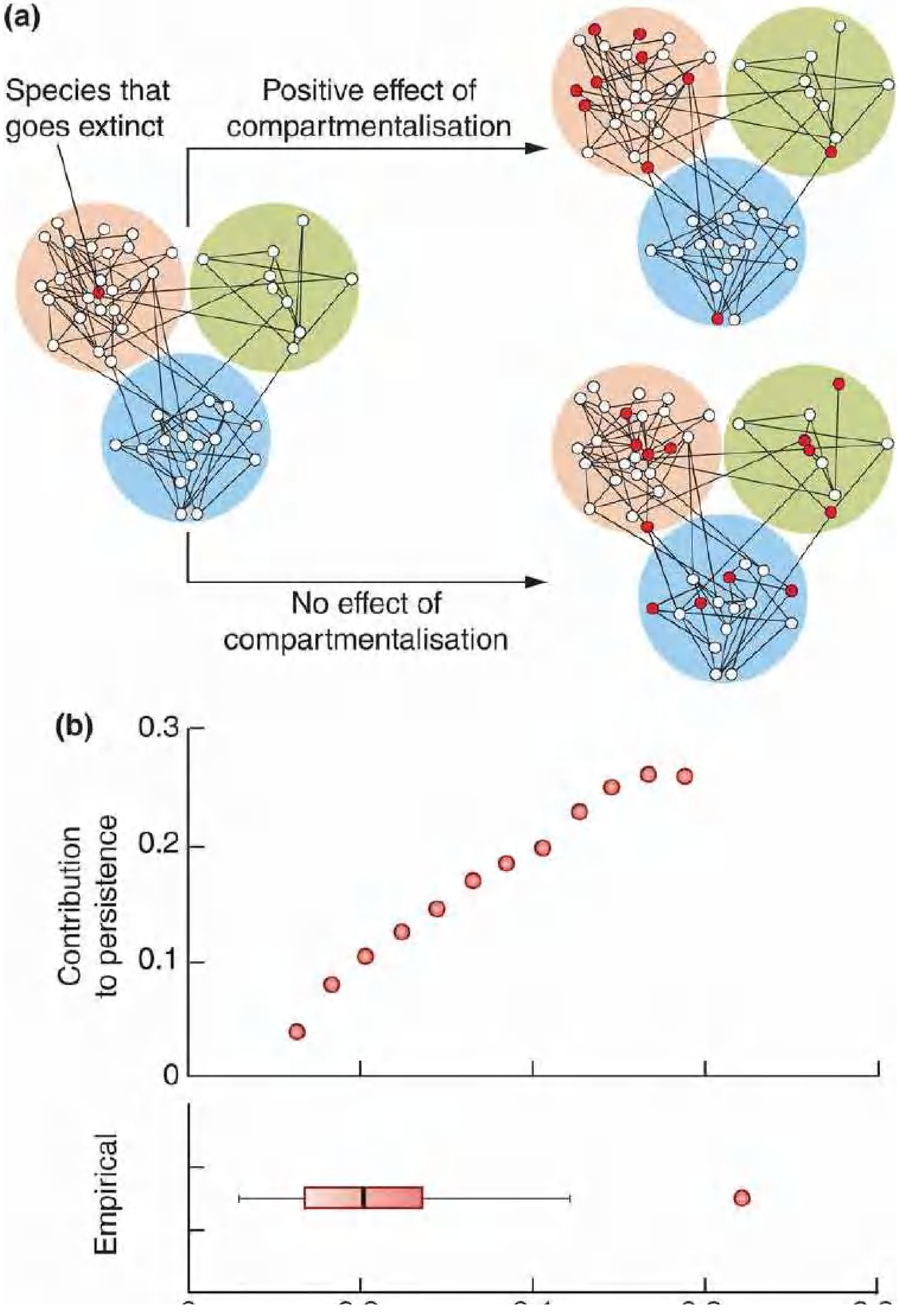
Source: (a) After Steiner *et al*. (2005). (b) After Vogt *et al*. (2006).

reconciling complexity with stability

An alternative approach, therefore, as noted above, is to accept that many natural communities are both complex and stable, and ask whether there are particular aspects of their structure that help explain this. Three aspects – the compartmentalisation of food webs, the organisation of trophic 'loops' and the length of food chains – are dealt with in the following three sections.

17.2.6 Compartmentalisation

A food web is compartmentalised if it is organised into subunits within which interactions are strong, but between which interactions are weak (Figure 17.16a). Some of the earliest analyses of food web stability suggested that compartmentalisation was a stabilising force (May, 1972), and subsequent studies have confirmed this. Indeed, a common theme among theoretical studies, generally, has been that food webs are stabilised by patterns of interaction strength that depart from the random patterns assumed in the earliest models (Neutel *et al*., 2002). Thus, for example, more compartmentalised webs are more robust to perturbation (the removal of one species), and indeed the contribution made by compartmentalisation to such stability is greater the greater that compartmentalisation is (Figure 17.16b). This occurs, simply, because the opportunities for knock-on effects are more limited in more compartmentalised webs: most subsequent extinctions occur in the same compartment as the original perturbation, rather than spreading evenly through the whole web (Figure 17.16c).


(a)
Species that goes extinct
Positive effect of compartmentalisation
No effect of compartmentalisation
(b)
0.3
0.2
0.1
0
Contribution to persistence
Empirical

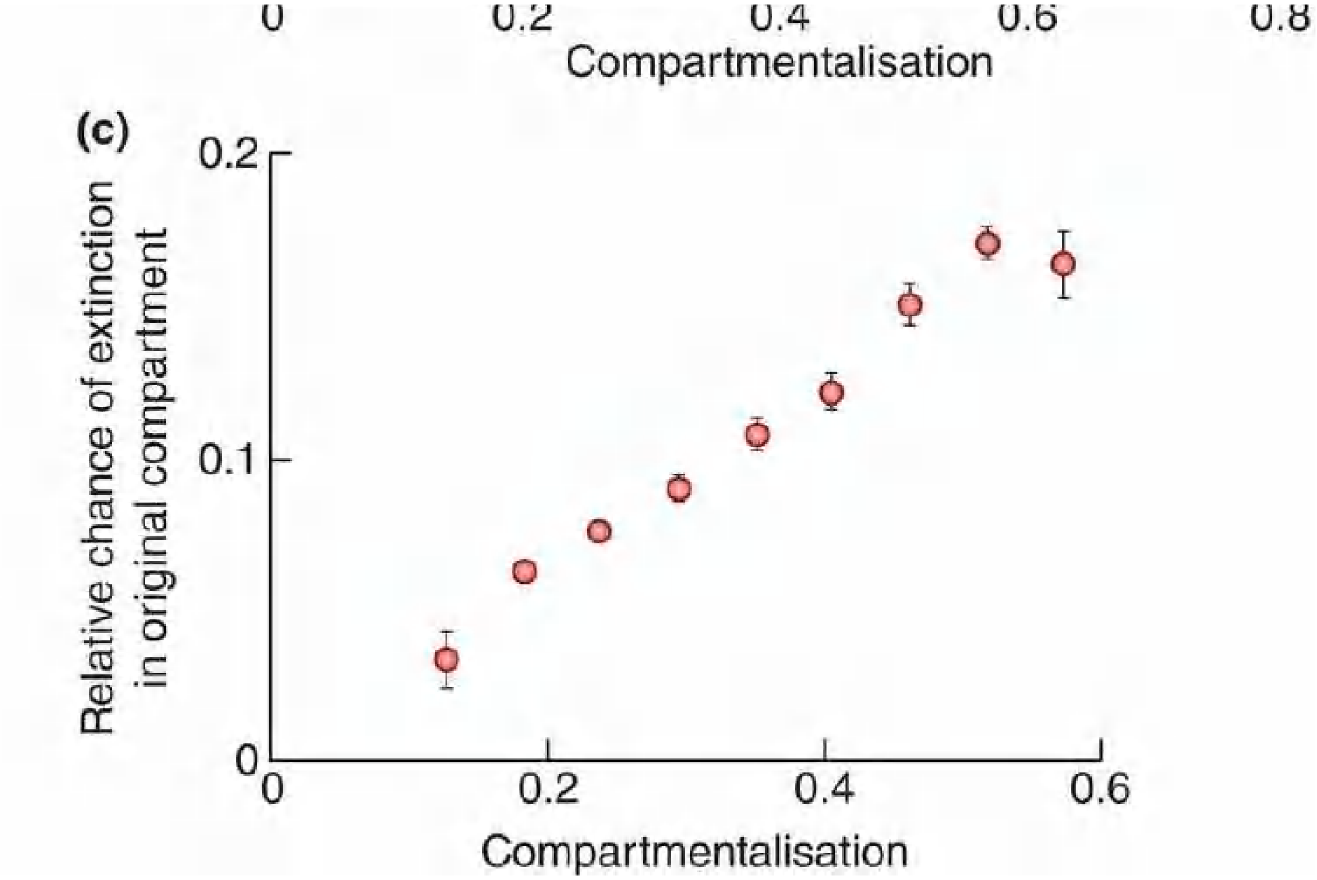


Figure 17.16 **Compartmentalisation stabilises food webs.** (a) Potential effects of compartmentalisation on the fate of a food web, following the initial extinction of one species (red dot). Subsequent extinctions are also in red. In the compartmentalised web, there are fewer extinctions more focused on one compartment. (b) Above: compartmentalised webs are more persistent (stable). Compartmentalisation in such webs is zero when species are connected at random and increases as more connections are confined within compartments. 'Persistence' describes the fraction of the original species remaining at the end of a model simulation following the removal of one species. The contribution to persistence is estimated from the proportion of persistence not accounted for by connectance and 'baseline' persistence. Below: box and whisker plots for estimated compartmentalisation from 15 actual (empirical) food webs. The mid-line is the median; the box covers the first-third quartiles; the whiskers span 1.5 times the interquartile range; the one dot is an outlier. (c) In compartmentalised webs, more extinctions are local. The effect of compartmentalisation on the relative chance of extinction in the same compartment as the original extinction is shown. Values greater than zero indicate that extinctions are more likely in the same compartment. Bars are SEs.

Source: After Stouffer & Bascompte (2011).

But do real food webs tend to be compartmentalised? And does any stability that such compartmentalisation confers account for the compartmentalisation? Real food webs certainly seem to be more compartmentalised than would be expected from making merely random connections between species (Figure 17.16b). It seems likely, however, that the compartmentalisation often emerges as a reflection of other aspects of community composition rather than being 'selected for' in its own right. One, probably typical example is an analysis of a Caribbean marine food web covering 3313 trophic interactions between 249 species, where five major compartments were identified. Members of different compartments tended to occupy different habitats (shore or off-shore; Figure 17.17a), they tended to feed on prey of different sizes (Figure 17.17b) and they tended to occupy different trophic levels (Figure 17.17c). Hence, some degree of compartmentalisation is inevitable, and it no doubt contributes to food web stability.

But there are no strong grounds for claiming that food webs tend to be compartmentalised *because* this confers stability.

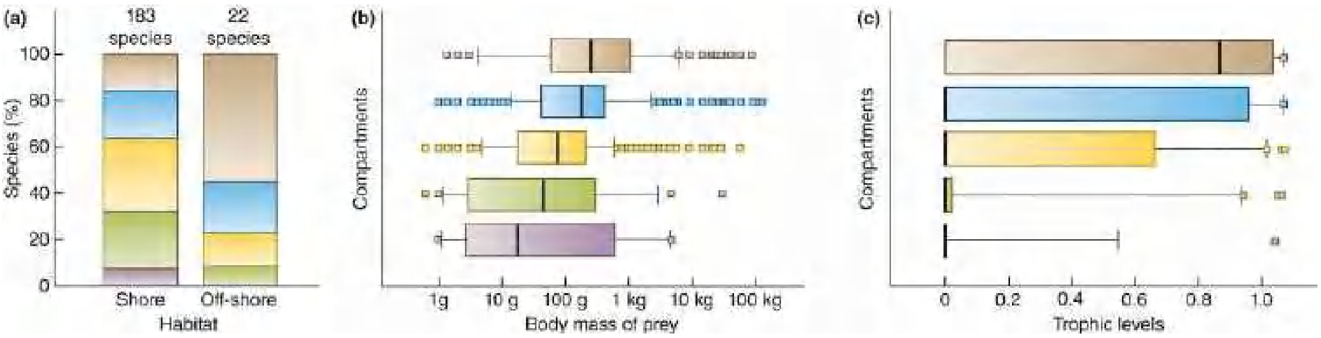


Figure 17.17 Separate compartments in a Caribbean marine food web arise because of similarities in preferred habitat and trophic relationships (a) The distribution of species allocated to different compartments of a Caribbean food web (distinguished by their colour) between two major habitat types. (b) Box and whisker plots of the prey consumed by species in the different compartments. The mid-line is the median; the box covers the first-third quartiles; the whiskers span the 10th and 90th percentiles; the small squares are outliers. (c) The ‘trophic level’ of fish species in the web allocated to different compartments, ranging from 0 (consume no fish) to 1 (top predators – consumed by no fish) with species that are both predators and prey of fish ranged in between.

Source: After Rezende *et al.* (2009).

17.2.7 Organisation of trophic loops

In a more general investigation of non-random patterns of interaction strengths in real food webs, Neutel *et al.* (2002) looked at seven documented webs of soil communities. Here we focus especially on one of these, as they did (the CPER or Central Plains Experimental Range web), though the results from the others were similar. At the heart of their analysis was an examination of trophic ‘loops’. These are pathways through a web, beginning and ending with the same focal species and visiting other species no more than once. The simplest, therefore, is a loop of length two that links a consumer and its resource: from the consumer to the resource (a negative, predatory interaction) and back from the resource to the consumer (a positive interaction). However, longer loops are possible, for example where there is omnivory, as shown in Figure 17.18a. In this case there are two loops of length three: one clockwise, one anticlockwise. The loops in Figure 17.18a also reflect the respective strengths of the different interactions, and there are two notable points about these. First, the *per capita* effects of predators on their prey (which is what the interaction strengths measure) are typically much greater (often by two orders of magnitude) than those of prey on their predators, simply because a single predator may typically kill many prey, whereas a single prey item is typically only a small proportion of the lifetime consumption of a predator. Second, biomass typically decreases as we move up the trophic levels. Hence the omnivore in Figure 17.18a focuses its consumption more on the more abundant lower trophic level than on the intermediate one, and the interaction strengths reflect that. Each loop can then be characterised by its ‘loop weight’: the geometric mean of the absolute values of its interaction strengths. The geometric mean of a set of items is the reciprocal of the mean of the reciprocals of their individual values and is more influenced by low than by high values. The weights of the loops in Figure 17.18a are therefore both relatively small – one because it contains two weak bottom-up links, and the other because it contains only one strong top-down link.

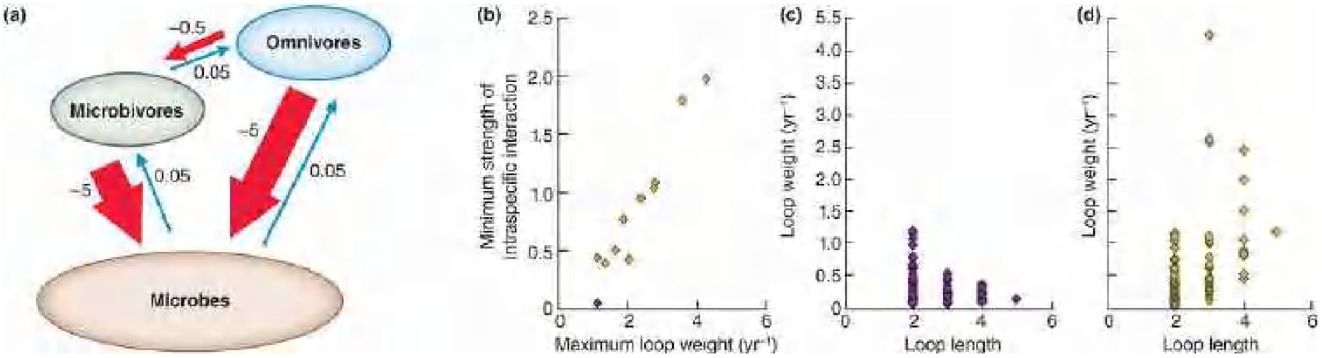


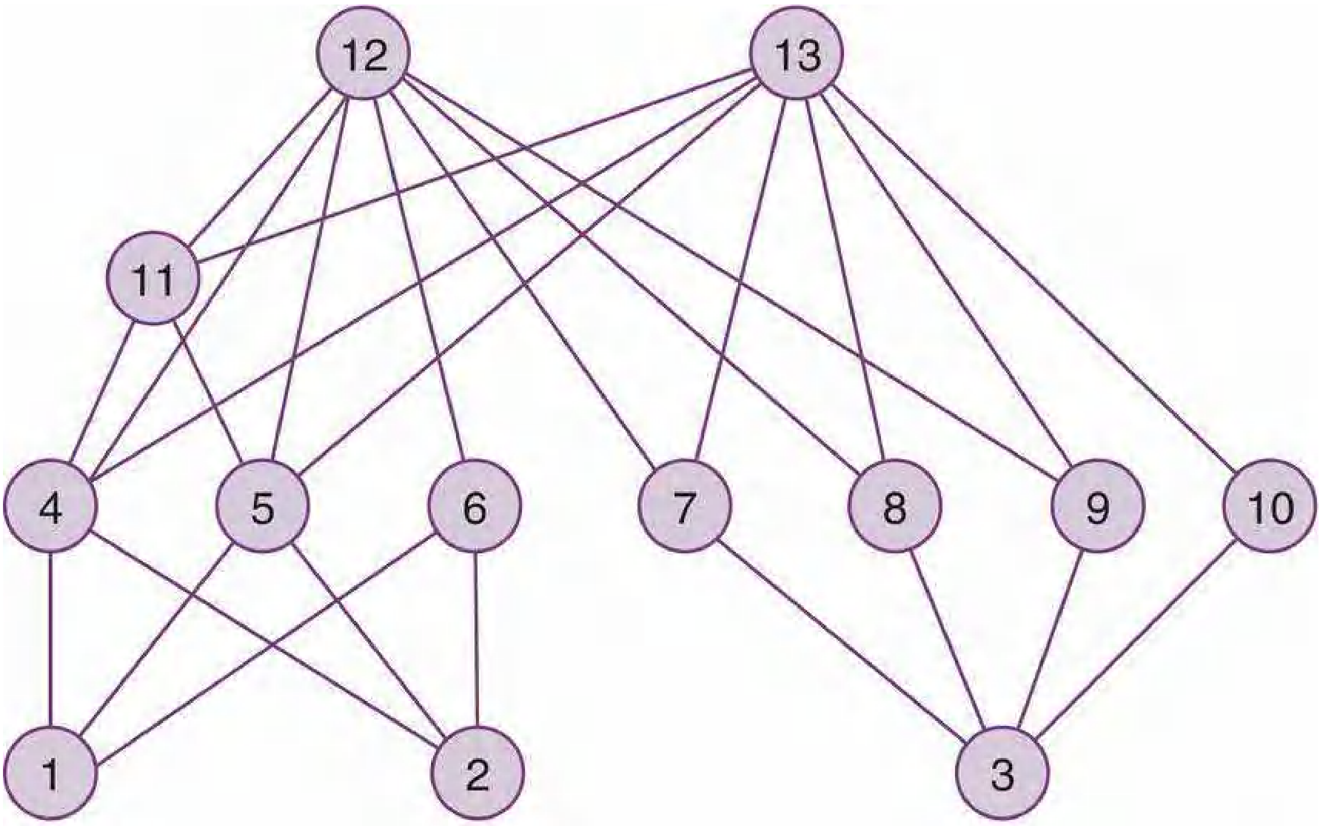
Figure 17.18 **Weak links in long trophic loops stabilise food webs.** (a) Loops and the calculation of their weights. The three (hypothetical) trophic compartments shown are linked by trophic interactions (positive from prey to predator (blue arrows); negative from predator to prey (red arrows)) the strengths of which are indicated. There are two loops of length three – from a compartment back to itself, clockwise or anticlockwise. The weight of each loop is the geometric mean of the absolute values of the three interaction strengths. (b) The relationship between the stability of the CPER food web and the maximum loop weight (purple diamond), and the same relationship for 10 typical randomisations of the web (green diamonds). Stability was measured as level of intraspecific interaction required to stabilise the web. Stability therefore *de*creased with maximum loop weight. (c) In the real food web, long loops tended to have a low weight. (d) In a typical randomisation of the food web, long loops tended to have a high weight.

Source: After Neutel *et al.* (2002).

Crucially, Neutel and her co-workers found that the stability of a web reflected its maximum loop weight (that is, the weight of its heaviest loop). To understand this, note first that web stability overall reflects a balance between the dampening, negative-feedback, stabilising effects of intraspecific interactions (see Chapter 5) and the destabilising, positive-feedback effects of interspecific, especially predator–prey, interactions (see Chapter 10). For randomly constructed versions of the CPER web (Figure 17.18b), the maximum loop weight was directly related to level of intraspecific interaction required to stabilise the web – that is, webs with smaller maximum loop weights required only relatively little intraspecific interaction for stability and hence were more stable. But the loop weights of longer loops in the real CPER web were very much lower than in its randomised counterparts (Figure 17.18c, d). Hence, the maximum loop length of the real web was itself much lower than in the random webs and the stability much higher (Figure 17.18b), with this stability arising from the real web containing a much higher proportion of long loops of low weight than expected by chance. Combined with similar results for the other six webs, this suggests that the stability of real food webs is reconciled with their complexity, at least in part, by their being organised into structures containing long loops with relatively many weak links – itself a reflection of patterns of omnivory and the distribution of biomass across the trophic levels.

17.2.8 Food chain length: the number of trophic levels

One aspect of community structure that has received particular attention has been food chain length. The term has sometimes been used to describe the number of species in the chain, and sometimes (as we do here) the number of links (Post, 2002). If we wish to calculate the food chain length for the community in Figure 17.19, for example, we can start with basal species 1 and trace four possible trophic pathways via species 4 to a top predator: 1–4–11–12, 1–4–11–13, 1–4–12 and 1–4–13, giving us four food chain lengths: 3, 3, 2 and 2. Including the 21 further chains, starting from basal species 1, 2 and 3, then gives us an average of all the possible food chain lengths of 2.32, and adding one to this gives us the number of trophic levels that can be assigned to the food web. Almost all communities described have consisted of between two and five trophic levels, and most of these have had three or four. What sets the limit on food chain length? And how can we account for variations in length?



Base-to-top food chains

1 4 11 12
1 4 11 13
1 4 12
1 4 13
1 5 11 12
1 5 11 13
1 5 12
1 5 13
1 6 12

2 4 11 12
2 4 11 13
2 4 12
2 4 13
2 5 11 12
2 5 11 13
2 5 12
2 5 13
2 6 12

3 7 12
3 7 13
3 8 12
3 8 13
3 9 12
3 9 13
3 10 13

Figure 17.19 **The calculation of food chain length.** Food web of an exposed intertidal rocky shore in Washington State, USA. The pathways of all possible base-to-top food chains are listed. The length of a chain (number of links) is one fewer than the number of taxa. The food chain length assigned to the web is then simply the average of these. The taxa are as follows: 1, detritus; 2, plankton; 3, benthic algae; 4, acorn barnacles; 5, *Mytilus edulis*; 6, *Pollicipes*; 7, chitons; 8, limpets; 9, *Tegula*; 10, *Littorina*; 11, *Thais*; 12, *Pisaster*; 13, *Leptasterias*.

Source: After Briand (1983).

productivity and chain length

The earliest attempts to answer these questions focused on an energetics argument, going back at least to Elton (1927). Of the radiant energy that reaches the earth, only a small fraction is fixed by photosynthesis and made available to either herbivores or detritivores. Thereafter, each feeding link amongst heterotrophs is similarly limited in its efficiency: at most 50%, sometimes as little as 1%, and typically around 10% of energy consumed at one trophic level is available as food to the next (see Chapter 20). A typical chain with just three or four trophic levels could arise, therefore, simply because a viable population of predators at a further trophic level could not be supported by the available energy, and similarly, variations in chain length could simply reflect variations in available energy – an argument we met earlier as part of Oksanen *et al.*'s (1981) Ecosystem Exploitation Hypothesis.

productive space and ecosystem size

On the other hand, this simple argument leaves unanswered the question of whether it is the amount of energy available per unit area or the total amount of available energy. Thus, Schoener (1989) proposed the 'productive space' hypothesis: that food chain length is determined by the productivity per unit area multiplied by the space (or volume) occupied by the ecosystem. A very small and isolated habitat, for example, no matter how productive locally, is unlikely to provide enough energy for viable populations at higher trophic levels. However, the suggestion that food chains will therefore be longer in larger ecosystems alerts us to the suggestion that this might also be true for other reasons. For example, larger ecosystems may provide greater opportunities for trophic diversity, adding links to a chain, or may be more stable because disturbances are confined to only a small part of the whole (compartmentalisation), allowing subsequent recovery.

dynamic stability

Indeed, a stability argument has been proposed more generally as a limitation on food chain length. Pimm and Lawton (1977), in particular, examined variously structured four-species Lotka–Volterra models and found that webs with more trophic levels had return times after a perturbation that were substantially longer than those with fewer levels – they were less resilient. This suggests that longer chains are therefore only likely to persist in environments with fewer disturbances. In fact, though, reflecting a pattern that we see connecting ecosystem models and stability more generally (see Section 17.2.3), it is possible to construct other models in which there is no such link between food chain length and stability, or even an *in*crease in stability in longer food chains (Sterner *et al.*, 1997). Predictions, therefore, are varied, but this does at least pose the question of whether there is some association, overall, between food chain length and levels of disturbance.

constraints on design and optimal foraging

There may also be evolutionary constraints on the anatomy or behaviour of predators that limit the lengths of food chains. To feed on prey at a given trophic level, a predator has to capture, subdue and consume it. In general, predators are larger than their prey (not true, though, of grazing insects and parasites – see later), and body size tends to increase at successive trophic levels (Cohen *et al.*, 2003). It may be impossible, therefore, to imagine a predator that is, for example, both fast enough to catch an eagle and big enough to kill it.

Also, consider the optimal diet of a carnivore. Would it do best to feed on herbivores or other carnivores? The herbivores are more abundant and less well protected. The advantages to feeding lower down in the food chain are therefore obvious. However, if all species did this, competition would intensify, and feeding higher in the food chain would become more advantageous. Nonetheless, analyses of models comparing foragers that were ‘adaptive’ (switching their preferences to prey species as they became more abundant) with ‘non-adaptive’ foragers (with fixed preferences, unable to switch) found that adaptive foraging did usually shorten chain length (Figure 17.20). What’s more, whereas chain length increased with productivity in the models for non-adaptive foragers (as predicted more generally), with adaptive foragers, chain length actually decreased slightly because increased productivity enhanced the attractiveness of prey at lower levels and encouraged consumers to switch to them (Figure 17.20).

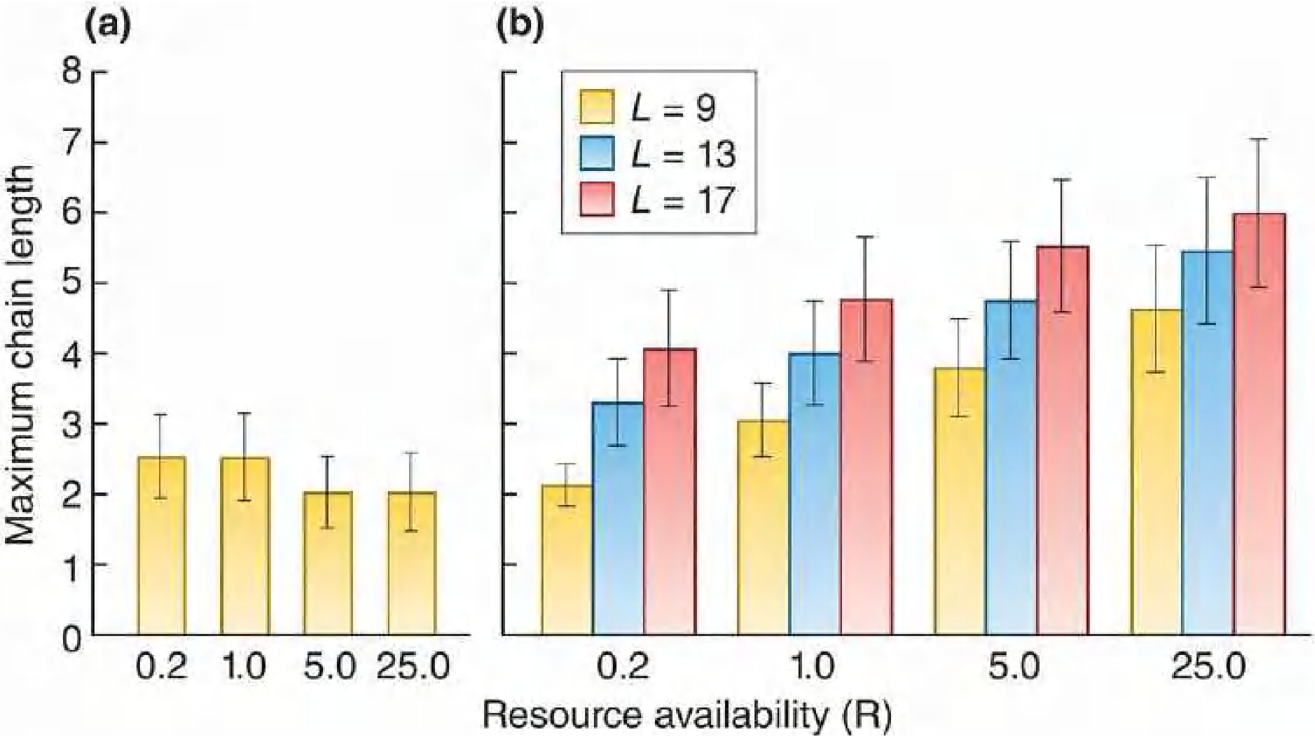


Figure 17.20 Adaptive foraging shortens food chain length in a simulation model, but food chain length increases with productivity only with non-adaptive foraging. (a) Adaptive foraging. Maximum food chain length in a simulation model, run 1000 times, each for 10 000 time steps, in which a food web comprising 10 species was built up by giving potential foragers the opportunity, at each step, to link to prey if it was energy-efficient for them to do so. This was done at each of four levels of resource availability, as indicated. The number of trophic links, *L*, established by the foragers in the 1000 simulations ranged between nine and 17. Bars are SDs. (b) Non-adaptive foraging. Maximum food chain lengths in related simulation models in which foragers were unable to switch their preferences. To match the adaptive model, this was done for *L* = 9, = 13 and = 17.

Source: After Kondoh & Ninomiya (2009).

the evidence: sometimes supportive of predictions, sometimes not

Taken overall, therefore, we have several predictions about food chain length, but each bears caveats. Food chain length is predicted to be greater in more productive environments, though whether this should be productivity per unit area or total productivity is uncertain, and adaptive foraging may, in any case, hide or even reverse this effect. Food chain length should be greater in larger ecosystems, though finding such an association may still leave the underlying reason uncertain. And food chain length may be greater in less disturbed environments, though by no means all analyses of model ecosystems predict this relationship. Early analyses of natural systems tended to combine studies from a wide range of habitats, but found no evidence of increased chain length in more productive systems, nor of an effect of disturbance (Briand & Cohen, 1987 ; Schoenly *et al*., 1991). On the other hand, studies where resource availability was manipulated experimentally, especially in very simple microcosms, and especially where the decreases take productivity below around 10 g carbon m^{-2} $year^{-1}$, have more commonly shown food chain length to decrease with decreased productivity (Post, 2002). For example, in an experiment using water-filled containers as analogues of natural tree-holes, a 10- or 100-fold reduction from a 'natural' level of energy input (leaf litter) reduced maximal food chain length by one link, because in this simple community of mosquitoes, midges, beetles and mites, the principal predator – a chironomid midge *Anatopynia pennipes* – was usually absent from the less productive habitats (Figure 17.21). This, though, is arguably a case of simple experiments being closer to simple models than to natural systems, where, within a much larger food web, there is a far greater chance that effects of productivity would be compensated for by other changes in structure rather than food chain length.

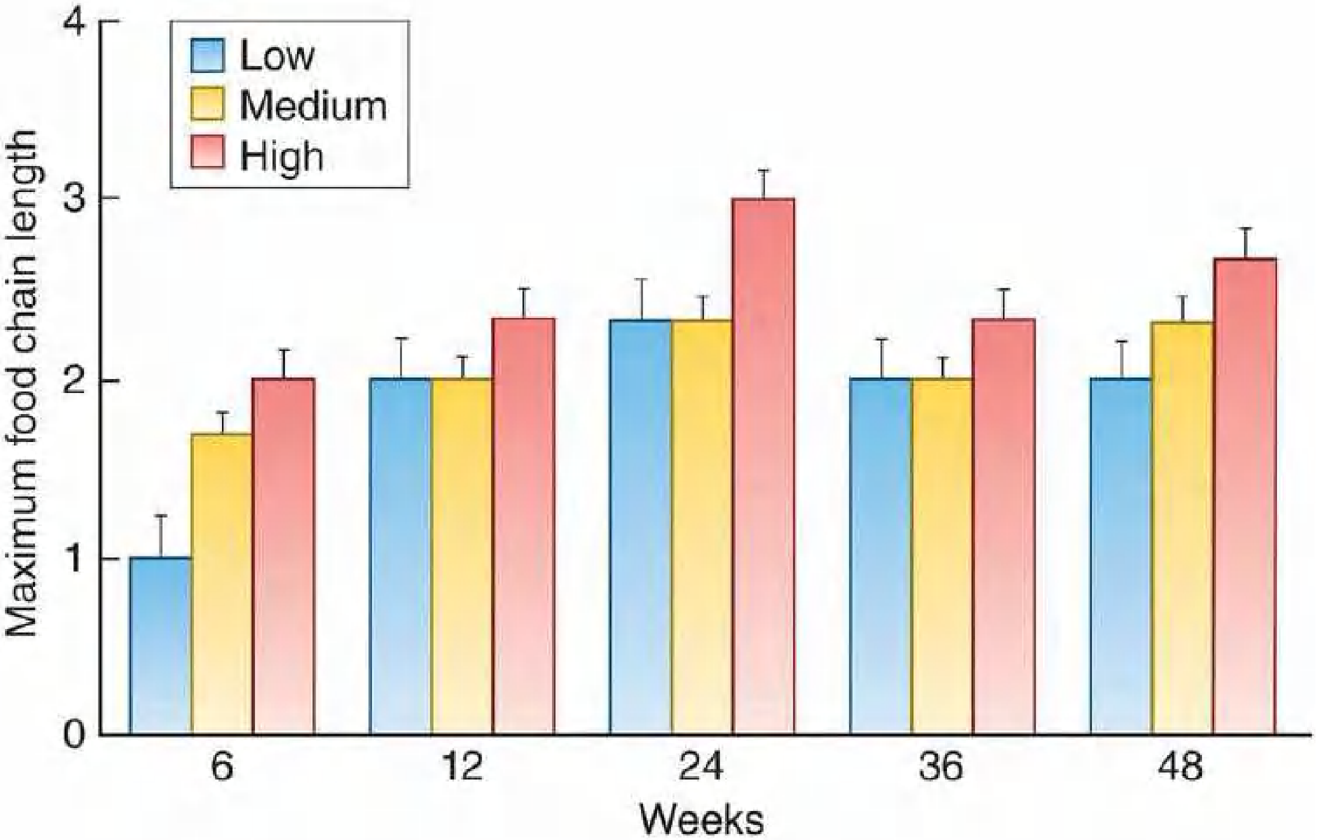
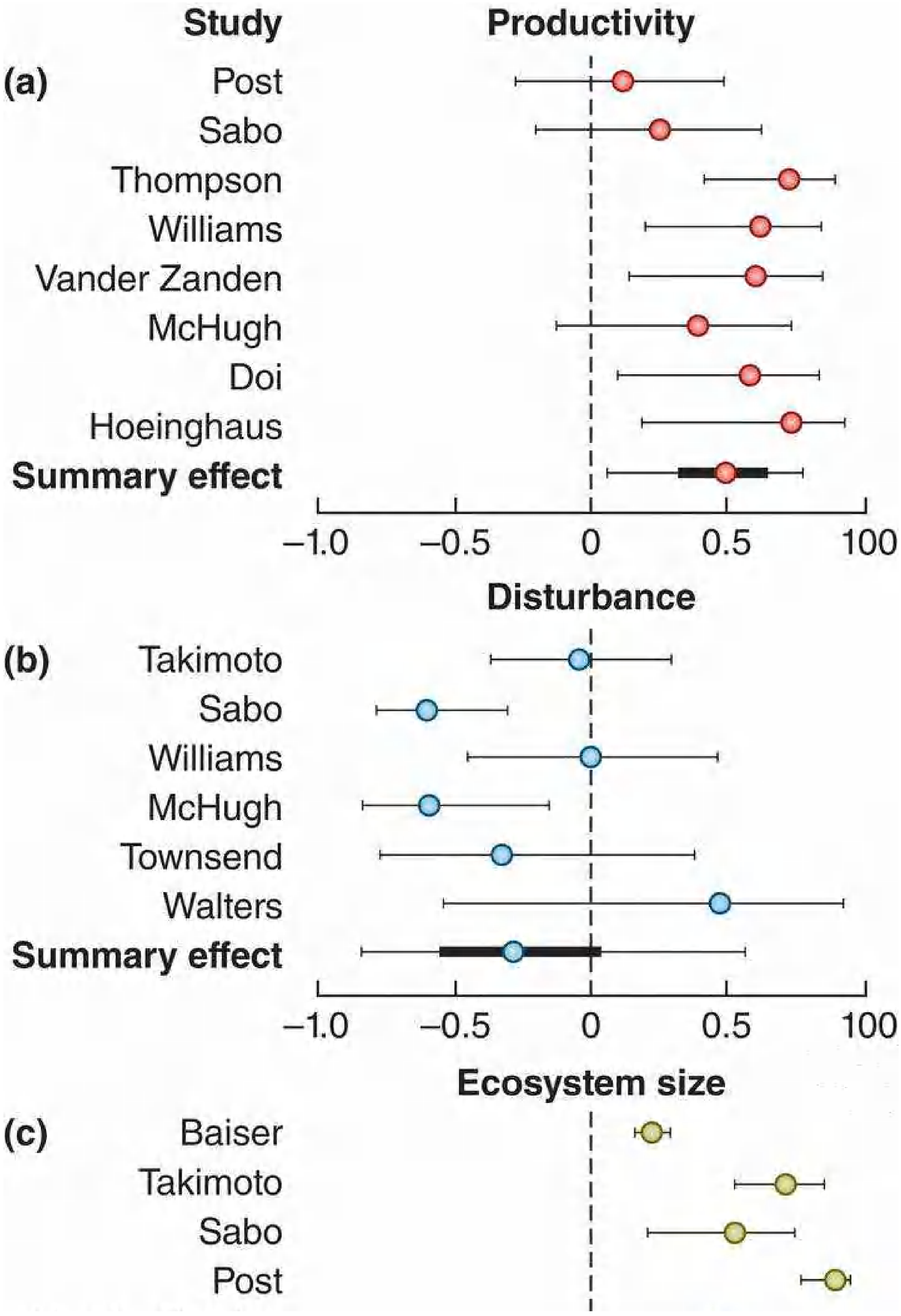


Figure 17.21 Food chains are longer when productivity is higher in simple artificial tree-hole communities. The artificial tree-holes were water-filled containers, also containing crushed leaf litter, that were available, over a 48-week period, for colonisation by animals living in real tree-holes nearby in New South Wales. Australia. The maximum food chain lengths were noted (means and SEs shown). This was carried out at three levels of productivity: low (an initial 0.6 g of litter per container only), medium (an initial 0.6 g of litter, plus 0.06 g every 6 weeks) and high (an initial 6 g of litter, plus 0.6 g every 6 weeks – close to natural levels).

Source: After Jenkins *et al.* (1992).

In contrast to the earlier analyses, however, a more recent meta-analysis, focusing on studies where more similar, natural systems were compared (only ponds, only communities in water-filled pitcher plants, etc.) produced more supportive results (Figure 17.22). There was overall support, in particular, for increases in chain length with both productivity and ecosystem size. Notably, though, in three of the studies there was a significant ecosystem size effect and also a significant productivity effect, but there were none in which there was a productivity effect but no size effect. This is more consistent with the productive space hypothesis than with an effect of productivity *per se*. Moreover, in a further three studies there was a significant ecosystem size effect but no productivity effect, supporting the idea that ecosystem size effects may arise for a variety of reasons.


Study
Productivity
(a)
Post
Sabo
Thompson
Williams
Vander Zanden
McHugh
Doi
Hoeinghaus
Summary effect
–1.0
–0.5
0
0.5
100
Disturbance
(b)
Takimoto
Sabo
Williams
McHugh
Townsend
Walters
Summary effect
–1.0
–0.5
0
0.5
100
Ecosystem size
(c)
Baiser
Takimoto
Sabo
Post

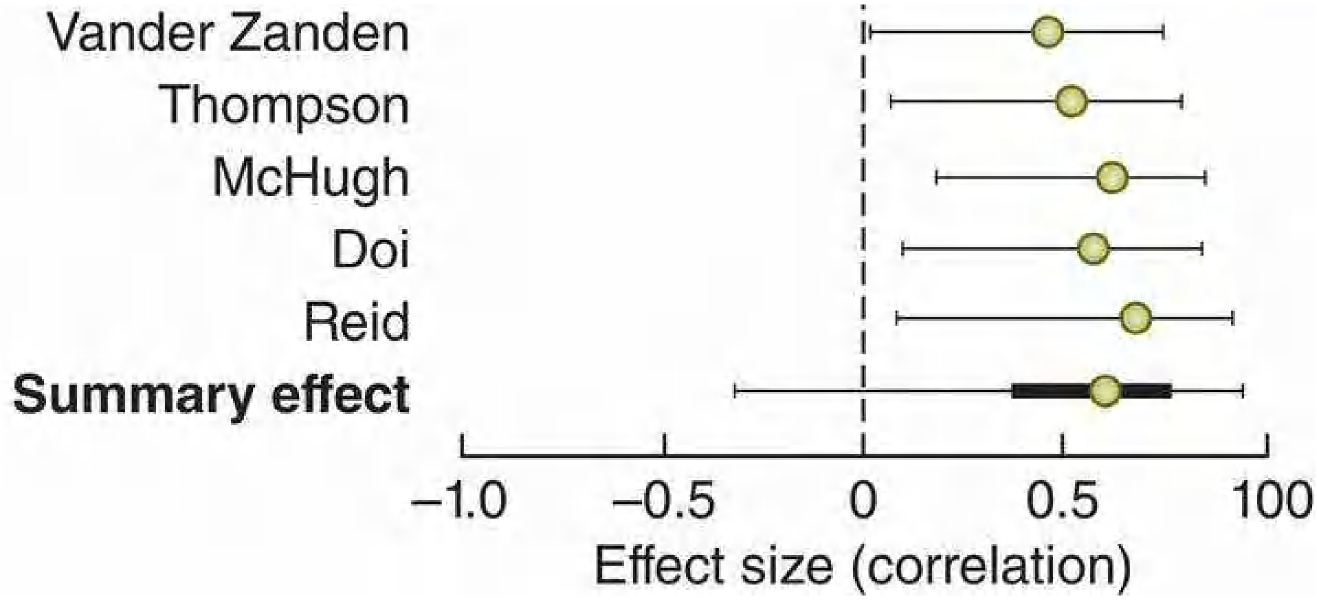


Figure 17.22 A meta-analysis supports the importance of productive space and ecosystem size, but not disturbance, in determining food chain length. Individual studies are indicated by the names of their first authors to the left. Each is a comparison of similar, natural communities allowing food chain length to be correlated with (a) levels of productivity, (b) levels of disturbance and (c) ecosystem size. For individual studies, the mean effect size and 95% CI is shown. For the summaries of all studies, the overall mean is shown with the 95% CI (thick bar) and the 95% prediction interval – the range within which future tests would lie with a 95% probability (longer bar).

Source: After Takimoto & Post (2013).

On the other hand, while some studies found that chains were shorter in more disturbed ecosystems, by no means all did, and there was no significant effect overall. This is perhaps another case where effects are most likely to be apparent in simple systems similar to those envisaged by the models that predicted them. The experiment in Figure 17.21, for example, is one case where a disturbance did shorten the food chain as the apex predator was lost. By contrast, in a study of terrestrial food webs on islands in the Bahamas that we have discussed previously (see Figures 16.17 and 17.2), there was a clear effect of ecosystem size (Figure 17.23), but disturbance (where a community was exposed to rather than protected from hurricanes and storms) tended simply to lead to a lizard, *Anolis sagrei*, being replaced as an apex predator by an orb spider, *Argiope argentata*, but not to a reduction in food chain length. For studies of food chain length overall, therefore, the two longest established hypotheses – energy per unit area and dynamic fragility – have, if anything, the least support.

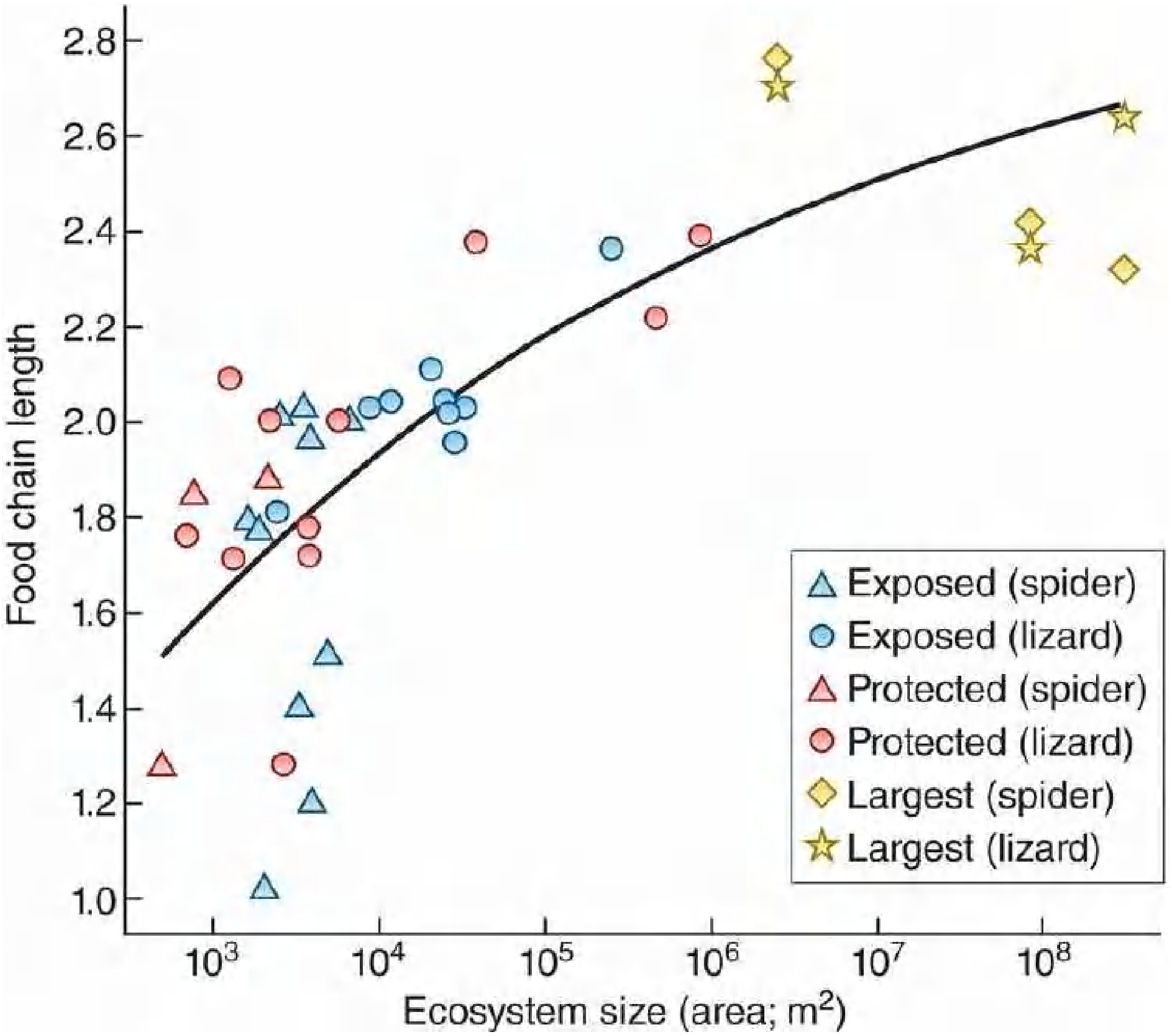


Figure 17.23 Food chain length in the Bahamas increases with ecosystem size but is not affected by disturbance. The best-fitting statistical model to account for variations in food chain length (the curve) includes ecosystem size (food chain length = 2.99 [1 − exp(−0.26 log area)]; R^2 = 0.55) but not disturbance, as indicated by the distribution of data relative to the curve from islands exposed to storms (high disturbance), protected islands (low disturbance) and the largest islands, which were also protected from storms, whether the apex predator was a lizard or a spider.

Source: After Takimoto *et al.* (2008).

17.2.9 Parasites in food webs

Historically, studies of food webs have tended to ignore parasites, such that when a food chain is described as having four trophic levels, these would typically be a plant, a herbivore, a predator that eats the herbivore, and a top predator that eats the intermediate predator, without any acknowledgement that the top predator is itself almost certainly attacked by parasites, which are themselves attacked by pathogens, and so on. More recently there have been an increasing number of studies seeking to rectify the neglect.

Including parasites in a food web modifies the web in a number of ways. Most obviously, as we have already noted, it tends to increase the maximum length of food chains. But in addition, parasites, even when added at the top of existing food chains, may themselves be prey, either consumed in their own right (for example, as part of a cleaner–customer mutualism – see Section 13.3.1) or when their own hosts are being preyed upon, referred to as *concomitant predation* (Johnson *et al.*, 2010). It has been estimated, for example, that in one estuarine food web, 44% of the trophic links involve the predation of parasites (Lafferty *et al.*, 2006).

parasites: often ignored, but are they important?

Ignoring parasites is clearly wrong, in the same way as it would be wrong to ignore any other substantial component of the communities we study. A crucial question, though, especially when so many studies *have* failed to include parasites, is whether their inclusion would affect food webs in broadly the same way as the inclusion of a similar number of species with trophic relationships similar to those in the existing web, or whether there are changes specifically attributable to the added species being parasites. Dunne *et al.* (2013) addressed this question using data from seven published food webs for which they could compare: (i) the web including only free-living species; (ii) the same web with parasites included but no concomitant links; and (iii) the web with parasites and concomitant links. As expected, including parasites, with or without concomitant links, made the food webs more diverse and more complex – there were more species, more trophic links, and so on. Crucially, though, these changes were simply 'scale-dependent', that is, precisely what would be expected from the inclusion of additional species. There were some structural changes for which the parasites, as parasites, could be held responsible – for example, a higher proportion of community modules characteristic of parasitism – but overall, it seems not that we need to rethink the fundamentals of food web analysis, but rather that we must ensure that we have methods that will handle the more diverse and more complex food webs that the inclusion of parasites will inevitably generate.

One relatively well-established consequence of the inclusion of parasites in food webs is that the robustness of the web is decreased – that is, the tendency for secondary extinctions to follow primary extinctions (Figure 17.24). As noted earlier (see Section 17.2.5), robustness tends generally to increase with both diversity (species richness) and complexity (connectance), and hence the question arises again: 'Is the decreased robustness a consequence simply of including additional, connected species, or does it reflect characteristics of parasites in particular?' In this case, where species removals were simulated from five published webs that included parasitic links, it seems that the parasite characteristics themselves contribute. Starting with versions of the food webs in which the parasites were left out, robustness was decreased slightly by the inclusion of parasites (from 48% to 47% on average) but was actually decreased more (from 48% to 45%) by the inclusion instead of top predators at random in numbers equivalent to the parasites (Figure 17.24). However, robustness was decreased much further (down to 41%) by acknowledging that many parasites require multiple species of host for the completion of their life cycle, so that the loss of any one of them would lead to their own extinction (Figure 17.24). On the other hand, if adding new species to a food web decreases robustness largely by leading to the loss of the species that have just been added, this reinforces the view that while rectifying the neglect of parasites in food web studies is important, it may not lead to a revolution in our understanding of food web dynamics.

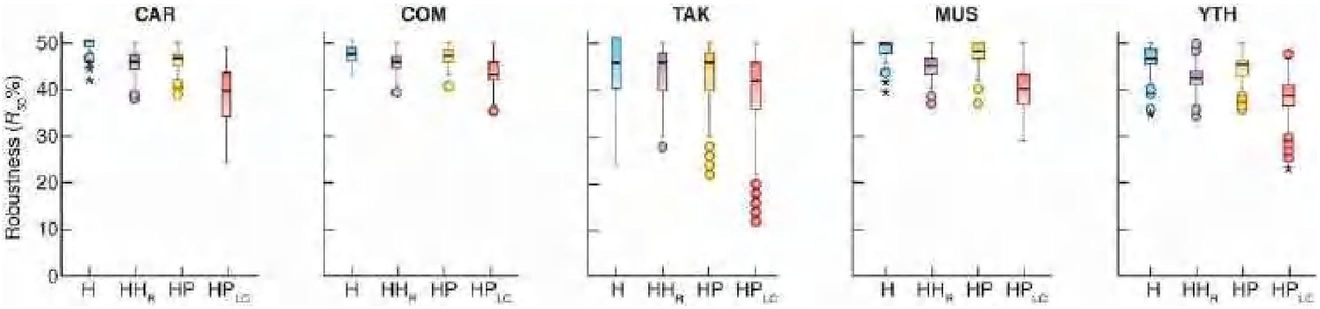


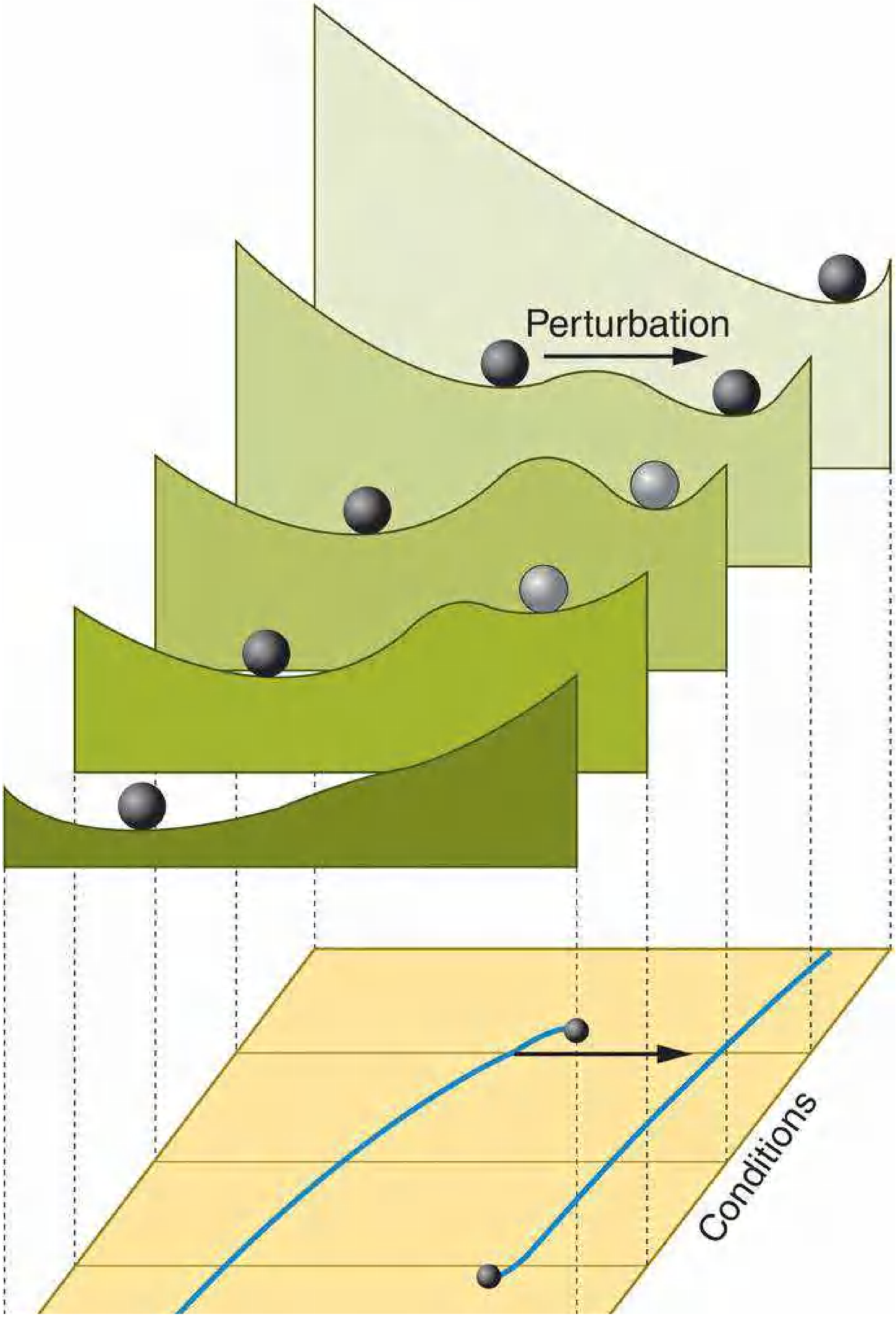
Figure 17.24 **Parasites decrease the robustness of food webs (slightly).** Food web robustness (the percentage of species that need to be removed for 50% of species to be lost ultimately) is shown for five food webs, as indicated (for details of the food webs see Chen *et al.* (2011)), showing medians, quartiles, outliers (circles) and extreme values (stars), based on the results of 1000 simulated sets of removals in each case. H signifies the web with only predator–prey (not parasite–host) links included. HH_R added top predators to this, at random, in numbers equivalent to the parasites that had been excluded. HP included parasite–host links as per the original web, but did not reflect the reliance of many parasites on a multiplicity of hosts for their continued existence. HP_{LC} included this.

Source: After Chen *et al.* (2011).

17.3 Regime shifts

tipping points

We close this chapter by returning to a topic touched on earlier in the book (see Section 14.7) but that is especially relevant to community stability. Up to now, we have been assuming implicitly that small disturbances will be relatively easy for a community to resist and will generally have little effect, whereas large disturbances are more likely to be irresistible and to lead to profound changes. It is now widely acknowledged, however, that in some cases, in terms of stability, communities may lie close to a *tipping point*, such that no more than a small disturbance is required to transform the community from one state into a very different one. The community would be said to have experienced a *regime shift*. The underlying characteristics of a stability landscape that can give rise to such shifts are illustrated figuratively, and described, in Figure 17.25. This stability landscape can be applied to whole communities, but also, for example, to ‘domino effects’ in spatially explicit metacommunities, where the state of each subcommunity has a dependence on the states of nearby subcommunities to which it is connected (Scheffer *et al.*, 2012). The figure also highlights another key feature of such systems, namely the difficulty, once a regime shift has taken place, of reversing that shift: only a small disturbance or change of conditions may be required to provoke the shift but a much more profound change may be necessary to restore the system to its former state.


Perturbation
Conditions

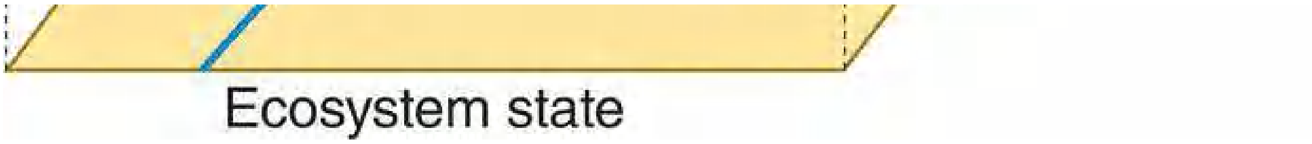


Figure 17.25 Figurative representation of a stability landscape giving rise to a regime shift in the state of an ecosystem. Imagine environmental conditions changing gradually from the front to the back of the figure. The five landscapes represent the stability of the state of the ecosystem. Initially it rests stably to the left (black ball). As conditions change, it remains in that state, and small disturbances will not shift it, despite the appearance of an alternative stable state to the right (grey ball). However, a small change in conditions from the third to the fourth landscape would be enough to allow a small disturbance to carry the ecosystem from one stable state to the other (left to right, two black balls), and a further small change in conditions would make the right-hand state as stable as the left-hand one was initially. The changing state of the ecosystem is also traced on the base of the figure. Note that over a wide range of conditions the ecosystem can exist in either of two stable states, and which it occupies depends largely on its previous state – a phenomenon known as 'hysteresis', where the state of a system depends on its history.

Source: After Scheffer & Carpenter (2003).

evidence of regime shifts

Evidence from field data for regime shifts comes in at least two forms (Scheffer & Carpenter, 2003). The most direct is from time series, where the jump from one state to another is actually captured (Figure 17.26a). Alternatively, what is ostensibly the same habitat at different locations may support very different communities such that key community characteristics exhibit a multimodal distribution (Figure 17.26b). As ever, experimental (as opposed to observational) evidence, while difficult to obtain, can be better at making a direct link between cause (a small change in the environment) and effect (a large change in the community). This, too, may be of various types: for example, small initial differences diverging to much larger, stable differences during the course of an experiment (Figure 17.26c), an experimental disturbance triggering a shift in state (Figure 17.26d) or direct observation of the hysteresis we saw in Figure 17.25 – a community behaving very differently in response to conditions changing in one direction to its response when they change in the opposite direction (Figure 17.26e).

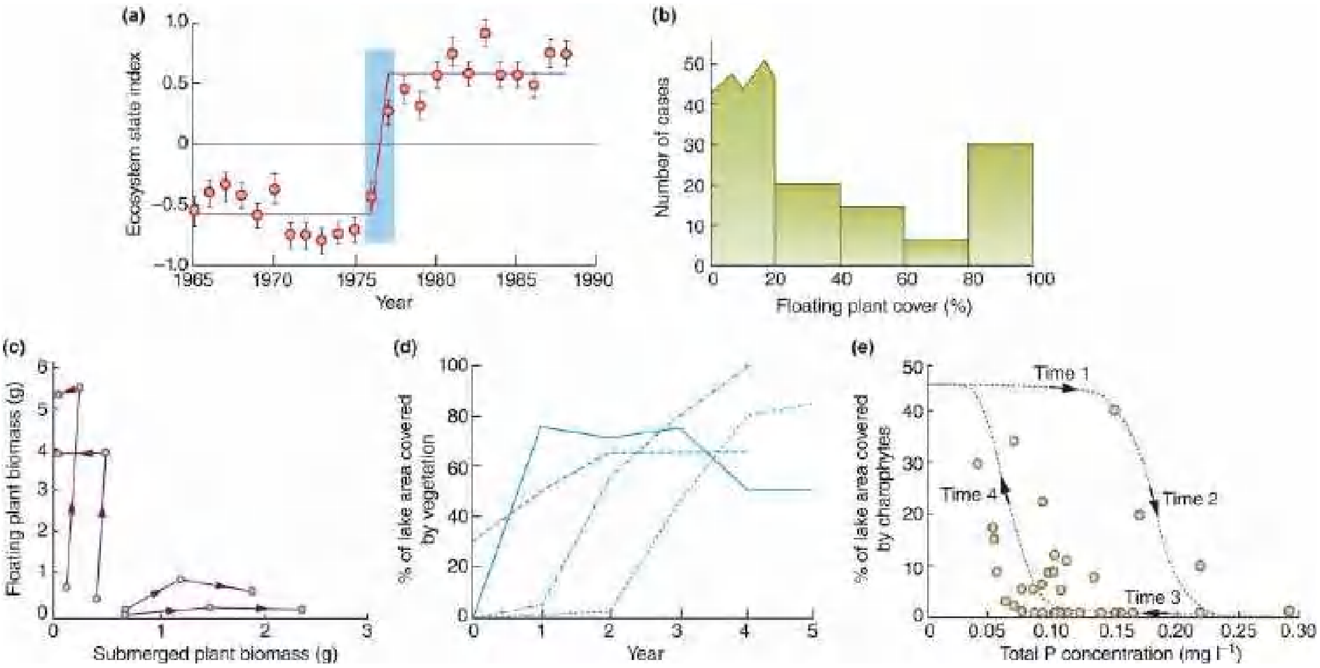


Figure 17.26 Evidence for regime shifts in the states of aquatic communities and ecosystems. (a) Step change in the state of the ecosystem of the North Pacific in 1977, where ecosystem state was a composite of 31 physical and 69 biological time series. (b) Bimodal frequency distribution of the percentage cover of free-floating plants in 158 Dutch ditches. (c) Experiment from the same study as in (b), and hence accounting for the results there, in which a floating plant (*Lemna*) and a submerged plant (*Elodea*) were grown together under the same conditions, but at three different starting abundances. Arrows show the changing state of the system over time. (d) Shifts in the vegetation cover of four lakes in the Netherlands and Denmark triggered by the removal of fish at time zero. (e) Hysteresis in the response of charophyte algae in a Dutch lake to a decrease and subsequent increase in phosphorus concentration. Details of the original references for the examples can be found in Scheffer and Carpenter (2003).

Source: After Scheffer & Carpenter (2003).

anticipating regime shifts

Given the suddenness of these changes and the difficulties in reversing them, it is understandable that effort is being invested in uncovering early warning signs of impending regime shifts. Chief among the candidates for this is ‘critical slowing down’, and from Figure 17.25 it is easy to see why. Often, as tipping points are approached, the landscape ‘flattens’ – an equilibrium remains stable in the sense that disturbances away from it are reversed, but resilience, the speed of that reversal, declines. Reliable detection of critical slowing down, however, or of other diagnostics, remains a major challenge (Scheffer *et al.*, 2012).

slow regime shifts

Tied to the problem of anticipating regime shifts can be the problem of detecting them once they have happened. Again, Figure 17.25 helps us understand this. A system may pass a tipping point and find itself in a new domain of attraction *en route* to a new state. That passage may be effectively inexorable and irreversible. But it may still be slow and hence difficult to detect, at least immediately (Hughes *et al.*, 2013). An apparent example is shown in Figure 17.27: the transition throughout many of the shallow seas of the Caribbean, over 25–30 years, from communities dominated by hard corals to ones in which coral cover is less than 10%. There are likely to have been a succession of drivers over this period: overfishing, pollution, disease and climate change, alongside more acute shocks to the system from hurricanes and coral bleaching. It is as important for us to acknowledge these slow transitions while we are living on borrowed time as for us to anticipate rapid transitions that may stop the clock.

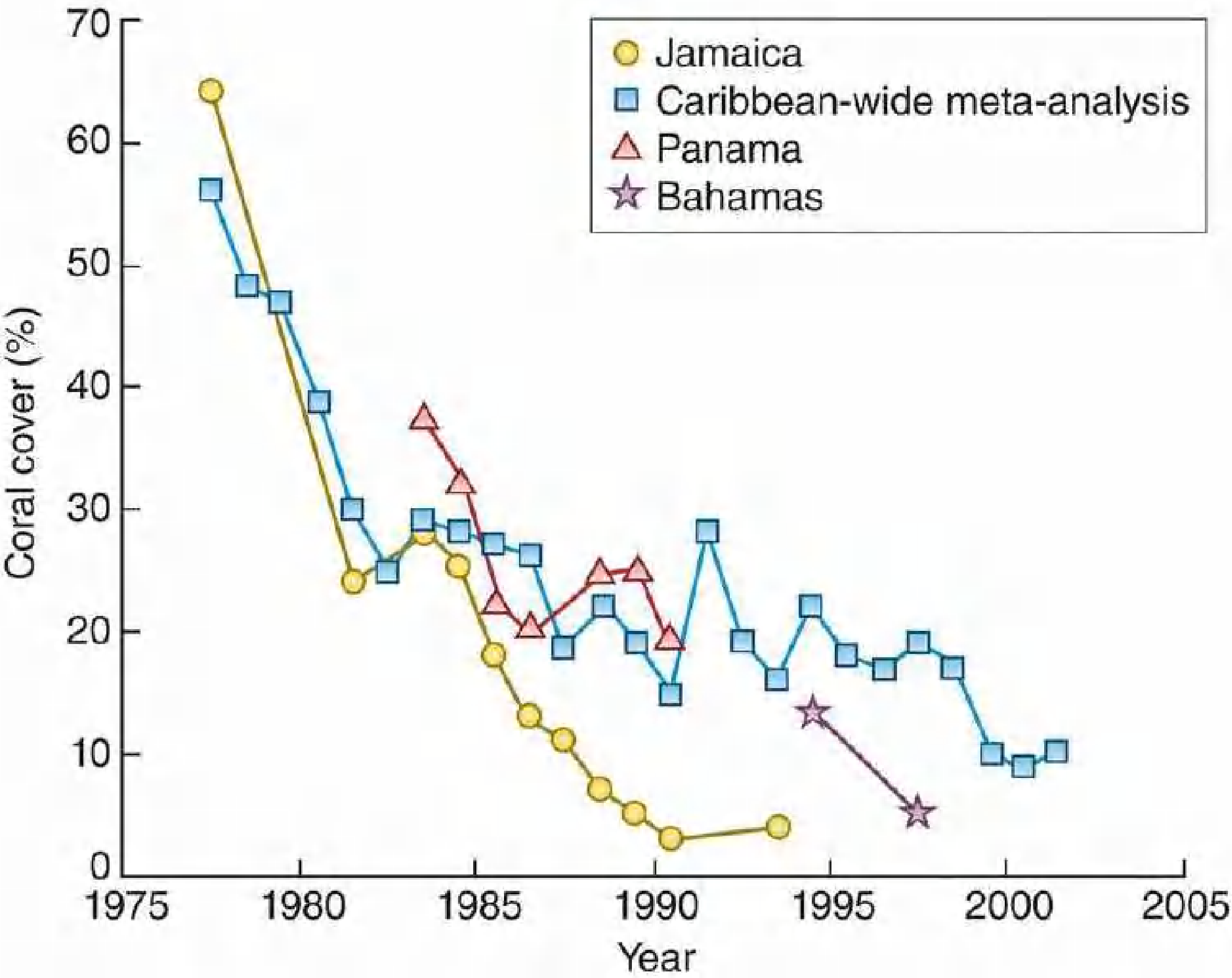


Figure 17.27 Slow regime shifts in the Caribbean from coral-dominated to coral-depleted communities. Data from Jamaica, a Caribbean-wide meta-analysis, Panama and the Bahamas. References to the original studies can be found in Hughes *et al*. (2013).

Source: After Hughes *et al*. (2013).

APPLICATION 17.5 A permafrost tipping point in the global ecosystem?

The very nature of tipping points and regime shifts makes them causes for concern. Any change that threatens the well-being of ourselves or of the natural world more generally is problematic. But change that occurs unexpectedly, or too rapidly for us to be able to respond effectively, or requires a particularly profound response in order to reverse it, is especially problematic. One example, at the global scale, is the effect of the thawing of permafrost on the progression of climate change (Schuur *et al.*, 2015). Permafrost is ground, including soil and rock, that remains below 0°C for two or more years, though in practice we think of it as having been frozen for centuries, at least. By its nature it has existed in high-latitude and high-altitude regions, but the area of permafrost that has attracted most attention has been the vast region affected by it in the northern Arctic and sub-Arctic zones (Figure 17.28). The critical issue is that vast quantities of carbon from the dead bodies of animals and plants are stored in the permafrost – more than a thousand billion tons in surface layers down to 3 m, with close to a further billion tons below that (Schuur *et al.*, 2015). For as long as this remains frozen, it also remains inert. But once it thaws, it becomes susceptible to microbial decomposition, releasing greenhouse gases into the atmosphere – CO_2 where aerobic decomposition is possible, but also CO_2 and methane (CH_4) from deeper, anaerobic layers (see Section 3.6).

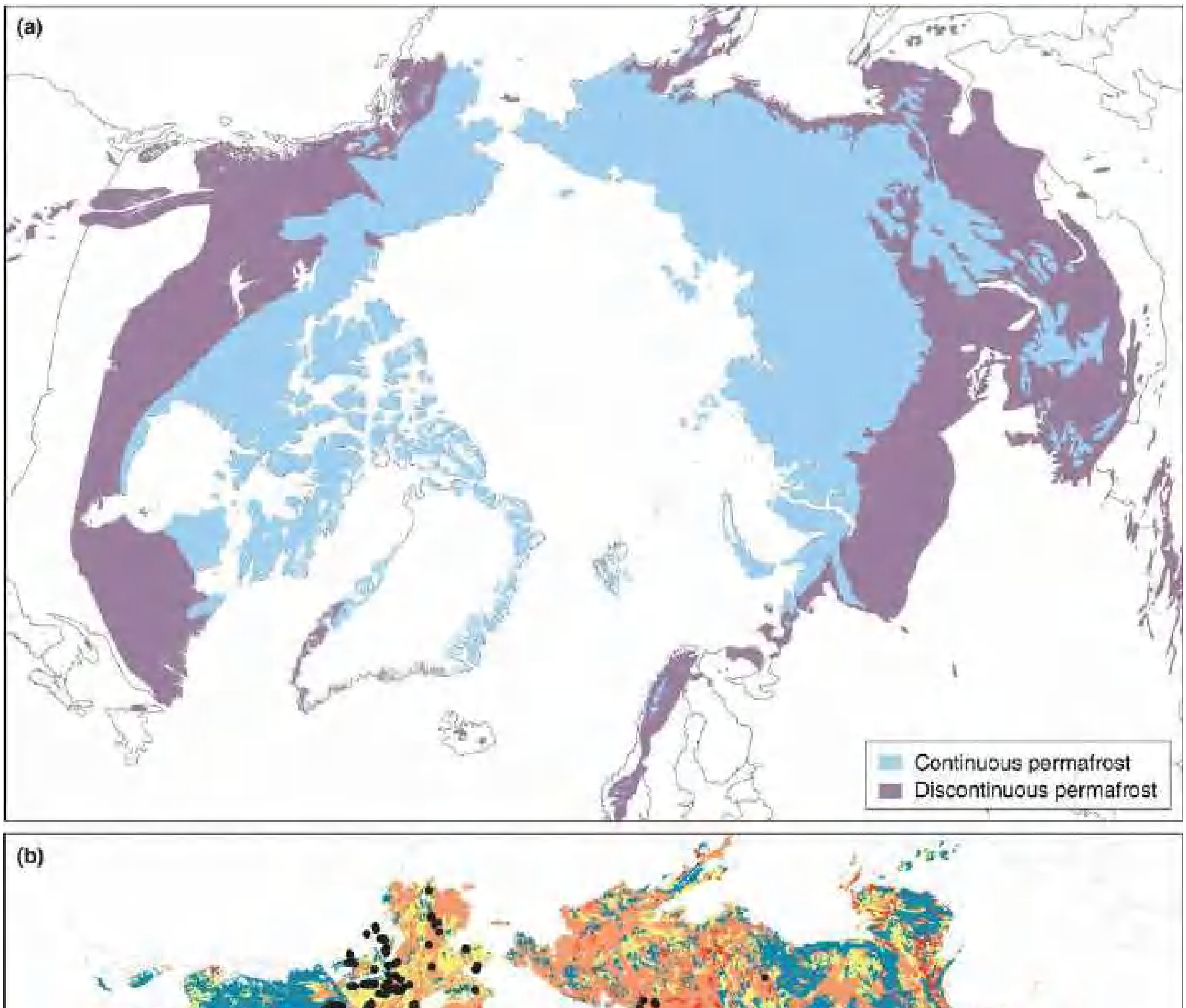

(a)
Continuous permafrost
Discontinuous permafrost
(b)
Soil organic carbon storage (0–3 m):
0.1–30 kg m^{-2}
30–50 kg m^{-2}
50–100 kg m^{-2}
100–260 kg m^{-2}

Figure 17.28 **Distribution of permafrost at high northern latitudes and the associated distribution of organic carbon stored in the soil that could be released to the atmosphere through global warming.** (a) The distribution of continuous (>90% of locations affected) deep permafrost (>3 m) and discontinuous (<90% of locations affected) deep permafrost. (b) Estimated distribution of the soil organic carbon pool in these regions at depths of 0–3 m. Dots show locations of inventory points.

Source: After Schuur *et al.* (2015).

The potential for a regime shift is easy to see. An initial regime exists in which the permafrost is inert. Global warming, the consequence of increases in the atmospheric concentration of greenhouse gases, leads to thawing of the permafrost. This thawing releases greenhouse gases, which accelerate the rate of global warming, leading to further thawing, further release of gases, further warming and so on, such that we arrive at a new regime in which those high-latitude soils are an active part of the global carbon cycle. Of course, compensatory processes are also likely. The warming will encourage previously impossible plant growth in those same regions, sequestering some, at least, of the newly released carbon. The question, then, is how will this process play out, overall? What will be the consequences for the global climate and the global biosphere?

Remarkably, in part because of our ignorance of many aspects of its extent, its composition, and so on, the Intergovernmental Panel on Climate Change (IPCC)'s Fifth Assessment Report (IPCC, 2014; see Section 22.2) included no consideration of carbon emissions from permafrost. There is little doubt that future projections will do so. In the mean time, there is an urgent need for further data, but the data available do allow future scenarios to be modelled. Some suggest a rapid switch into that new regime (Whiteman *et al.*, 2013). Others project a much more gradual regime shift, giving us some opportunity, at least, to respond (Chadburn *et al.*, 2017). All agree that the thawing of permafrost will compound what is already arguably the greatest threat facing the biosphere.

Chapter 18

Patterns in Community Composition in Space and Time

18.1 Introduction

A community is an assemblage of species populations that occur together in space and time. Community ecology seeks to understand the manner in which groupings of species are distributed in nature, and the ways these groupings can be influenced by their abiotic environment and by interactions among species populations. The challenge for community ecologists is to discern and explain patterns arising from this multitude of influences. In Chapters 16 and 17, we focused on the role of species interactions in determining community patterns and on the manner in which species modules assemble into food webs. Here we place the community into a broader context and pay particular increased attention to the role of abiotic factors.

the search for rules of community assembly

The species that assemble to make up a community can be viewed as the outcome of passing through a series of filters: (i) dispersal constraints; (ii) environmental constraints; and (iii) internal dynamics (Figure 18.1a) (Belyea & Lancaster, 1999). Chapter 16 dealt in detail with the internal dynamics of population interactions including competition, facilitation, predation and parasitism.

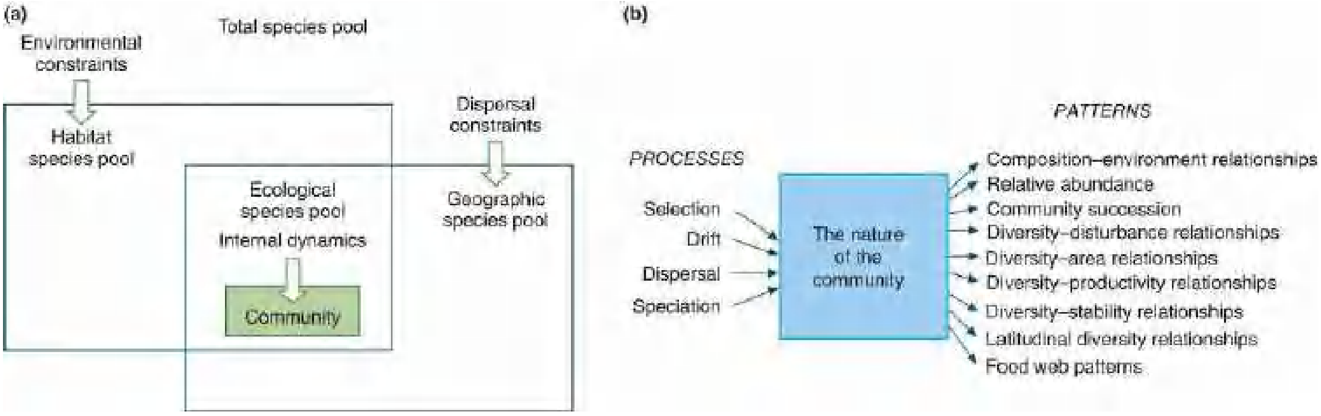


Figure 18.1 The relationships among four types of species pools and four classes of process. (a) The total pool of species in a region, the geographic pool (species able to arrive at a site), the habitat pool (species able to persist under the abiotic conditions of the site), the ecological pool (the overlapping set of species that can both arrive and persist) and the community (the pool that remains in the face of biotic interactions). (b) A wide variety of community patterns may be understood in terms of four classes of process.

Source: (a) After Belyea & Lancaster (1999) and Booth & Swanton (2002). (b) After Vellend (2010).

The concept can be extended by dividing internal dynamics into two classes – 'selection' (deterministic interactions such as those between competitors or predators and prey) and 'drift' (stochastic or random changes in species relative abundances) – and also by adding the process of 'speciation', recognising that different geographic areas have different evolutionary histories that have resulted in different species pools (Hubbell, 2001 ; Vellend, 2010). Ecologists search for rules of community assembly, and with Vellend's four classes of process in mind (dispersal, selection, drift and speciation) we discuss community assembly rules in this chapter.

Science at the community level poses daunting problems because the species catalogue may be enormous and complex. A first step is usually to search for patterns, such as the repeated grouping of similar growth forms in different places, or repeated trends in species richness along different environmental gradients. Recognition of patterns leads, in turn, to the forming of hypotheses about the causes of these patterns. The hypotheses may then be tested by making further observations or by doing experiments.

A community can be defined at any scale within a hierarchy of habitats. At one extreme, broad patterns in the distribution of community types can be recognised on a global scale. The temperate forest biome is one example; its range in North America is shown in Figure 18.2. At this scale, ecologists usually recognise climate as the overwhelming factor that determines the limits of vegetation types. At a finer scale, the temperate forest biome in parts of New Jersey is represented by communities of two species of tree in particular, beech and maple, together with a very large number of plants, animals and microorganisms. Study of the community may be focused at this scale. On an even finer habitat scale, the characteristic invertebrate community that inhabits water-filled holes in beech trees may be studied, or the biota in the gut of a deer in the forest. Amongst these various scales of community study, no one is more legitimate than another. The scale appropriate for investigation depends on the sorts of questions that are being asked.

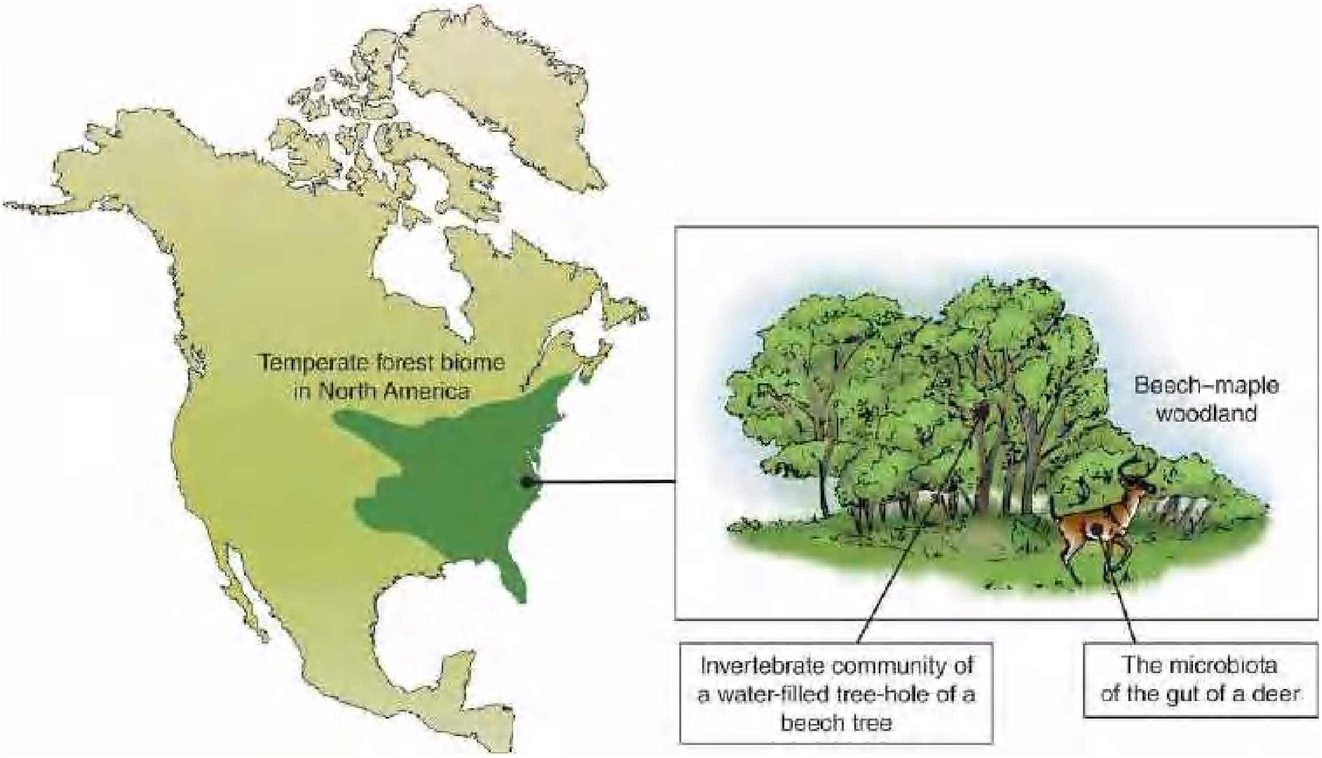


Figure 18.2 **A community can be defined at any scale.** We can identify a hierarchy of habitats, nesting one into the other: a temperate forest biome in North America; a beech–maple woodland in New Jersey; a water-filled tree hole; or a mammalian gut. The ecologist may choose to study the community that exists on any of these scales.

communities can be recognised at a variety of levels – all equally legitimate

Community ecologists sometimes consider all of the organisms existing together in one area, although it is rarely possible to do this without a large team of taxonomists. Others restrict their attention to a single taxonomic group (e.g. birds, insects or trees), or a group with a particular activity (e.g. herbivores or detritivores).

The rest of this chapter is in seven sections. We start by explaining how the structure of communities can be measured and described (Section 18.2). Then we focus on patterns in community structure: in space (Section 18.3), in time (Sections 18.4 and 18.5) and in a combined spatiotemporal setting (Section 18.6). Then, in a final section, particular attention will be paid to the concept of the metacommunity – a set of communities linked by dispersal (analogous to the metapopulations discussed in Chapter 6).

18.2 Description of community composition

species richness: the number of species present in a community

One way to characterise a community, as we saw in a number of studies described in Chapter 16, is simply to count or list the species that are present. This sounds a straightforward procedure that enables us to describe and compare communities by their species 'richness' (i.e. the number of species present). In practice, though, it is often surprisingly difficult, partly because of taxonomic problems, but also because only a subsample of the organisms in an area can usually be counted. The number of species recorded then depends on the number of samples that have been taken, or on the volume of the habitat that has been explored. The most common species are likely to be represented in the first few samples, with rarer species added as more samples are taken. At what point can we stop taking further samples? Ideally, the investigator should continue to sample until the number of species reaches a plateau (Figure 18.3). At the very least, the species richnesses of different communities should be compared on the basis of the same sample sizes (in terms of area of habitat explored, time devoted to sampling or number of individuals or modules included in the samples). Statistical approaches to standardise estimates of richness from different communities include random resampling of the pool of *N* samples to provide a mean richness for a standardised sampling effort for each community (rarefaction) or extrapolation to a theoretical asymptote (see Cayuela *et al*., 2015 for details). The analysis of species richness in contrasting situations figures prominently in Chapter 19.

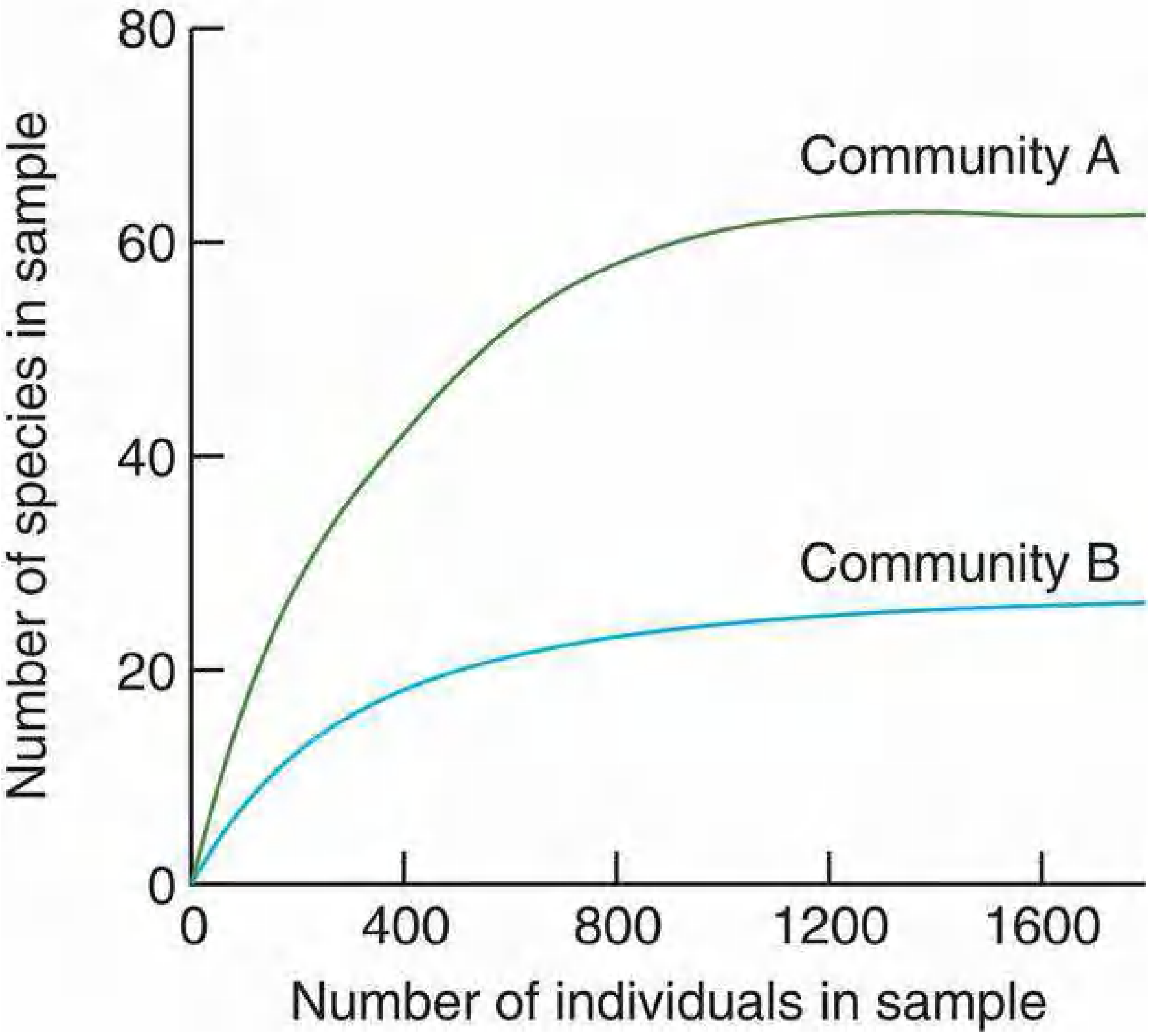


Figure 18.3 The relationship between species richness and the number of individual organisms: total species richness in community A is much greater than in community B.

18.2.1 Diversity indices

diversity incorporates richness, commonness and rarity

An important aspect of community structure is completely ignored, though, when the composition of the community is described simply in terms of the number of species present. It misses the information that some species are rare and others common. Consider a community of 10 species with equal numbers in each, and a second community, again consisting of 10 species, but with more than 50% of the individuals belonging to the most common species and less than 5% in each of the other nine. Each community has the same species richness, but the first, with a more ‘equitable’ distribution of abundances, is clearly more *diverse* than the second. Richness and equitability combine to determine community diversity.

Simpson’s diversity index

The simplest measure of the character of a community that takes into account both abundance (or biomass) patterns and species richness, is Simpson’s diversity index. This is calculated by determining, for each species, the proportion of individuals or biomass that it contributes to the total in the sample (the proportion is P_i for the ith species):

$$\text{Simpson's index, } D = \frac{1}{\sum_{i=1}^{S} P_i^2} \tag{18.1}$$

where S is the total number of species in the community (i.e. the richness). As required, for a given richness, D increases with equitability, and for a given equitability, D increases with richness.

‘equitability’ or ‘evenness’

Equitability can itself be quantified (between 0 and 1) by expressing Simpson’s index, D, as a proportion of the maximum possible value D would assume if individuals were completely evenly distributed amongst the species. In fact, $D_{\max} = S$. Thus:

$$\text{equitability, } E = \frac{D}{D_{\max}} = \frac{1}{\sum_{i=1}^{S} P_i^2} \times \frac{1}{S}. \tag{18.2}$$

Shannon’s diversity index

Another index that is frequently used and has essentially similar properties is the Shannon diversity index, H. This again depends on an array of P_i values. Thus:

$$\text{diversity, } H = -\sum_{i=1}^{S} P_i \ln P_i \tag{18.3}$$

and:

$$\text{equitability, } J = \frac{H}{H_{\max}} = \frac{-\sum_{i=1}^{S} P_i \ln P_i}{\ln S}. \tag{18.4}$$

An example of an analysis of diversity is provided by a uniquely long-term study that started in 1856 in grassland at Rothamsted in England (Crawley *et al.*, 2005). Experimental plots received a fertiliser treatment once every year, whilst control plots did not. Figure 18.4 shows how species diversity (H) and equitability (J) of the grass species changed over a century. Whilst the unfertilised area remained essentially unchanged, the fertilised area progressively declined in diversity and equitability. One possible explanation is that high nutrient availability leads to high rates of population growth and a greater chance of the most productive species coming to dominate and, perhaps, competitively exclude others, a process we discuss more fully in Section 19.3.1.

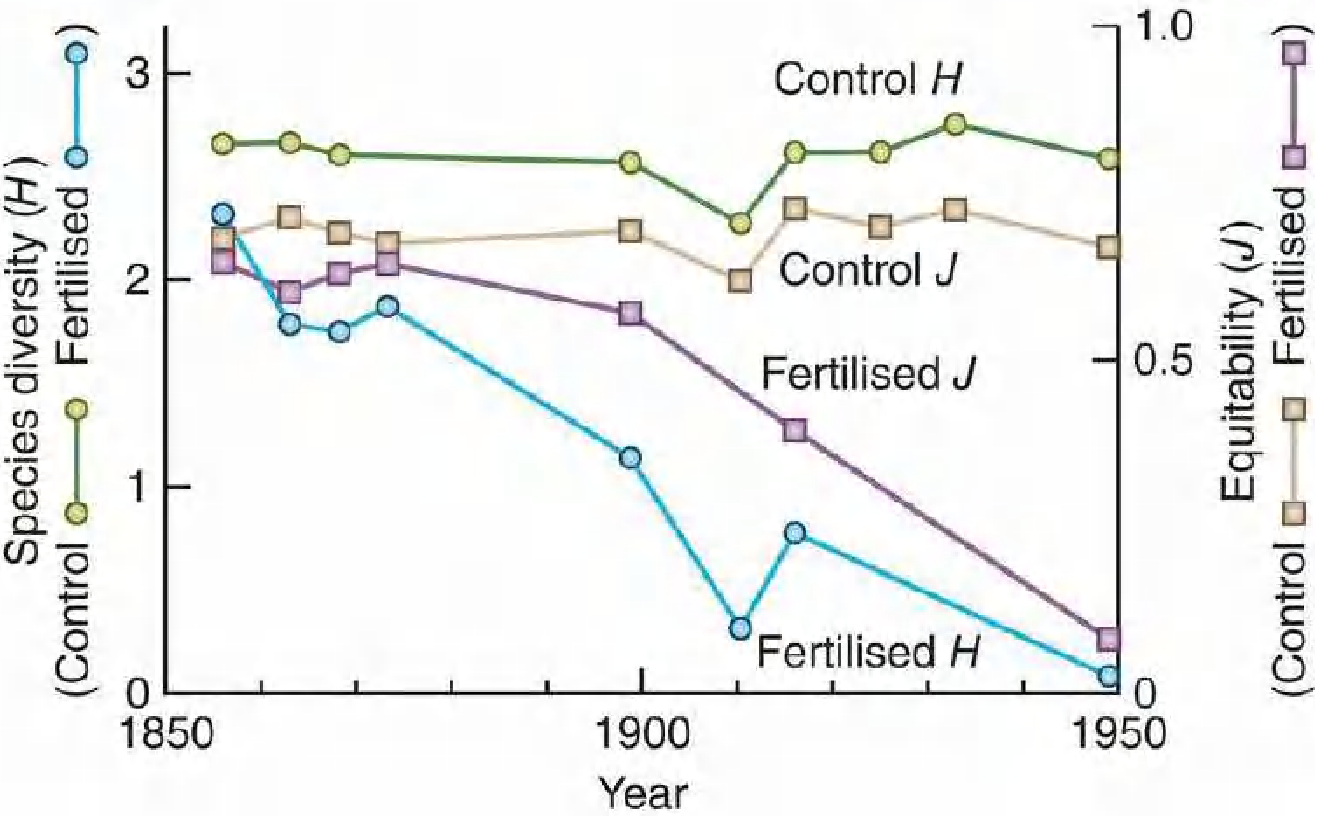


Figure 18.4 Species diversity (*H*) and equitability (*J*) decline progressively in a fertilised grassland plot. Indices calculated for a control plot and a fertilised plot in the long-term Rothamsted ‘Parkgrass’ experiment.

Source: After Tilman (1982).

Hill numbers

Increasingly, a case has been made for using *Hill numbers* to measure diversity. These have a long history (Hill, 1973), but having suffered a period of neglect, their advantages have more recently been recognised (Chao *et al*., 2014). The formula for a Hill number, *H*, is:

$$ {}^{q}H = \left(\sum_{i=1}^{S} P_i^{q} \right)^{1/(1-q)} , \tag{18.5} $$

where the terms are as previously defined with the exception of the exponent, *q*. One advantage they have is that by varying *q*, other measures of diversity can be recovered as special cases. Thus, when $q = 0$, *H* is simply species richness, *S*; when $q = 2$, *H* is the Simpson’s index, *D*; and, following some algebraic manipulation, when $q = 1$, *H* is the exponential of the Shannon index, *H*. Thus, developments in the methodologies for using one are more readily transferable to the others. Further advantages are discussed by Chao *et al*. (2014).

α-, β- and γ-species richness

The ideas of species richness (and diversity) need to be extended by acknowledging that their measurement depends on the scale of study. So far we have considered what is essentially a single homogeneous habitat patch. When dealing with a landscape of habitat patches containing different sets of species, it is important to separate out the total species richness of the region, referred to as γ-richness (similarly, the diversity of the whole region would be γ-diversity), the average species richness *within* patches (α-richness) and the between-patch component of regional richness (β-richness). Thus, γ-richness is the sum of α-richness and β-richness. If every patch has identical species lists, β-richness is zero and γ-richness equals α-richness. But β-richness will make a contribution to γ-richness wherever there is heterogeneity in the distribution of species among patches. Hui and McGeoch (2014) have recently proposed a fourth index, $\boldsymbol{\zeta}$ (zeta) diversity, where $\boldsymbol{\zeta}_i$ is the mean number of species shared by i sites. When $i = 1$, $\boldsymbol{\zeta}$ species richness ($\boldsymbol{\zeta}_1$) is simply the mean number of species across all sites. As more sites are added, the number of shared species ($\boldsymbol{\zeta}_i$) declines monotonically.

18.2.2 Rank–abundance diagrams

An even more complete picture of the distribution of species abundances in a community makes use of the full array of P_i values by plotting P_i against rank. Thus, the P_i for the most abundant species is plotted first, then the next most common, and so on until the array is completed by the rarest species of all. A rank–abundance diagram can be drawn for the number of individuals, or for the area of ground covered by different sessile species, or for the biomass contributed to a community by the various species. Indeed, we have already seen in Section 16.2.4 (and Figure 16.7) how these, and the closely-related species abundance distributions (McGill *et al*., 2007), have been used in judging the ability of neutral theory to account for key patterns in community structure.

rank–abundance models may be based on statistical or biological arguments

A range of the many equations that have been fitted to rank–abundance diagrams is shown in Figure 18.5. Two of these are statistical in origin (the log series and log-normal) with no foundation in any assumptions about how the species may interact with one another. The others take some account of the relationships between conditions, resources and species abundance patterns (niche-orientated models) and are more likely to help us understand the mechanisms underlying community organisation. These include the geometric series, in which each species in sequence takes a constant proportion (k) of individuals; thus, if k is 50% the most common species has 50% of individuals, the next has 25%, etc.

Tokeshi (2009) presents a full range of niche-orientated models and we focus on four of them. The *dominance–pre-emption model*, which produces the least equitable species distribution, has successive species pre-empting a dominant portion (50% or more) of the remaining niche space (this is close to the geometric series). A somewhat more equitable distribution is represented by the *random fraction model*, in which successive species invade and take over an arbitrary portion of the niche space of any species previously present. The *MacArthur fraction model*, on the other hand, assumes that species with larger niches are more likely to be invaded by new species; this results in a more equitable distribution than the random fraction model. Finally, the *dominance–decay model* is the inverse of the dominance–pre-emption model, in that the largest niche in an existing assemblage is always subject to a subsequent (random) division. Thus, in this model the next invading species is supposed to colonise the niche space of the species currently most abundant, yielding the most equitable species abundances of all the models.

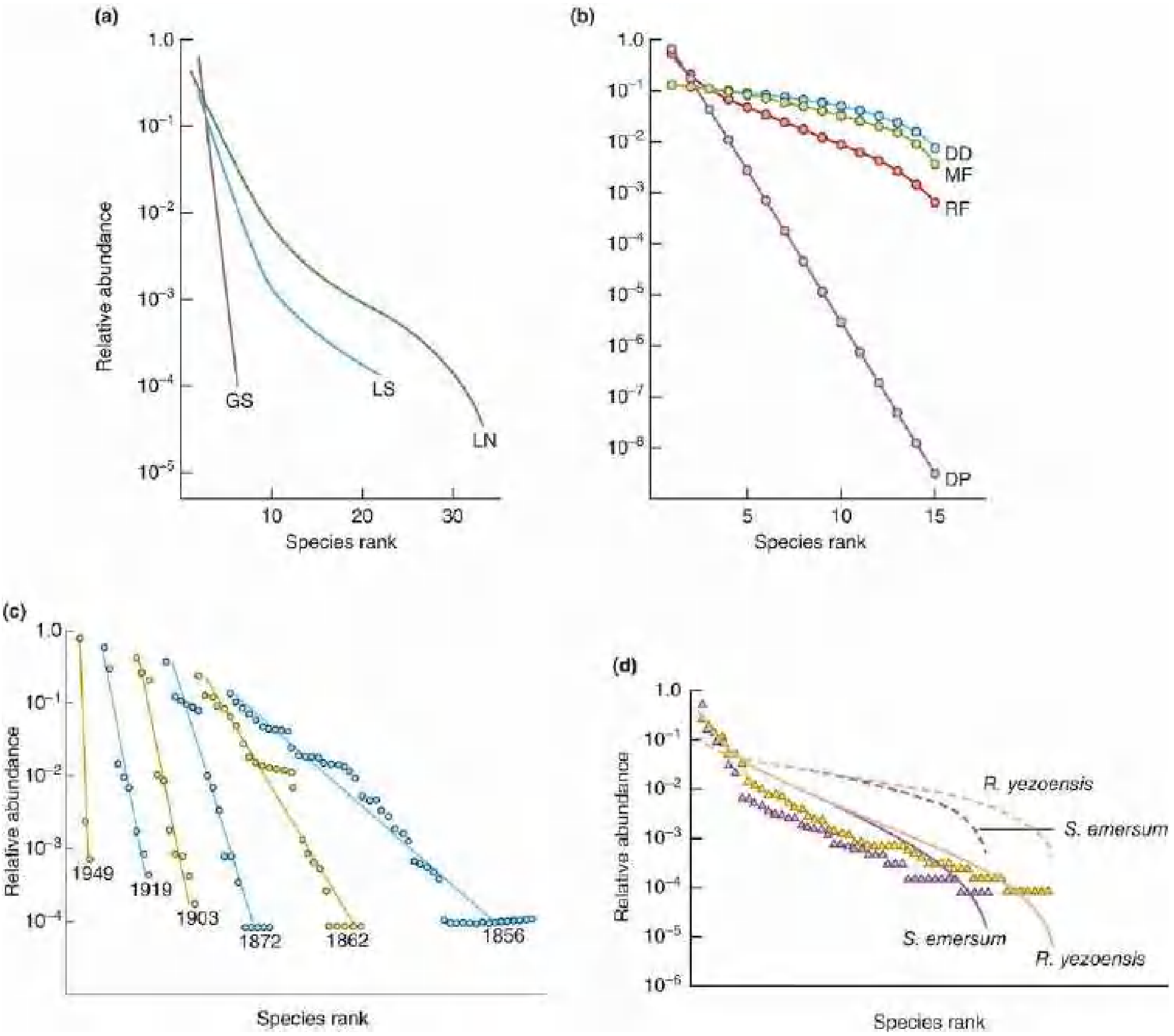


Figure 18.5 Rank–abundance patterns of various models. (a) GS, geometric series; LN, log-normal; LS, log series. (b) DD, dominance decay; DP, dominance pre-emption; MF, MacArthur fraction; RF, random fraction. (c) Change in the relative abundance pattern (geometric series fitted) of plant species in the experimental grassland in Figure 18.4 subjected to continuous fertiliser from 1856 to 1949. (d) Comparison of rank–abundance patterns for invertebrate species living on a structurally complex stream plant *Ranunculus yezoensis* (yellow symbols) and a simple plant *Sparganium emersum* (purple symbols); fitted lines represent the MacArthur fraction model (dashed lines) and the random fraction model (solid lines).

Source: (a–c) After Tokeshi (2009). (d) After Taniguchi *et al.* (2003).

the value of community indices

Rank–abundance diagrams, like indices of richness, diversity and equitability, should be viewed as abstractions of the highly complex structure of communities that may be useful when making comparisons. The idea is that the best-fitting model should give us clues as to underlying processes. Indeed, some studies have successfully focused attention on a change in dominance/evenness relationships in relation to environmental change. Figure 18.5c shows how, assuming a geometric series can be appropriately applied, dominance steadily increased, whilst species richness decreased, during the long-term Rothamsted grassland experiment described above. Figure 18.5d shows how invertebrate species richness and equitability were both greater on an architecturally complex stream plant (*Ranunculus yezoensis*), which provides more potential niches, than on a structurally simple plant (*Sparganium emersum*). The rank–abundance diagrams of both are closer to the random fraction model than the MacArthur fraction model, suggesting that the probability of niche invasion/division is not related to original niche size.

18.2.3 Community size spectra

species functional traits

As we saw in Section 17.2.4, much recent research effort has been devoted to understanding the link between species richness and ecosystem functioning (productivity, decomposition and nutrient dynamics). Unravelling patterns in functioning can benefit from a focus not only on the relative abundance of different taxa but also on the relative representation within a community of different functional traits (basal metabolic rate, reproductive mode, degree of specialisation, frost tolerance, diet, etc.) (Cadotte *et al.*, 2011). After all, it is the combination of functional traits, not the detailed taxonomy, that drives ecosystem functioning. Among the functional traits that community ecologists study, body size is particularly important because of its links to many life history attributes (e.g. longevity and reproductive rate) and ecological attributes (e.g. competitiveness and vulnerability to predators). This has led to a search for patterns in the size spectra of communities – the frequency distributions of sizes of individuals present – an approach that pays no attention to species identity or species richness.

Vertebrate communities often exhibit multimodal size spectra, as seen for breeding birds in North America (Figure 18.6a). The pattern holds across spatial scales, from local communities to regional and continental species pools, and several or many species are represented in each mode. Multimodality may reflect the irregular availability of resources along the body size axis or limits to the similarity of competing species that can coexist (Thibault *et al.*, 2011; and see Section 16.2.4). In other cases size spectra are unimodal, as for some insect and deep-sea gastropod communities (Figure 18.6b), while a third common pattern is the monotonic decline exhibited by tree assemblages (Figure 18.6c) and many aquatic and soil communities.

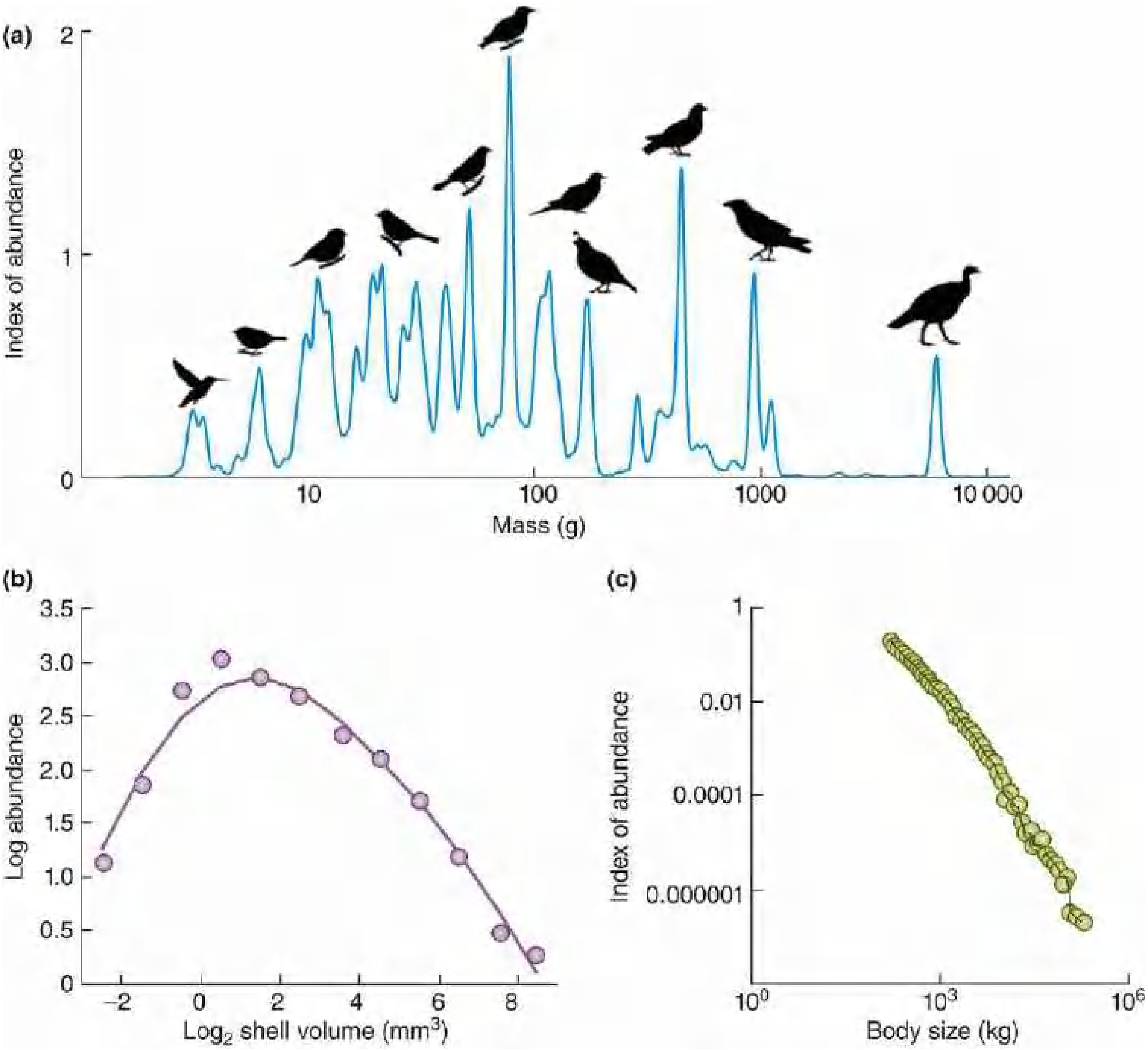


Figure 18.6 Multimodal, monomodal and monotonic size spectra. (a) The multimodal size spectrum of breeding birds in North America: the pattern is consistent whether the scale of study is the local community or the whole continental species pool, as illustrated here. Silhouettes represent common species or species groups that contribute to the corresponding peaks. (b) The monomodal size spectrum of deep-sea gastropods in the western North Atlantic. (c) Monotonically declining size spectrum for trees on Barro Colorado Island, Panama.

Source: (a) From Thibault *et al*. (2011). (b) From McClain (2004). (c) From White *et al*. (2007).

APPLICATION 18.1 Size spectra as indicators of human impairment of communities

Size spectra of assemblages of marine fish typically become steeper following the exploitation of fisheries, with poorer representation of the large individuals preferred by the market. Such a consistent change in pattern allows estimates of distance from the non-impacted reference state, which can then be used by managers as indicators of the impaired status of fish communities and of marine ecosystems generally (Figure 18.7). Mulder and Elser (2009) discovered that, just as with fish communities, the slopes of the size spectra of grassland soil communities (bacteria, fungi, nematodes, mites, springtails and enchytraeid worms) become steeper in more intensively managed pasture. It may turn out that the steepening in slope of monotonically declining size spectra signals a general effect of human pressures of different kinds (Petchey & Belgrano, 2010).

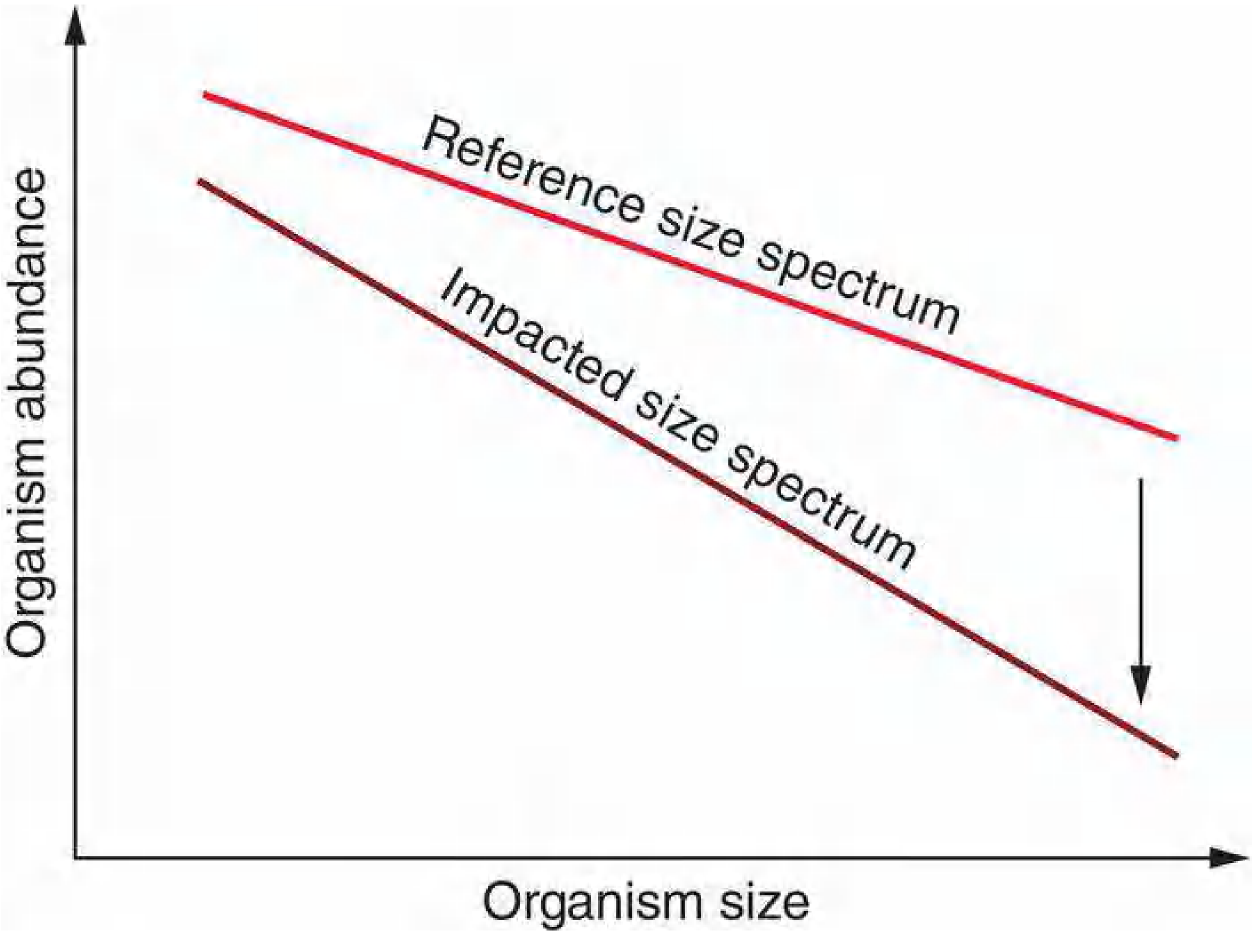


Figure 18.7 Under environmental pressures size spectra become steeper. Under pressures such as exploitation of fish by a fishery, or the consequences for the soil community of intensive management of pasture, size spectra become steeper.

Source: From Petchey & Belgrano (2010).

18.3 Community patterns in space

18.3.1 Gradient analysis

Figure 18.8 shows a variety of ways of describing the distribution of vegetation in the Great Smoky Mountains of Tennessee, USA, in the classic study of Whittaker (1956). Figure 18.8a is a subjective analysis that acknowledges that the vegetation of particular areas differs in a characteristic way from that of other areas. It could be taken to imply that the various communities are sharply delimited. Figure 18.8b, while highlighting two potentially influential environmental gradients (altitude and moisture), gives the same impression of sharp boundaries. Note that both Figure 18.8a and b are based on descriptions of the *vegetation*. However, Figure 18.8c changes the focus by concentrating on the pattern of the distributions of the individual *species* (expressed as a percentage of all tree stems present). It is then immediately obvious that there is considerable overlap in their abundance – there are no sharp boundaries. The various tree species are strung out along the gradient with the tails of their distributions overlapping. Many other gradient studies have produced similar results.

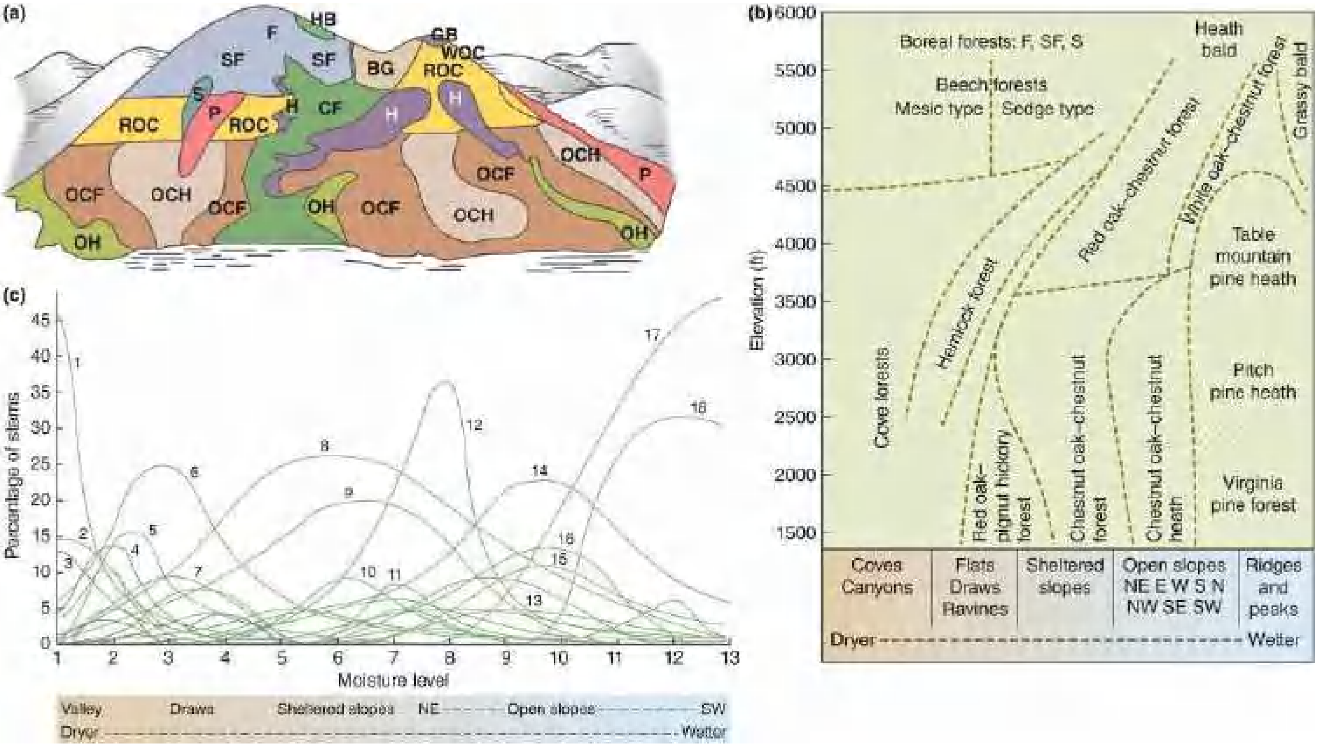


Figure 18.8 Three contrasting descriptions of distributions of the characteristic dominant tree species of the Great Smoky Mountains, Tennessee. (a) Topographic distribution of vegetation types on an idealised west-facing mountain and valley. (b) Idealised graphic arrangement of vegetation types according to elevation and aspect. (c) Distributions of individual tree populations (percentage of stems present) along the moisture gradient. Vegetation types: BG, beech gap; CF, cove forest; F, Fraser fir forest; GB, grassy bald; H, hemlock forest; HB, heath bald; OCF, chestnut oak–chestnut forest; OCH, chestnut oak–chestnut heath; OH, oak–hickory; P, pine forest and heath; ROC, red oak–chestnut forest; S, spruce forest; SF, spruce–fir forest; WOC, white oak–chestnut forest. Selected species: 1, *Halesia monticola*; 6, *Tsuga canadensis*; 8, *Acer rubrum*; 12, *Quercus montana*; 14, *Oxydendrum arboreum*; 17, *Pinus virginiana*.

Source: After Whittaker (1956).

the choice of gradient is almost always subjective

Perhaps the major criticism of gradient analysis as a way of detecting pattern in communities is that the choice of the gradient is almost always subjective. The investigator searches for some feature of the environment that appears to matter to the organisms and then organises the data about the species concerned along a gradient of that factor. It is not necessarily the most appropriate factor to have chosen. The fact that the species from a community can be arranged in a sequence along a gradient of some environmental factor does not prove that this factor is the most important one. It may only imply that the factor chosen is more or less loosely correlated with whatever really matters in the lives of the species involved. Gradient analysis is only a small step on the way to the objective description of communities.

18.3.2 The ordination of communities

Formal statistical techniques have been devised to take the subjectivity out of community description. These allow the data from community studies to sort themselves, without the investigator putting in any preconceived ideas about which species tend to be associated with each other or which environmental variables correlate most strongly with the species distributions. One commonly used technique is ordination.

> in ordination, communities are displayed on a graph with those most similar in composition closest together

Ordination is a mathematical treatment that allows communities to be organised on a graph so that those that are most similar in both species composition and relative abundance will appear closest together, whilst communities that differ greatly in the relative importance of a similar set of species, or that possess quite different species, appear far apart. Figure 18.9a shows the application of a particular ordination technique, canonical correspondence analysis (CCA), to the microscopic rotifer communities of a number of lakes in the North Island of New Zealand (Duggan *et al.*, 2002). CCA also allows the community patterns to be examined in terms of a variety of physicochemical factors (Figure 18.9a). Thus, the compositions of the communities lead to their being spread out along the two axes of the ordination graph, and the relationship between the environmental variables and these axes can then be displayed. Obviously, the success of the method now depends on having sampled an appropriate set of environmental variables. This is a major snag in the procedure – we may not have measured the qualities in the environment that are most relevant.

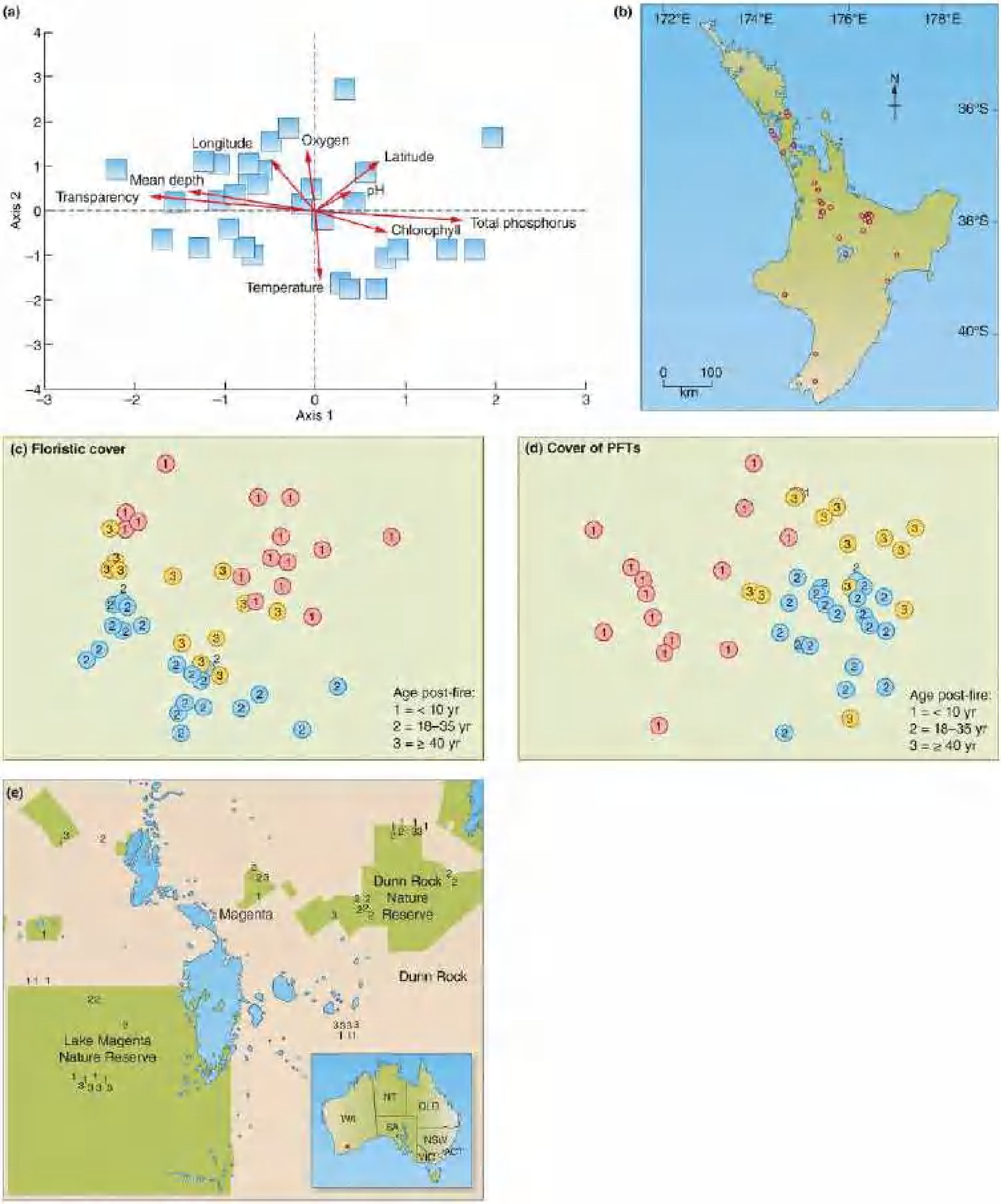

(a)
Axis 2
Axis 1
Longitude
Oxygen
Latitude
Mean depth
pH
Transparency
Chlorophyll
Total phosphorus
Temperature
(b)
172°E
174°E
176°E
178°E
36°S
38°S
40°S
N
0
100
km
(c) Floristic cover
Age post-fire:
1 = < 10 yr
2 = 18–35 yr
3 = ≥ 40 yr
(d) Cover of PFTs
Age post-fire:
1 = < 10 yr
2 = 18–35 yr
3 = ≥ 40 yr
(e)
Magenta
Dunn Rock Nature Reserve
Dunn Rock
Lake Magenta Nature Reserve
WA
NT
QLD
SA
NSW
VIC
ACT

Figure 18.9 Examples of ordination of community composition. (a) Results of canonical correspondence analysis (ordination) of the community composition of rotifers (assessed as relative abundance) in 31 New Zealand lakes. The complete species pool consisted of 78 species. The positions of the communities in ordination space are shown for each lake (squares) and for associated environmental factors (red arrows). (b) The position of the studied lakes around the North Island of New Zealand. (c) Results of non-metric multidimensional scaling (ordination) of the community composition of 'mallee-heath' vegetation at 45 sites in south-western Australia (assessed as percent ground cover). The complete species pool consisted of 305 species. Each site is classified as 1–3 in terms of time since the vegetation was last subject to fire. (d) Results of the equivalent analysis for community composition measured in terms of the representation (% cover) of plant functional traits (PFTs). (e) The position of the Australian study sites in the Lake Magenta area. *Source*: (a) After Duggan *et al.* (2002). (c–e) After Gosper *et al.* (2012).

Lakes that have been subject to a greater level of runoff of agricultural fertilisers or input of sewage are described as eutrophic. These tend to have high total phosphorus concentrations, leading to higher chlorophyll levels and lower transparency (a greater abundance of phytoplankton cells). Figure 18.9a shows that the rotifer communities are strongly influenced by these three indicators of the eutrophication to which the lakes are subject (operating along axis 1 of the ordination). The level of eutrophication, however, is not the only significant factor in accounting for rotifer community composition. The communities are also differentiated on axis 2 along which dissolved oxygen concentration and lake temperature are operating (themselves negatively related because oxygen solubility declines with increasing temperature).

ordination can generate hypotheses ...

What do these results tell us? First, they emphasise that under a particular set of environmental conditions, a predictable association of species is likely to occur: community ecologists have more than just a totally arbitrary and ill-defined set of species to study. Second, the correlations with environmental factors, revealed by the analysis, give us some specific hypotheses to test about the relationship between community composition and underlying environmental factors. (Remember that correlation does not necessarily imply causation. For example, dissolved oxygen and community composition may vary together because of a common response to another environmental factor.)

... or test hypotheses

Gosper *et al.* (2012) used a similar ordination approach, but this time to test some specific hypotheses about the determinants of community composition. The highly diverse Australian mallee-heath communities (dominated by shrubs and *Eucalyptus* trees) are fire-prone and the researchers predicted that community composition would reflect the time since a site was last subject to burning. Figure 18.9c shows the ordination of 45 mallee-heath communities according to their taxonomic composition and shows that more recently burnt sites (<10 years) are indeed predictably different from those that have remained unburned for longer periods. Gosper and colleagues further hypothesised that particular functional traits would underlie this temporal community pattern, including the ability to re-sprout after fire, the persistence and location of propagules on plants or in soil, plant height and longevity. Figure 18.9d therefore shows an alternative ordination based not on taxonomy but on the representation of these and other influential functional traits. It turns out that community differentiation along the time axis was even clearer than in the taxonomic ordination.

18.3.3 Problems of boundaries in community ecology

are communities discrete entities with sharp boundaries?

There may be communities that are separated by clear, sharp boundaries, where groups of species lie adjacent to, but do not intergrade into, each other. If they exist, they are exceptional. The meeting of terrestrial and aquatic environments might appear to be a sharp boundary but its ecological unreality is emphasised by the otters or frogs that regularly cross it and the many aquatic insects that spend their larval lives in the water but their adult lives as winged stages on land or in the air. On land, quite sharp boundaries occur between the vegetation types on acidic and basic rocks where outcrops meet, or where serpentine (a term applied to a mineral rich in magnesium silicate) and non-serpentine rocks are juxtaposed. However, even in such situations, minerals are leached across the boundaries, which become increasingly blurred. The safest statement we can make about community boundaries is probably that they do not exist, but that some communities are much more sharply defined than others.

the community: not so much a superorganism ...

In the early years of ecology as a science, there was considerable debate about the nature of the community. Clements (1916) conceived of the community as a sort of *superorganism* whose member species were tightly bound together both now and in their common evolutionary history. Thus, individuals, populations and communities bore a relationship to each other resembling that between cells, tissues and organisms. In contrast, the *individualistic* concept devised by Gleason (1926) and others saw the relationship of coexisting species as simply the results of similarities in their requirements and tolerances, and partly the result of chance (or 'drift' as highlighted in Figure 18.1b). Taking this view, community boundaries need not be sharp, and associations of species would be much less predictable than one would expect from the superorganism concept. The current view is close to the individualistic concept. Results of direct gradient analysis and ordination all indicate that a given location, by virtue mainly of its physical characteristics, possesses a reasonably predictable association of species. However, a given species that occurs in one predictable association is also quite likely to occur with another group of species under different conditions elsewhere.

... more a level of organisation

Whether or not communities have more or less clear boundaries is an important question, but it is not the fundamental consideration. Community ecology is the study of the *community level of organisation* and it is not necessary to have discrete boundaries between communities to study community ecology.

18.4 Community patterns in time

Just as the relative importance of species varies in space, so their patterns of abundance may change with time. In either case, a species can occur only where and when: (i) it is capable of reaching a location; (ii) appropriate conditions and resources exist there; and (iii) competitors, predators and parasites do not preclude it. A temporal sequence in the appearance and disappearance of species therefore seems to require that the probability of arrival (dispersal) and/or conditions, resources and the influence of enemies themselves vary with time.

For many organisms, and particularly short-lived ones, their relative importance in the community changes with time of year as the individuals act out their life cycles against a background of seasonal change. Sometimes community composition shifts because of externally driven physical change, such as the build up of silt in a coastal salt marsh leading to its replacement by forest. In other cases, temporal patterns are simply a reflection of changes in key resources, as in the sequence of heterotrophic organisms associated with faecal deposits or dead bodies as they decompose (see Figure 11.2). The explanation for such temporal patterns is relatively straightforward and will not concern us here. Nor will we dwell on the variations in abundance of species in a community from year to year as individual populations respond to a multitude of factors that influence their reproduction and survival (dealt with in Chapters 5, 6 and 8–14).

Our focus will be on patterns of community change that follow a disturbance, defined as a relatively discrete event that removes organisms (Townsend & Hildrew, 1994) or otherwise disrupts the community by influencing the availability of space or food resources, or by changing the physical environment (Pickett & White, 1985). Such disturbances are common in all kinds of community. In forests, they may be caused by high winds, lightning, earthquakes, elephants, lumberjacks or simply by the death of a tree through disease or old age. Agents of disturbance in grassland include frost, burrowing animals and the teeth, feet, dung or dead bodies of grazers. On rocky shores or coral reefs, disturbances may result from severe wave action during hurricanes, tidal waves, battering by logs or moored boats or the fins of careless scuba divers.

succession can occur because some potential colonisers are competitively dominant

Accepting that some species are competitively superior to others, an initial coloniser of an opening left by a disturbance cannot necessarily maintain its presence there. In such a *dominance-controlled* community (Yodzis, 1986), disturbances lead to reasonably predictable sequences of species because different species have different strategies for exploiting resources – early species are good colonisers and fast growers, whereas later species can tolerate lower resource levels and grow to maturity in the presence of early species, eventually outcompeting them. These situations are more commonly known by the term *ecological succession*, defined as the *non-seasonal, directional and continuous pattern of colonisation and extinction on a site by species populations*. We shall see, however, that important as competitive dominance is, it is not the only driving force behind successions.

18.4.1 Primary and secondary successions

primary succession: an exposed landform uninfluenced by a previous community

Our focus is on successional patterns that occur on newly exposed landforms. If the exposed landform has not previously been influenced by a community, the sequence of species is referred to as a primary succession. Lava flows and pumice plains caused by volcanic eruptions (see Section 18.4.2), craters caused by the impact of meteors (Cockell & Lee, 2002), substrate exposed by the retreat of a glacier (Crocker & Major, 1955) and freshly formed sand dunes (see Section 18.4.3) are examples. In cases where the vegetation of an area has been partially or completely removed, but where well-developed soil and seeds and spores remain, the subsequent sequence of species is termed a secondary succession. The loss of trees locally as a result of disease, high winds, fire or felling may lead to secondary successions, as can cultivation followed by the abandonment of farmland (so-called old-field successions – see Section 18.4.4).

secondary succession: vestiges of a previous community are still present

Successions on newly exposed landforms, such as lava flows and sand dunes, typically take several hundreds of years to run their course. However, a precisely analogous process occurs amongst the animals and algae on recently denuded rock walls in the marine subtidal zone, and this succession takes only a decade or so (Hill *et al*., 2002). The research life of an ecologist is sufficient to encompass a subtidal succession but not that following glacial retreat. Fortunately, however, information can sometimes be gained over the longer timescale. Often, successional stages in time are represented by community gradients in space. The use of historic maps, carbon dating or other techniques may enable the age of a community since exposure of the landform to be estimated. A series of communities currently in existence, but corresponding to different lengths of time since the onset of succession, sometimes referred to as a *chronosequence*, can be inferred to reflect succession. However, whether or not different communities that are spread out in space really do represent various stages of succession must be judged with caution (Johnson & Miyanishi, 2008). We must remember, for example, that in northern temperate areas the vegetation we see may still be undergoing recolonisation and responding to climatic change following the last ice age (see Chapter 1).

18.4.2 Primary succession on volcanic lava

facilitation: early successional species on volcanic lava pave the way for later ones

A primary succession on basaltic volcanic flows on Miyake-jima Island, Japan, was inferred from a known chronosequence (16, 37, 125 and >800 years old) (Figure 18.10a). In the 16-year-old flow, soil was very sparse and lacking in nitrogen; vegetation was absent except for a few small alder trees (*Alnus sieboldiana*). In the older plots, 113 taxa were recorded, including ferns, herbaceous perennials, lianas and trees. There is evidence in this primary succession of roles for colonisation and competition but also facilitation. Thus, the species interactions discussed in Chapter 16 can figure prominently in successions. Of most significance were: (i) the successful colonisation of the bare lava by the nitrogen-fixing alder; (ii) the facilitation (through improved nitrogen availability) of mid-successional *Prunus speciosa* and the late-successional evergreen tree *Machilus thunbergii*; (iii) the formation of a mixed forest and the shading out of *Alnus* and *Prunus*; and (iv) finally, the replacement of *Machilus* by the longer lived *Castanopsis sieboldii* (Figure 18.10b).

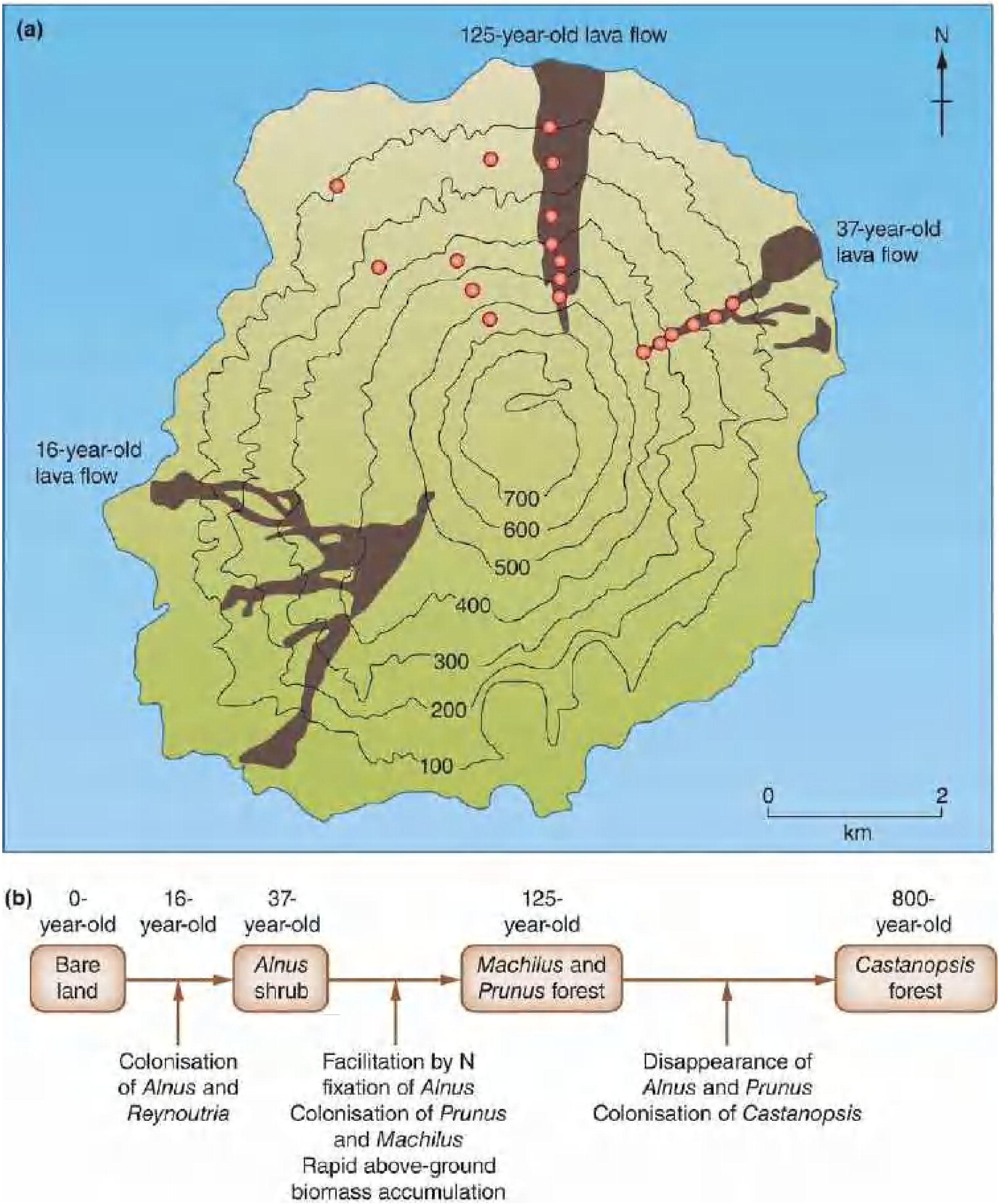


Figure 18.10 **Distribution of sampling sites on lava flows of different ages and the main features of the primary succession.** (a) Vegetation was described on 16-, 37- and 125-year-old lava flows on Miyake-jima Island, Japan. Analysis of the 16-year-old flow was non-quantitative (no sample sites shown). Sample sites on the other flows are shown as red circles. Sites outside the three flows are at least 800 years old. (b) The primary succession in relation to lava age.

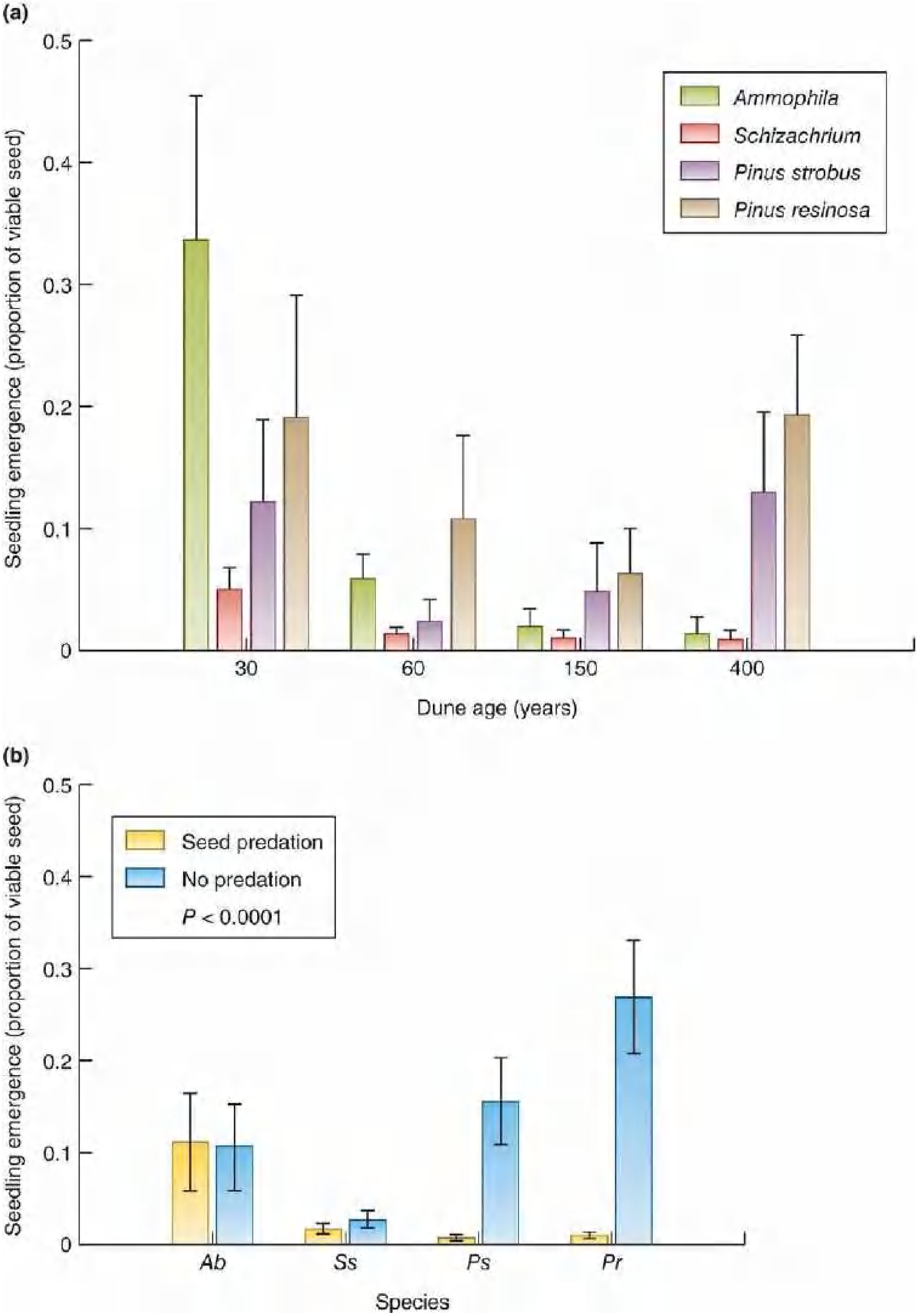
Source: After Kamijo *et al*. (2002).

18.4.3 Primary succession on coastal sand dunes

An extensive chronosequence of dune-capped beach ridges has been studied on the coast of Lake Michigan in the USA. Thirteen ridges of known age (30–440 years old) show a clear pattern of primary succession to forest (Lichter, 2000). The dune grass *Ammophila breviligulata* dominates the youngest, still-mobile dune ridge, but shrubby *Prunus pumila* and *Salix* spp. are also present. Within 100 years, these are replaced by evergreen shrubs such as *Juniperus communis* and by prairie bunch grass *Schizachyrium scoparium*. Conifers such as *Pinus* spp., *Larix laricina*, *Picea strobus* and *Thuja occidentalis* begin colonising the dune ridges after 150 years, and a mixed forest of *Pinus strobus* and *P. resinosa* develops between 225 and 400 years. Deciduous trees such as the oak *Quercus rubra* and the maple *Acer rubrum* do not become important components of the forest until 440 years.

importance of seed availability rather than facilitation in sand dune succession

It used to be thought that early successional dune species facilitated the later species by adding organic matter to the soil and increasing the availability of soil moisture and nitrogen (as in the volcanic primary succession). However, experimental seed addition and seedling transplant experiments have shown that later species are capable of germinating in young dunes (Figure 18.11a). While the more developed soil of older dunes may improve the performance of late successional species, their successful colonisation of young dunes is mainly constrained by limited seed dispersal, together with seed predation by rodents (Figure 18.11b). *Ammophila* generally colonises young, active dunes through horizontal vegetative growth. *Schizachrium*, one of the dominants of open dunes before forest development, has rates of germination and seedling establishment that are no better than *Pinus*, but its seeds are not preyed upon. Also, *Schizachyrium* has the advantage of quickly reaching maturity and can continue to provide seeds at a high rate. These early species are eventually competitively excluded as trees establish and grow. Lichter (2000) considers that dune succession in this case is better described in terms of the transient dynamics of colonisation and competitive displacement, rather than the result of facilitation by early species (improving soil conditions) followed by competitive displacement.



(a)
Seedling emergence (proportion of viable seed)
0.5
0.4
0.3
0.2
0.1
0
Ammophila
Schizachrium
Pinus strobus
Pinus resinosa
30
60
150
400
Dune age (years)
(b)
Seedling emergence (proportion of viable seed)
0.5
0.4
0.3
0.2
0.1
0
Seed predation
No predation
P < 0.0001
Ab
Ss
Ps
Pr
Species

Figure 18.11 Later successional species are capable of germinating in young dunes but are not well represented because of limited dispersal and seed predation. (a) Seedling emergence (means + SE) from added seeds of species typical of different successional stages on dunes of four ages. (b) Seedling emergence of the four species (*Ab*, *Ammophila breviligulata*; *Ss*, *Schizachrium scoparium*; *Ps*, *Pinus strobus*; *Pr*, *Pinus resinosa*) in the presence and absence of rodent predators of seeds.

Source: After Lichter (2000).

18.4.4 Secondary successions in abandoned fields

abandoned old fields: succession to forest in North America ...

Successions on old fields have been studied particularly along the eastern part of the USA where many farms were abandoned by farmers who moved west after the frontier was opened up in the 19th century (Tilman, 1987 , 1988). Most of the precolonial mixed conifer–hardwood forest had been destroyed, but regeneration was swift. In many places, a series of sites that were abandoned for different, recorded periods of time are available for study. The typical sequence of dominant vegetation is: annual weeds, herbaceous perennials, shrubs, early successional trees and late successional trees.

... but to grassland on the Island of Öland in Sweden

The cessation of arable cultivation on the Baltic Island of Öland has led, because of centuries of low-intensity grazing management, not to succession to forest but to semi-natural grassland. By studying grassland habitat patches that have existed continuously for 5–270+ years, Purschke *et al.* (2013) were able to compare patterns of taxonomic and functional diversity through the succession and to pinpoint the functional traits that played particularly prominent roles. Between early (5–15 years) and early-mid-successional stages (16–50 years), species richness increased but there was no significant increase in functional diversity (Figure 18.12a). In this phase of succession, it seems that the environmental and dispersal filters, highlighted in Figure 18.1a, were prominent and were selecting for sets of functionally similar species. These were typically capable of expanding rapidly within the community, with high values for seed production, leaf size and canopy height, and were more likely to have high potential for long-distance dispersal (e.g. by wind) or long-term persistence in the seed bank, as compared with later successional species (Figure 18.12c–f). In contrast to the early successional pattern, between late-mid (51–270 years) and late successional stages (more than 270 years of continuous grassland) there was no further increase in species richness, but functional diversity increased markedly (Figure 18.12b), indicating that functionally redundant species were replaced by species that were functionally more different to those that remained. This pattern is consistent with the idea that abiotic filtering dominates early in succession, while biotic filtering (e.g. competitive exclusion and resource partitioning; see Chapter 16) becomes prominent later. Such biotic filtering can be expected to select for functionally dissimilar species as the range of available niches broadens and they become more fully occupied.

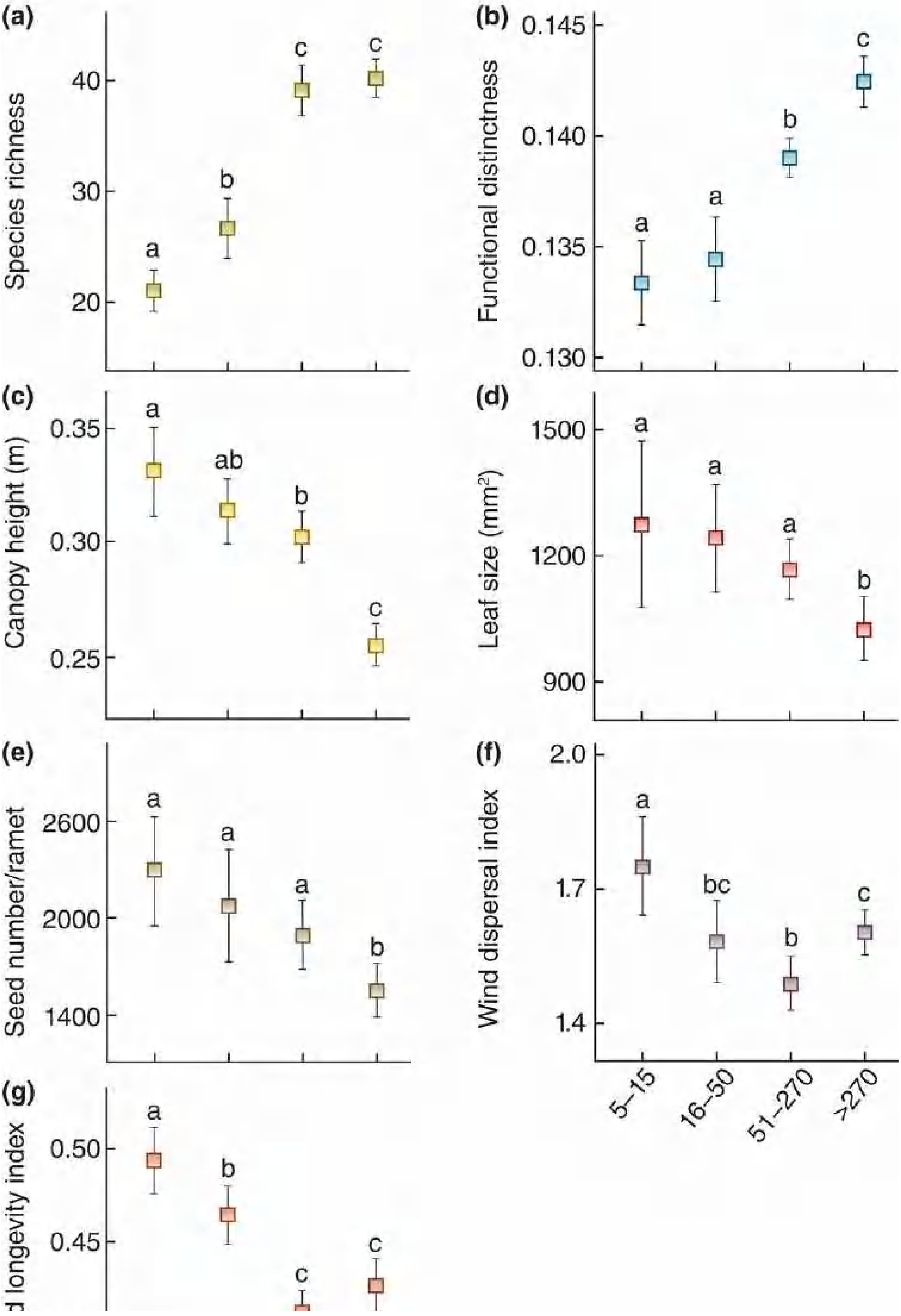

(a)
Species richness
20
30
40
a
b
c
c
(b)
Functional distinctness
0.130
0.135
0.140
0.145
a
a
b
c
(c)
Canopy height (m)
0.25
0.30
0.35
a
ab
b
c
(d)
Leaf size (mm²)
900
1200
1500
a
a
a
b
(e)
Seed number/ramet
1400
2000
2600
a
a
a
b
(f)
Wind dispersal index
1.4
1.7
2.0
a
bc
b
c
5–15
16–50
51–270
>270
(g)
d longevity index
0.45
0.50
a
b
c
c

Figure 18.12 **Successional patterns in an arable grassland chronosequence**. (a) Plant species richness and (b) functional diversity ('functional distinctness' assessed as the mean of functional trait distances between all species-pairs) of four successional stages in an arable–grassland chronosequence in Sweden (5–15, 16–50, 51–270 and >270 years of grassland continuity). (c–g) How the mean representation of selected functional traits changes during succession. Data points (means with SEs) with the same letter are not significantly different.

Source: After Purschke *et al*. (2013).

18.5 The mechanisms underlying succession

18.5.1 A species replacement model of succession

forest succession can be represented as a tree-by-tree replacement model

A model of succession developed by Horn (1981) sheds some light on the successional process. In his so-called Markov chain model, Horn recognised that in a hypothetical forest community it would be possible to predict changes in tree species composition simply by knowing for each tree species the probability that, within a particular time interval (say 50 years), an individual of any one species would be replaced by another of the same species or of a different species. More complex, individual-based models work along similar lines. For example, Holm *et al*. (2012) modelled 300 years of succession in a subtropical dry forest in Puerto Rico. Their model assumes the forest is made up of many small patches, some recently opened gaps and others at different successional stages, and follows the progress of each individual tree as it establishes from seed, grows (annual diameter increase), dies and regenerates. Making use of demographic databases for each species, establishment and growth are first simulated under optimal conditions and then moderated according to environmental constraints operating in each patch (light availability, soil moisture, etc.). Figure 18.13 shows how, starting from bare ground, the commonest 18 species (of a total of 37) change in relative importance with time, a pattern that is close to what happens in the real succession.

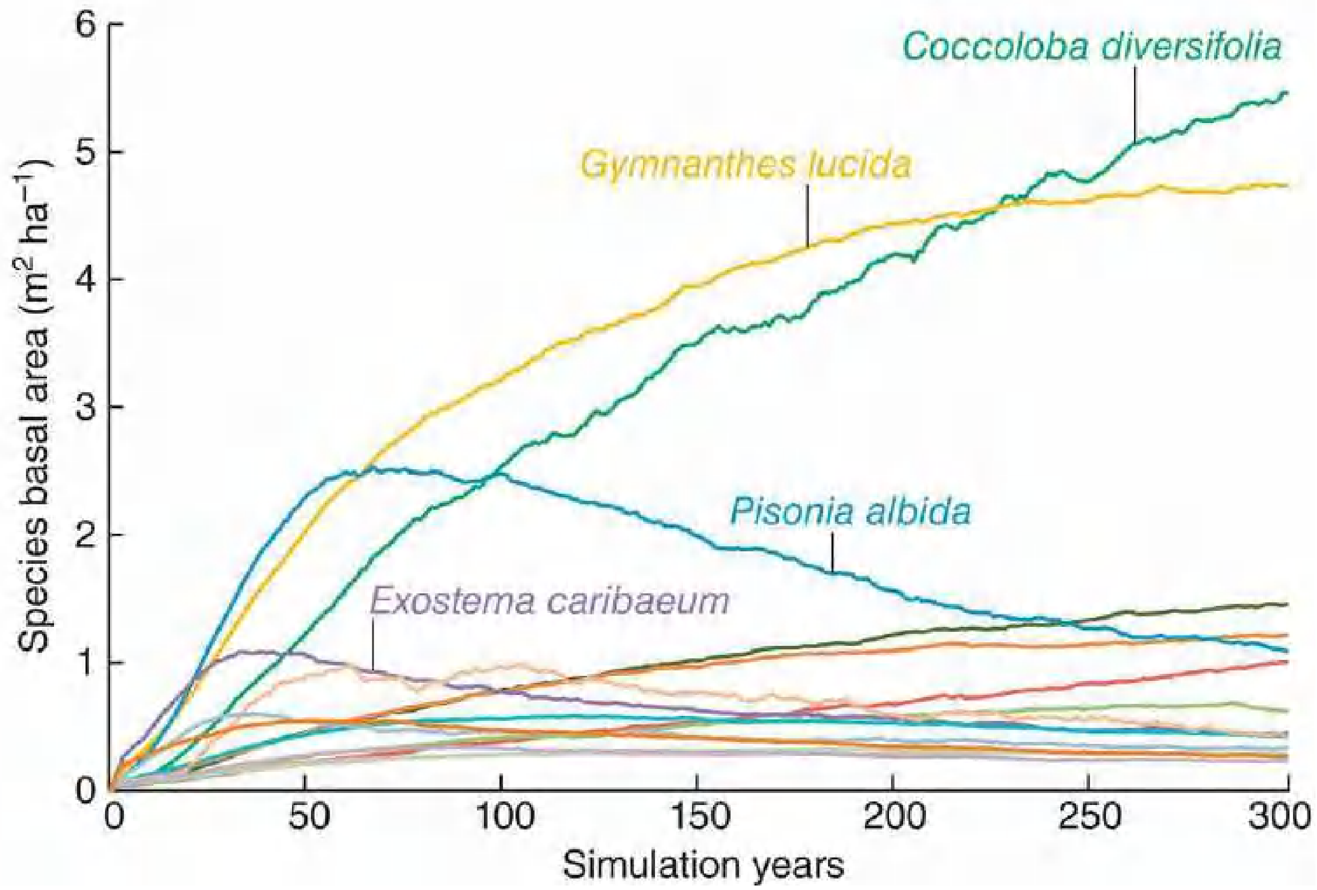


Figure 18.13 **Predicted changes in relative abundance of forest trees in Puerto Rico**. Relative abundance (assessed in terms of tree basal surface area per hectare of forest), starting from bare ground. Early successional species include *Pisonia albida* and *Exostema caribaeum*, while the late successional *Gymnanthes lucida* and *Coccoloba diversifolia* dominate later.

Source: From Holm *et al.* (2012).

18.5.2 A trade-off between competition and colonisation

Despite the advantages of simple Markov models, a theory of succession should ideally not only predict but also explain. To do this, we need to consider the *biological* basis for the replacement values in the model, and here we have to turn to alternative approaches.

competition–colonisation trade-off

Rees *et al.* (2001) drew together a diversity of experimental, comparative and theoretical approaches to produce some generalisations about vegetation dynamics. Early successional plants have a series of correlated traits, including high fecundity, effective dispersal, rapid growth when resources are abundant, and poor growth and survival when resources are scarce. Late successional species usually have the opposite traits, including an ability to grow, survive and compete when resources are scarce. In the absence of disturbance, late successional species eventually outcompete early species, because they reduce resources beneath the levels required by the early successional species. Early species are successful if their dispersal ability and high fecundity permits them to colonise and establish in recently disturbed sites before late successional species can arrive. Rees and his colleagues refer to this as a *competition–colonisation trade-off*. The idea is strengthened by a further physiological inevitability. Huge differences in per capita seed production among plant species are inversely correlated to equally large variations in seed size; plants producing tiny seeds tend to produce many more of them than plants producing large seeds (see Section 7.3.3). Thus, Rees *et al.* (2001) point out that small-seeded species are good colonists (many propagules) but poor competitors (small seed food reserves), and vice versa for large-seeded species. This echoes an older, evolutionary perspective (Harper, 1977) in which either (i) a species reacts to the competitive selection pressures and evolves characteristics that enable it to persist longer in the succession, i.e. it responds to *K*-selection; or (ii) it may develop more efficient mechanisms of escape from the succession, and discover and colonise suitable early stages of succession elsewhere, i.e. it responds to *r*-selection (see Section 7.6.1). Thus, from an evolutionary point of view, good colonisers can be expected to be poor competitors and vice versa.

18.5.3 Successional niche models

the importance of resource availability – the successional niche

Rees *et al.* (2001) also noted that a successional sequence could occur because rapid growth under resource-rich conditions allows early species to temporarily outcompete late successional species even if they arrive at the same time. They referred to this second possibility as the *successional niche* (early conditions suit early species because of their niche requirements). An example is provided by the relations of three tree species that appear in sequence in successions in North America. Trembling aspen (*Populus tremuloides*) appears earlier than northern red oak (*Quercus rubra*) or sugar maple (*Acer saccharum*). Kaelke *et al.* (2001) compared the growth of seedlings of all three species when planted along a gradient of light availability. The aspen outgrew the others when relative light availability exceeded 5%. However, there was a rank reversal in relative growth rate in deep shade; here the oak and maple, typical of later stages of succession, grew more strongly and survived better than aspen (Figure 18.14). In other words, there are clear links between changing abiotic conditions and the outcome of species interactions (see Chapter 16).

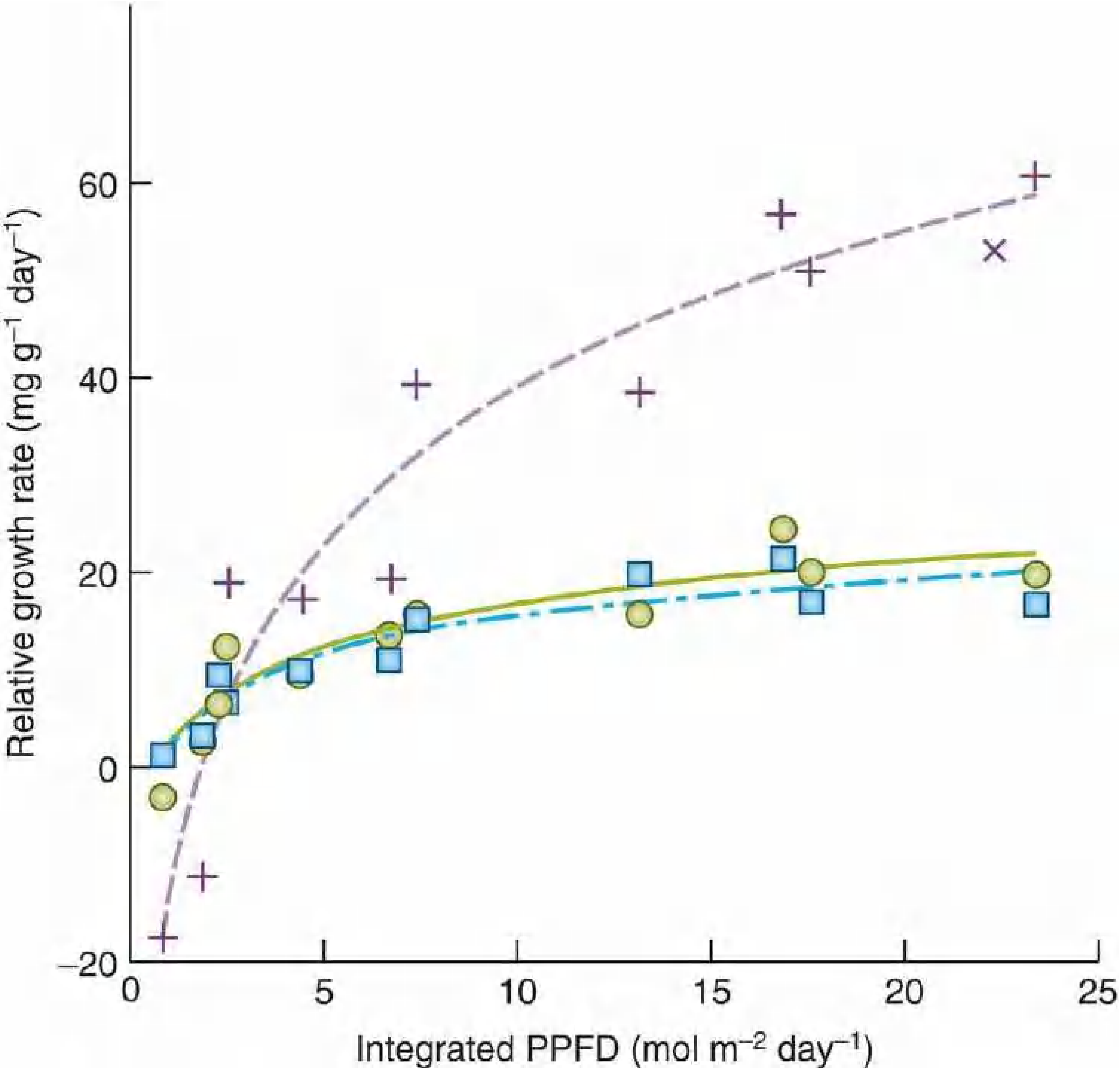


Figure 18.14 Example of a successional niche – early conditions suit early species because of their niche requirements. Relative growth rate (during the July–August 1994 growing season) of trembling aspen (purple cross), northern red oak (green circle) and sugar maple (blue square) in relation to photosynthetic photon flux density (PPFD).

Source: After Kaelke *et al.* (2001).

18.5.4 Facilitation

the importance of facilitation – but not always

Cases of competition–colonisation trade-offs and/or successional niche relations are prominent in virtually every succession that has been described. In addition, we have seen cases where early species may change the abiotic environment in ways (e.g. increased soil nitrogen) that make it easier for later species to establish and thrive. Thus, *facilitation* (see Chapter 13) has to be added to the list of biological phenomena underlying some successions. Examples of facilitation are widespread (Bonanomi *et al.*, 2011) but the converse is not uncommon; thus, many plant species alter the environment in a way that makes it more, rather than less, suitable for themselves

(Wilson & Agnew, 1992). For example, woody vegetation can trap water from fog or ameliorate frosts, improving the conditions for growth of the species concerned, whilst grassy swards can intercept surface flowing water and grow better in the moister soil that is created.

18.5.5 The role of animals

the importance of interaction with enemies

It follows from the idea of a competition–colonisation trade-off that recruitment of competitively dominant plants should be determined largely by the rate of arrival of their seeds. This means that herbivores that reduce seed production are more likely to reduce the density of dominant competitors than of subordinates. Recall that this is just what happened in the sand-dune study described in Section 18.4.3. In a similar vein, Carson and Root (1999) showed that by removing insect predators of seeds, the meadow goldenrod (*Solidago altissima*), which normally appears about five years into an old-field succession, became dominant after only three years. This happened because release from seed predation allowed it to outcompete earlier colonists more quickly. Thus, apart from competition–colonisation trade-off, successional niche and facilitation, we have to add a fourth mechanism – interactions with enemies – if we are to fully understand plant successions.

the place of animals in successions .... and microorganisms

The fact that plants dominate most of the structure and succession of communities does not mean that animals always follow the communities that plants dictate. This will often be the case, of course, because the plants provide the starting point for food webs (see Chapter 17) and determine much of the character of the physical environment in which animals live. But it is also sometimes the animals that determine the nature of the plant community, as we have seen for seed predators. Another example comes from neotropical forests in Peru and Venezuela where frugivorous bats play critical roles in the dispersal of early successional trees and in forest regeneration in general (Muscarella & Fleming, 2007).

More often though, animals, such as passerine bird species, are passive followers of successions amongst the plants. Arbuscular mycorrhizal fungi (see Section 13.9), which show clear changes in species composition in the soils associated with an old-field succession (Fitzsimons *et al.*, 2008), may also be passive followers of the plants. But this does not mean that the birds, which eat and disperse seeds, or the fungi, which affect plant growth and survival, do not influence the succession in its course. They probably do.

18.5.6 The role of functional traits

beyond just competitive ability: Noble and Slatyer’s ‘vital attributes’

We have already seen that particular functional traits can play a prominent role in determining the course of a succession (e.g. Figure 18.12). Noble and Slatyer (1981) called such properties *vital attributes*. For plants, the two most important relate to: (i) the method of recovery after disturbance (four classes are defined: vegetative spread, V; seedling pulse from a seed bank, S; seedling pulse from abundant dispersal from the surrounding area, D; no special mechanism with just moderate dispersal from only a small seed bank, N); and (ii) the ability of individuals to reproduce in the face of competition (defined in terms of tolerance T at one extreme and intolerance I at the other). Thus, for example, a species may be classed as SI if disturbance releases a seedling pulse from a seed bank but the plants are intolerant of competition. Seedlings of such a species could establish themselves only immediately after a disturbance, when competitors are rare. Of course, a seedling pulse fits well with such a pioneer existence. An example is the annual *Ambrosia artemisiifolia*, which often figures early in old-field successions. In contrast, the American beech (*Fagus grandifolia*) could be classed as VT (being able to regenerate vegetatively from root stumps, and tolerant of competition since it is able to establish itself and reproduce in competition with older or more advanced individuals of either its own or another species) or NT (if no stumps remain, it would invade slowly via seed dispersal). In either case, it would eventually displace other species but not itself be displaced. Noble and Slatyer argue that all the species in an area can be classified according to these two vital attributes (to which relative longevity might be added as a third). Given this information, quite precise predictions about successional sequences become possible.

Lightning-induced fires produce regular and natural disturbances in many ecosystems in arid parts of the world (refer back to Gosper *et al.* (2012) and Figure 18.9c, d) and two fire-response syndromes, analogous to two of Noble and Slatyer's disturbance recovery classes, can be identified. Re-sprouters have massive, deeply penetrating root systems, and survive fires as individuals, whereas re-seeders are killed by the fire but re-establish through heat-stimulated germination and growth of seedlings (Bell, 2001). The proportion of species that can be classified as re-sprouters is higher in forest and shrubland vegetation of south-west Western Australia (Mediterranean-type climate) than in more arid areas of the continent. Bell suggests that this is because the Western Australian communities have been subject to more frequent fires than other areas, conforming to the hypothesis that short intervals between fires (averaging 20 years or less in many areas of Western Australia) promote the success of re-sprouters. Longer intervals between fires, on the other hand, allow fuel loads to build up so that fires are more intense, killing re-sprouters and favouring the re-seeding strategy.

18.5.7 The nature of the climax

do successions reach a climax?

Do successions come to an end? It is clear that a stable equilibrium will occur if individuals that die are replaced on a one-to-one basis by young of the same species. At a slightly more complex level, Markov models (see earlier in this chapter) tell us that a stationary species composition should, in theory, occur whenever the replacement probabilities (of one species by itself or by any one of several others) remain constant through time.

A community that emerges, and persists, when a succession ends is known as a 'climax' community. The concept of the climax has a long history. One of the earliest students of succession, Frederic Clements (1916), is associated with the idea that a single climax will dominate in any given climatic region, being the end point of all successions, whether they happened to start from a sand dune, an abandoned old field or even a pond filling in and progressing towards a terrestrial climax. This *monoclimax* view was challenged by many ecologists, amongst whom Tansley (1939) was prominent. The *polyclimax* school recognised that a local climax may be governed by one factor or a combination of factors: climate, soil conditions, topography, fire and so on. Thus, a single climatic area could easily contain a number of specific climax types. Later still, Whittaker (1953) proposed his climax pattern hypothesis. This conceives a continuity of climax types, varying gradually along environmental gradients and not necessarily separable into discrete climaxes. (This is an extension of Whittaker's approach to gradient analysis of vegetation, discussed in Section 18.3.1.)

In fact, it is very difficult to identify a stable climax community in nature. Usually, we can do no more than note that the rate of change of succession slows down to the point where any change is imperceptible to us. Convergence to a climax may take only a few years in rocky intertidal communities. Old-field successions, on the other hand, might take 100–500 years to reach a ‘climax’, but in that time the probabilities of further fires or hurricanes are so high that a process of succession may rarely go to completion. If we bear in mind that forest communities in northern temperate regions, and probably also in the tropics, are still recovering from the last glaciation (see Chapter 1), it is questionable whether the idealised climax vegetation is often reached in nature.

APPLICATION 18.2 The application of succession theory to restoration

Given the widespread occurrence of successional processes, and the realisation that few if any communities are static over time, it is not surprising that succession theory has important applications in the fields of restoration of derelict land and vegetation management (Pickett *et al.*, 2008). In effect, managers can seek to circumvent one or more of the community assembly filters highlighted in Figure 18.1a, by relaxing dispersal constraints, environmental constraints or biotic constraints (involving facilitation, competition, predation, etc.).

Invoking the theory of competition–colonisation trade-offs

In many successions, the later stages are later, in part, because the more competitive late successional species are slower to colonise than their early successional counterparts – knowledge that can be applied by managers to speed up restoration projects.

Vast areas of tropical forest, like their temperate counterparts, have been replaced by agriculture. But when land is retired from production in the tropics, the early and mid-successional species tend to dominate succession for a century or more. The nearest late successional trees occur in distant forest fragments. Because their large seeds depend on birds, bats or primates for dispersal, they are in short supply, and of the few that arrive after succession has started, most die as seeds or seedlings. Martinez-Garza and Howe (2003) found that in Panama and Costa Rica when large-seeded late successional trees such as *Dipteryx panamensis* and *Genipa americana* are planted by hand in pastures, before succession has really started, they have a much higher probability of survival. Thus, the competition–colonisation trade-off can be circumvented by removing dispersal constraints of late successional species, bypassing up to 70 years of succession. In a similar vein, seed dispersal of bird-dispersed trees in early successional temperate forest in Chile was enhanced dramatically simply by the installation of artificial perches (Bustamante-Sanchez & Armesto, 2012).

Invoking successional niche theory

If species are restricted to a particular stage in succession because their niche requirements are met only at that time, successful restoration can be expected to depend on re-establishing appropriate conditions and resources or, in other words, relaxing environmental constraints.

Sand mining is a disruptive process and, when it ceases, the substrate is deficient not only in seeds but also in soil organic matter and plant nutrients. An additional problem in the subtropical Bongil Peninsula in New South Wales, Australia, has been invasion of recovering land by the aggressive perennial grass *Imperata cylindrica*. Cummings *et al.* (2005) performed experiments to identify the barriers limiting restoration of the native community of these invader-dominated, sand-mined sites. One barrier might have been poor establishment of native seedlings because of competition by *I. cylindrica*. However, competition was shown not to be the critical factor, in that burning of the grassland and weed control, coupled with native seed planting, failed to result in the regeneration of native woody cover. But when soil high in organic matter was added in a second experiment, this significantly improved the survival and growth of planted native species. Thus, the succession moved forward when appropriate niche conditions were created.

Invoking facilitation theory

The success of a later successional species can also be promoted by an earlier species that enhances local abiotic conditions or improves the supply of plant nutrients. Most forests in the Mediterranean area of Europe have disappeared after millennia of agricultural development and urbanisation. Restoration in such degraded shrub-dominated habitats has traditionally started with shrub removal, in the belief that these compete with newly planted tree seedlings. But all plants in Mediterranean environments are prone to suffer from high temperatures and low water availability, and Gómez-Aparicio *et al.* (2004) argued that in these harsh conditions the pioneer shrubs might well have a positive effect on the establishment of seedlings of mid-successional woody shrubs and trees, acting as 'nurse plants'. Thus, in a series of experiments in the Sierra Nevada area of south-east Spain, Gómez-Aparicio and his colleagues determined the average effects of nurse shrubs for different categories of target woody plants, different categories of nurse plants and different abiotic conditions (Figure 18.15).

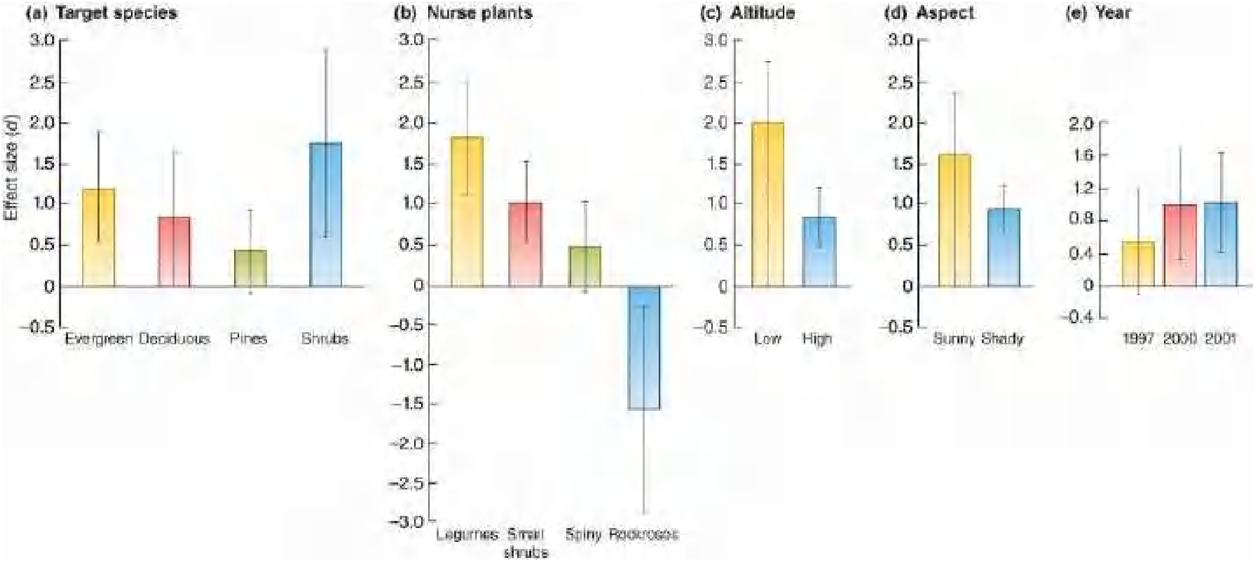


Figure 18.15 Nurse plants facilitate seedling survival in forest restoration. The mean effect size (*d* ± 95% confidence limits) of the presence of 'nurse' plants for the survival of seedlings of target Mediterranean forest species. Effect size is the difference in survival probability after 1 year in the presence and absence of nurse plants divided by the mean survival probability in their absence. Thus positive values indicate facilitation and larger positive values indicate a stronger facilitative effect. (a) All target species showed evidence of facilitation by potential nurse plants, of which (b) legumes proved the most effective (while rockroses turned out not to be 'nurse' plants at all). (c–e) The effectiveness of 'nurse' plants also varied with abiotic conditions.

Source: After Gómez-Aparicio *et al.* (2004).

Shrubs promoted the survival of seedlings of evergreen and deciduous trees like oak (*Quercus* spp.) and maple (*Acer opalus*), reflecting their late successional status and requirement for shade. However, mid-successional woody shrubs also benefited. In contrast, the shade-intolerant montane pines were not significantly facilitated (Figure 18.15a). Leguminous nurse plants (with their root nodules containing nitrogen-fixing bacteria) enhanced soil nutrient status in these nitrogen-limited soils and provided the strongest facilitation for woody species. But legumes, and most other nurse shrubs, no doubt also benefited seedlings by providing shade from the intense Mediterranean summer sun. Rockroses (*Cistus* spp.) turned out not to be nurse plants at all, reflecting their production of 'allelopathic' chemicals, reducing the success of neighbouring plants (Figure 18.15b). The advantage conferred by nurse plants was higher at low altitudes and on sunny slopes (Figure 18.15c, d), both situations where lower rainfall and higher temperatures cause intense summer droughts. Finally, facilitation was least noticeable in 1997, an uncharacteristically wet year in which seedlings survived well even in the absence of nurse plants (Figure 18.15e).

When pioneer species are facilitators of successional change, the appropriate management action is to leave them in place.

Invoking enemy interaction theory

Herbivores can be expected to speed up succession if they preferentially feed on pioneer species or to slow it down if their greatest negative influence is on later species. Thus, Lai and Wong (2005) improved the growth and survival of seedlings of thick-leaved oak (*Cyclobalanopsis edithiae*), a late successional species of the original primary forest of Hong Kong, by installing tree guards (plastic sleeves) to prevent access to browsing mammals (Figure 18.16). The installation of weed mats to reduce competition with pioneer plants also improved survivorship, but to a lesser extent than when combined with tree guards. From a cost–benefit point of view, the total cost per surviving seedling was reduced from US$6.76 in the control treatment (allowing access to browsing animals) to US$4.05, despite the cost of providing the tree guard (and weed mat).

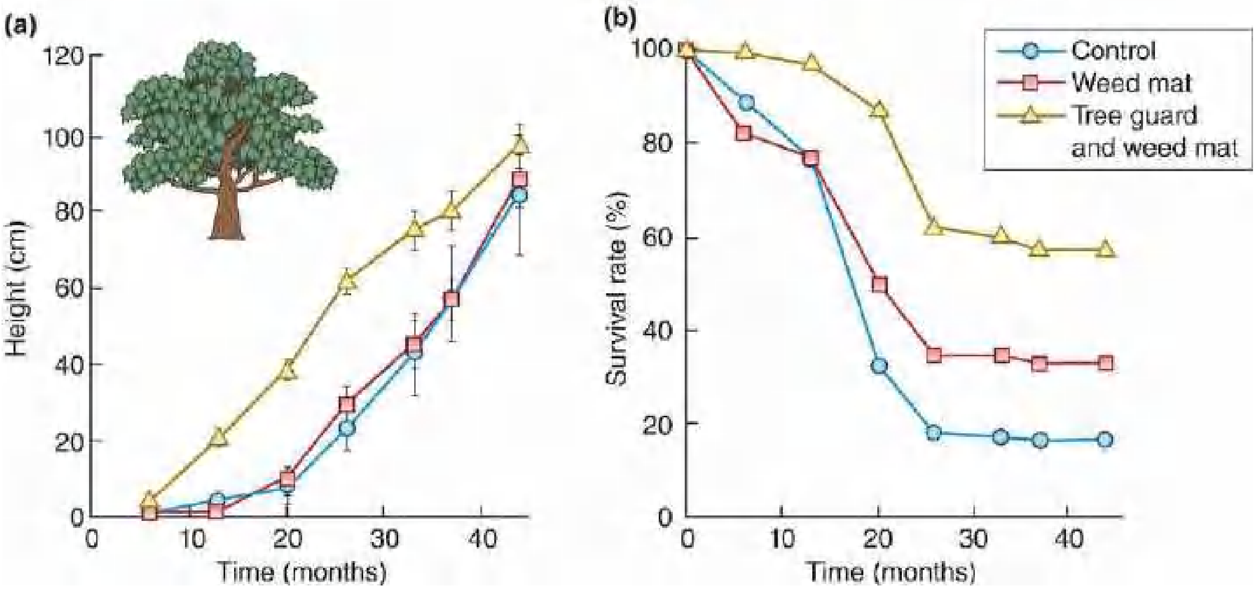


Figure 18.16 Protection from browsing mammals by a tree guard increased thick-leaved oak seedling growth and survivorship in restored Hong Kong forest. A hessian weed mat, to reduce competition with pioneer plants, improved (a) oak growth (means + SE) and (b) survivorship but to a lesser extent than when combined with a tree guard.

Source: After Lai & Wong (2005).

Managing succession to restore a cultural tradition

For its use in basketry, members of the Mohawk Nation of North America greatly value sweetgrass (*Anthoxanthum nitens*), a mid-successional species that once grew along the Mohawk River valley in New York State near the Kanatsiohareke farming community. The vegetation here is now dominated by exotic grasses, and the nearest gathering area for sweetgrass is more than 300 km away.

Shebitz and Kimmerer (2005) used four experimental treatments to assess the restoration potential of sweetgrass: (i) sweetgrass alone (weeded to remove competition); (ii) together with existing old-field vegetation; (iii) weeded and planted with hairy vetch (*Vicia villosa*) as a potential nurse plant; and (iv) weeded and planted with ryegrass (*Lolium multiflorum*) as an alternative nurse plant. The nitrogen-fixing ability of hairy vetch, an annual legume, might be expected to enhance sweetgrass growth, while the annual ryegrass has potentially beneficial weed-suppressing properties. Both species establish readily, but ryegrass has a tendency to persist through succession with possible negative consequences for sweetgrass and local biodiversity generally. Hairy vetch does not persist long in succession.

Sweetgrass biomass, height, reproduction rate and survivorship were greatest in plots that were weeded to eliminate competition and where hairy vetch was present: note that high sweetgrass abundance and tall blades are particularly desirable to the basketmakers. Ryegrass, in contrast, was not a nurse crop, actually reducing sweetgrass growth and reproduction. Hairy vetch can thus be used to bring back a valued cultural tradition by restoring the native sweetgrass into its place in old-field succession at Kanatsiohareke.

18.6 Communities in a spatiotemporal context

the idea of a successional mosaic

A forest, or a rangeland, or a rocky shore community that appears to have reached a stable community structure when studied at a large scale, will almost always be a mosaic of miniature successions. Every time a tree falls or a grass tussock or seaweed dies, an opening is created in which a new succession starts. One of the most seminal papers in the history of ecology was entitled ‘Pattern and process in the plant community’ (Watt, 1947). A crucial part of the pattern of a community is caused by the dynamic processes of deaths, replacements and microsuccessions that the broad view may conceal. Thus, although we can point to patterns in community composition in space (Section 18.3) and in time (Sections 18.4 and 18.5), it is often more meaningful to consider space and time together.

α-, β- and γ-species richness

When dealing with a successional mosaic, or any landscape of habitat patches containing different sets of species, we need to separate out the total species richness of the region, referred to as γ-richness, the average species richness *within* patches (α-richness) and the between-patch component of regional richness (β-richness), as we saw in Section 18.2.1. To reiterate, γ-richness is the sum of α-richness and β-richness and we noted that if every patch has identical species lists, β-richness is zero and γ-richness equals α-richness. However, β-richness contributes to γ-richness wherever there is heterogeneity in the distribution of species among patches.

18.6.1 Disturbance, gaps and dispersal

We have already seen that disturbances that open up gaps are common in all kinds of communities. The formation of gaps is obviously of considerable significance to sessile or sedentary species that have a requirement for open space, but gaps have also proved to be important for mobile species such as invertebrates on the beds of streams (Matthaei & Townsend, 2000). A single patch without migration is, by definition, a closed system, and any extinction caused by disturbance would be final. However, extinction within a patch in an open system is not necessarily the end of the story because of the possibility of reinvasion from other patches.

Fundamental to this perspective is recognition of the importance of migration between habitat patches. This may involve adult individuals, but very often the process of most significance is the dispersal of immature propagules (seeds, spores, larvae) and their recruitment to populations within habitat patches. The order of arrival and relative recruitment levels of individual species may determine or modify the nature and outcome of population interactions in the community.

This is another case where we must acknowledge that a plurality of processes is usually at work in determining community structure (see Section 16.4).

disturbance scale and phasing

Some disturbances are synchronised, or phased, over extensive areas. A forest fire may destroy a large tract of a mature forest community. The whole area then proceeds through a more or less synchronous succession, with diversity increasing through the early colonisation phase and falling again through competitive exclusion as the mature stage is approached (β-diversity is zero). Other disturbances are much smaller and produce a patchwork of habitats. If these disturbances are unphased, the resulting community comprises a mosaic of patches of different ages. This mosaic is much richer in species (both β-diversity and α-diversity contribute to γ-diversity) than an extensive area undisturbed for a very long period and occupied by just one or a few late successional species. Confirming this, Towne (2000) monitored the plant species that established in prairie grassland where large ungulates had died (mainly bison, *Bos bison*). Scavengers remove most of the body tissue, but copious amounts of body fluids and decomposition products seep into the soil. The flush of nutrients combined with death of the previous vegetation produces a competitor-free, disturbed area where resources are unusually abundant. The patches are also exceptional because the soil has not been disturbed (as it would be after a ploughed field is abandoned or a badger makes a burrow); thus, the colonising plants do not derive from the local seed bank. The unusual nature of the disturbed patches means that many of the pioneer species are rare in the prairie as a whole, and carcass sites contribute to species diversity and community heterogeneity for many years.

18.6.2 The frequency of gap formation

The influence that disturbances have on a community depends strongly on the frequency with which gaps are opened up. One influential idea, the *intermediate disturbance hypothesis* (Connell, 1978 ; see also the earlier account by Horn, 1975), proposed that the highest diversity is maintained at intermediate levels of disturbance. The argument runs that soon after a severe disturbance, propagules of a few pioneer species arrive in the open space, and if further disturbances occur frequently most gaps will not progress beyond the pioneer stage and the diversity of the community as a whole will be low. As the interval between disturbances increases, however, the successional state of different patches will be more varied, and many will be in mid-succession, where diversity is often high. Hence, the overall diversity will itself be high. But then at very low frequencies of disturbance, most of the community for most of the time will reach and remain in a late successional state, with competitive exclusion having reduced diversity.

boulders on a rocky shore that vary in disturbability ...

The influence of the frequency of gap formation was the subject of a classic study by Sousa (1979a , 1979b) in an intertidal algal community associated with boulders of various sizes in southern California. Wave action disturbs small boulders more often than large ones. A class of mainly small boulders (which required a force of less than 49 Newtons to move them) had a monthly probability of movement of 42%. An intermediate class (which required a force of 50–294 N) had a much smaller monthly probability of movement, 9%. Finally, the class of mainly large boulders (which required a force >294 N) moved with a probability of only 0.1% per month.

Communities on boulders in each of the three size/disturbability classes were assessed on four occasions. Table 18.1 shows that the percentage of bare space decreased from the small to large boulder classes, indicating the effects of the greater frequency of disturbance on the former. Mean species richness was lowest on the regularly disturbed small boulders, which tended to be dominated by the rapidly colonising green alga *Ulva* spp. and barnacles (*Chthamalus fissus*). The highest levels of species richness were consistently recorded on the intermediate boulder class where both rapid colonisers and strong competitors occurred together. The largest boulders had lower mean species richness than the intermediate class, because of competitive dominance by the brown alga *Gigartina canaliculata*, although a complete monoculture was achieved on only a few boulders.

Table 18.1 Seasonal patterns in bare space and species richness on boulders in each of three classes. The classes of boulders are categorised according to the force (in Newtons) required to move them. *Source*: After Sousa (1979b).

			Species richness		
Census date	**Boulder class (N)**	**Percentage bare space**	**Mean**	**Standard error**	**Range**
November 1975	<49	78.0	1.7	0.18	1–4
	50–294	26.5	3.7	0.28	2–7
	>294	11.4	2.5	0.25	1–6
May 1976	<49	66.5	1.9	0.19	1–5
	50–294	35.9	4.3	0.34	2–6
	>294	4.7	3.5	0.26	1–6
October 1976	<49	67.7	1.9	0.14	1–4
	50–294	32.2	3.4	0.40	2–7
	>294	14.5	2.3	0.18	1–6
May 1977	<49	49.9	1.4	0.16	1–4
	50–294	34.2	3.6	0.20	2–5
	>294	6.1	3.2	0.21	1–5

... provide support for the intermediate disturbance hypothesis

This study deals with a single community conveniently composed of identifiable patches (boulders) that become gaps (when overturned by waves) at short, intermediate or long intervals. Recolonisation occurs mainly from propagules derived from other patches in the community. Because of the pattern of disturbance, this mixed boulder community is more diverse than would be one with only large boulders.

further support from a study of streams....

Disturbances in small streams often take the form of bed movements during periods of high discharge. Because of differences in flow regimes and in the substrates of stream beds, some stream communities are disturbed more frequently and to a larger extent than others. This variation was assessed in 54 stream sites in the Taieri River in New Zealand (Townsend *et al.*, 1997) by recording the average percentage of the stream bed that moved between sampling occasions (assessed on five occasions during one year, using painted particles of sizes characteristic of the stream bed in question). The pattern of richness of insect species conformed to the intermediate disturbance hypothesis (Figure 18.17).

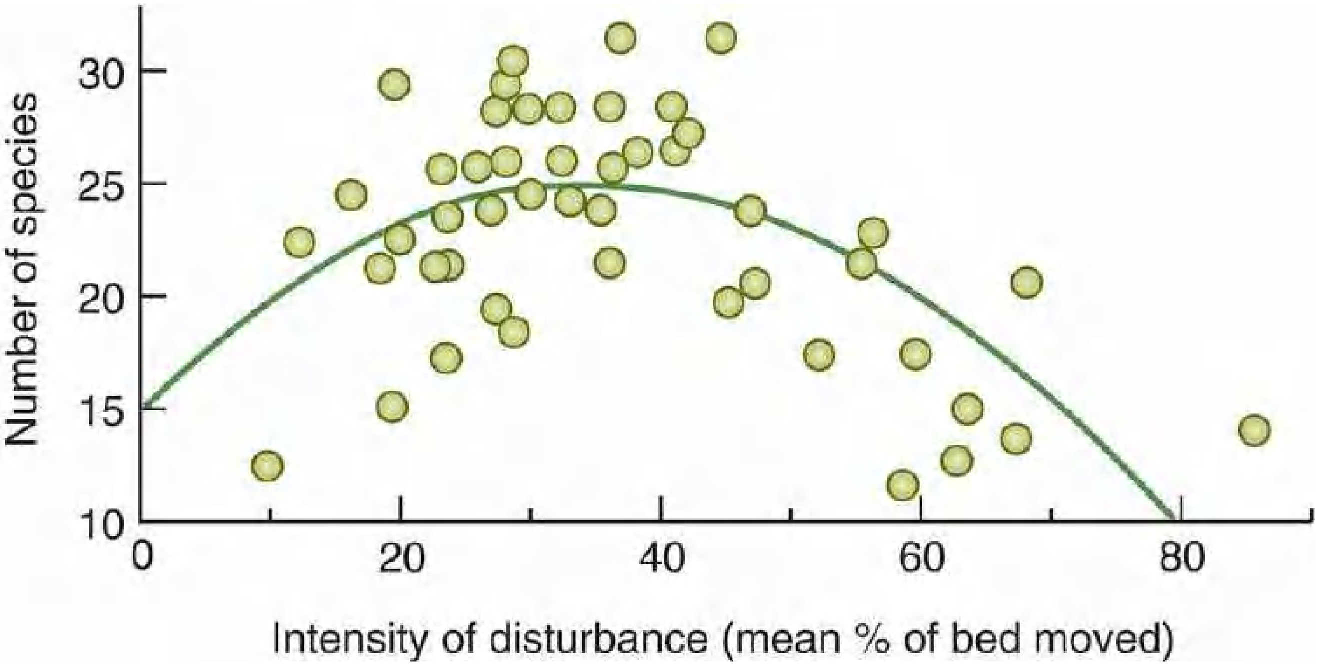


Figure 18.17 **Support for the intermediate disturbance hypothesis.** Relationship between invertebrate species richness and intensity of disturbance measured at 54 stream sites in the Taieri River, New Zealand, as the average percentage of the streambed that moved between sampling occasions (polynomial regression fitted, relationship significant at $P < 0.001$).

Source: After Townsend *et al*. (1997).

... but the pattern is far from general

While a proportion of observational (25%) and experimental (15%) studies have provided support for the intermediate disturbance hypothesis, many have described linear increases or decreases of diversity with increasing disturbance, a handful have produced U-shaped relationships (diversity minimal at intermediate disturbance) and the rest have reported no pattern (Hughes et al., 2007). A partial explanation may be that some studies have only incorporated the rising (or falling) limb of the full range of disturbance levels. On the other hand, perhaps reality is too complex to be neatly captured by a universal catchall explanation.

18.6.3 Formation and filling of gaps

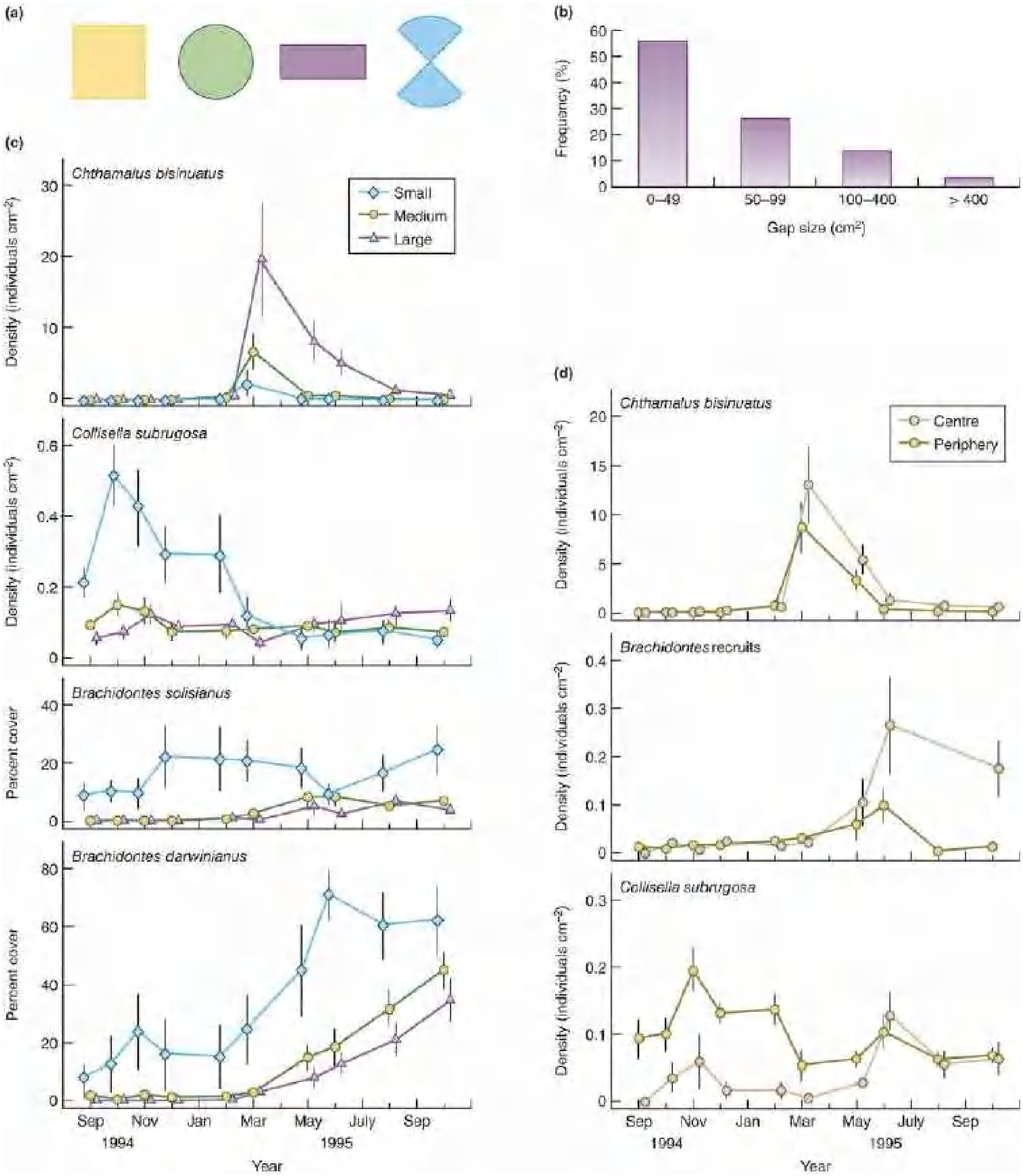
Gaps of different sizes may influence community structure in different ways because of contrasting mechanisms of recolonisation. The centres of very large gaps are most likely to be colonised by species producing propagules that travel relatively great distances. Such mobility is less important in small gaps, since most recolonising propagules will be produced by adjacent established individuals. The smallest gaps of all may be filled simply by lateral movements of individuals around the periphery.

Intertidal beds of mussels provide excellent opportunities to study the processes of formation and filling-in of gaps. In the absence of disturbance, mussel beds may persist as extensive monocultures. More often, they are an ever-changing mosaic of many species that inhabit gaps formed by the action of waves. Gaps can appear virtually anywhere, and may exist for years as islands in a sea of mussels. The size of these gaps at the time of formation ranges from the dimensions of a single mussel to hundreds of square metres. In general, a mussel or group of mussels becomes infirm or damaged through disease, predation, old age or, most often, the effects of storm waves or battering by logs. Gaps begin to fill as soon as they are formed.

In their experimental study of mussel beds of *Brachiodontes solisianus* and *B. darwinius* in Brazil, Tanaka and Magalhaes (2002) aimed to determine the differential effects of patch size and perimeter : area ratio on the dynamics of succession. In an experiment on one moderately exposed shoreline, they created square gaps with different areas (because of identical shapes, the bigger squares had smaller perimeter : area ratios) (Table 18.2). On a nearby and physically very similar shore, they created patches of four different shapes and chose areas for each that produced identical perimeter : area ratios (Figure 18.18a). Note that a circle has the most perimeter per unit area of any shape. The gap sizes were within the range observed for natural gaps, which did not differ on the two shores (Figure 18.18b).

Table 18.2 Measures of area, perimeter and perimeter: area ratio. Values used for the experimental gaps created in two experiments on semi-exposed shores in south-east Brazil. *Source*: From Tanaka & Magalhaes (2002).

	Area (cm^2)	Perimeter (cm)	Perimeter : area ratio
Patch size effects			
Square	25	20	0.8
Square	100	40	0.2
Square	400	80	0.2
Patch shape effects			
Square	100.0	40.0	0.4
Circle	78.5	31.4	0.4
Rectangle	112.5	45.0	0.4
Sector	190.1	78.6	0.4



[**Figure 18.18**](#) **Colonisation patterns in patches of different size and shape.** (a) The four shapes used in patch shape experiments: square, circle, rectangle and 'sector' (see [Table 18.2](#)). (b) Size distribution of natural gaps in the mussel beds. (c) Mean abundances (± SE) of four colonising species in experimentally cleared small, medium and large square gaps. (d) Recruitment of three species at the periphery (within 5 cm of the gap edge) and in the centre of 400 cm² square gaps.

Source: After Tanaka & Magalhaes ([2002](#)).

colonisation of gaps in mussel beds, ...

Higher densities of the herbivorous limpet *Collisella subrugosa* occurred in the small gaps in the first 6 months after gap formation (Figure 18.18c). Small gaps, compared with medium and large gaps, were also most quickly colonised by lateral migration of the two mussel species, but with *Brachidontes darwinianus* predominating. The larger gaps had higher densities of the barnacle *Chthamalus bisinuatus* and sheltered more limpets at their edges, while central areas had more *Brachiodontes* recruited from larvae after 6 months (Figure 18.18d). The gaps with identical perimeter : area ratios showed very similar patterns of colonisation despite their different sizes, emphasising that colonisation dynamics are mainly determined by distance from adjacent sources of colonists.

... in grassland ...

The pattern of colonisation of gaps in mussel beds is repeated in almost every detail in the colonisation of gaps in grassland caused by burrowing animals or patches killed by urine. Initially, leaves lean into the gap from plants outside it. Then colonisation begins by clonal spread from the edges, and a very small gap may close up quickly. In larger gaps, new colonists may enter as dispersed seed, or germinate from the seed bank in the soil. Over 2–3 years the vegetation begins to acquire the character that it had before the gap was formed.

... and in mangrove forest

The gaps produced in forests vary greatly in size. Lightning-induced gaps in mangrove forest in the Dominican Republic, for example, range from 200 to 1600 m^2 or more (Figure 18.19). Lightning almost always kills groups of trees in a 20–30 m circle, and the trees remain as standing dead for several years. In a forest dominated by red mangrove (*Rhizophora mangle*) and white mangrove (*Laguncularia racemose*), and with some black mangrove (*Avicennia germinans*), Sherman *et al*. (2000) compared the performance of the three species in lightning gaps and under forest canopy. Seedling density did not differ in gaps and intact forest, but sapling density and the growth rates of all three species were much higher in the gaps (Table 18.3). However, gap regeneration was dominated by *R. mangle* because its mortality rate was much lower in gaps than was the case for the other species. Sherman *et al*. (2000) note that the peat mat on the forest floor usually collapses after lightning damage, resulting in increased levels of standing water. They suggest that the success of *R. mangle* in gaps is due to their higher tolerance of flooding conditions.

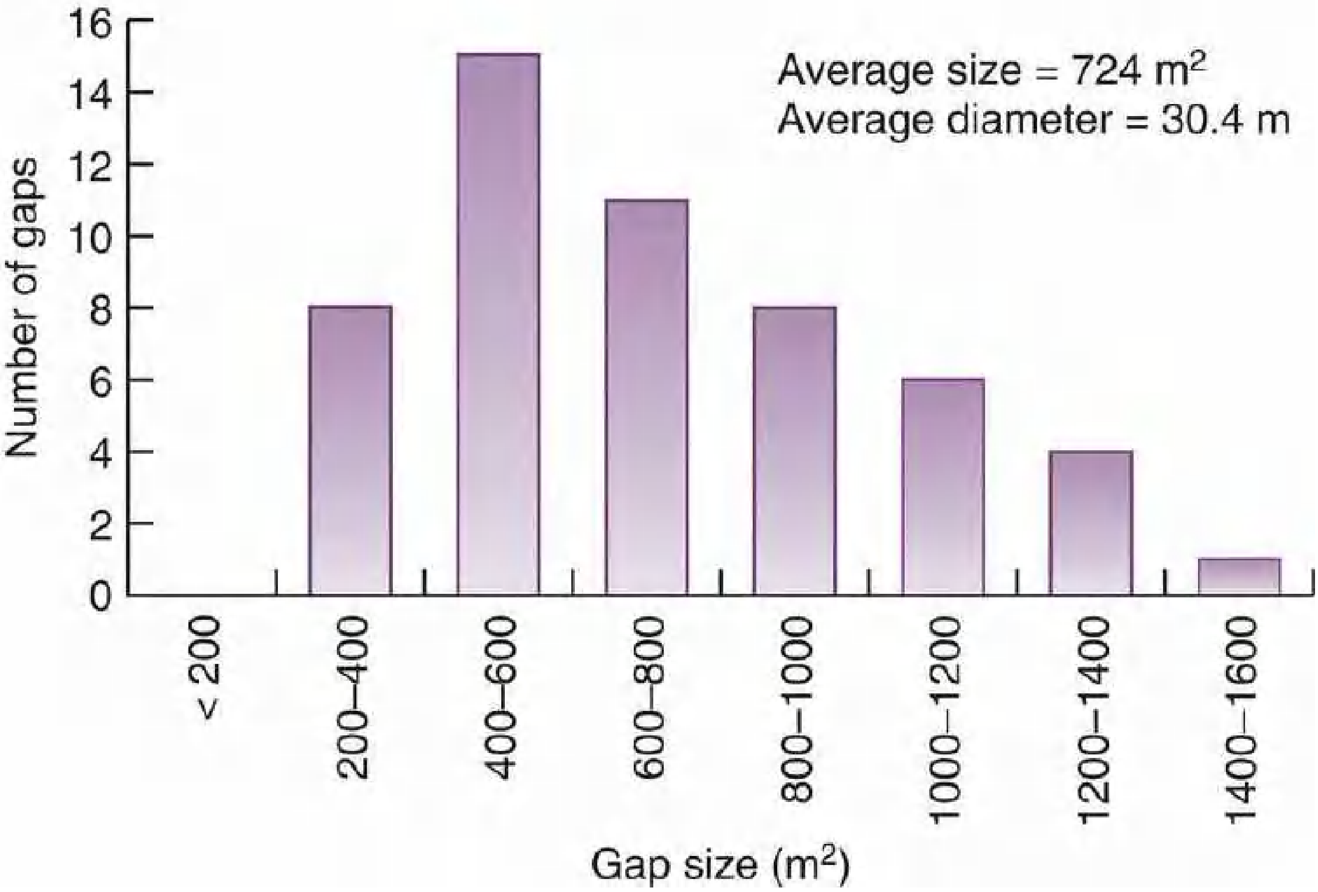


Figure 18.19 Frequency distribution of gaps created by lightning in a tropical mangrove forest in the Dominican Republic.

Source: After Sherman *et al.* (2000).

Table 18.3 Contrasting features of three mangrove species. Initial size, and growth and mortality rates over a 1-year period of saplings of three mangrove species in lightning-induced gaps and under intact forest canopy. *Source*: After Sherman *et al.* (2000).

	Initial sapling diameter (cm ± SE)		**Growth rate–diameter increment (cm ± SE)**		**Mortality (%)**	
	Gaps	**Canopy**	**Gaps**	**Canopy**	**Gaps**	**Canopy**
Rhizophora mangle	1.9 ± 0.06	2.3 ± 0.06	0.58 ± 0.03	0.09 ± 0.01	9	16
Laguncularia racemosa	1.7 ± 0.11	1.8 ± 0.84	0.46 ± 0.04	0.11 ± 0.06	32	40
Avicennia germinans	1.3 ± 0.25	1.7 ± 0.45	0.51 ± 0.04	–	56	88

Organisms other than plants can also be overrepresented in gaps. In a study of tropical rainforest in Costa Rica, Levey (1988) found that nectarivorous and frugivorous birds were much more abundant in treefall gaps, reflecting the fact that understory plants in gaps tend to produce more fruit over a longer period than conspecifics fruiting under a closed canopy.

founder-controlled communities – a competitive lottery to fill gaps ...

We have discussed a variety of cases where species differ in their colonising ability and competitive status so that the original coloniser of a patch cannot necessarily maintain its position there (termed dominance-controlled communities; see Section 18.4). In founder-controlled communities (Yodzis, 1986), on the other hand, all species are both good colonists and essentially equal competitors; thus, within a patch opened by disturbance, a competitive lottery rather than a predictable succession is to be expected. If a large number of species are approximately equivalent in their ability to invade gaps, are equally tolerant of the abiotic conditions and can hold the gaps against all comers during their lifetime, then the probability of competitive exclusion may be much reduced in an environment where gaps are appearing continually and randomly. A further condition for coexistence is that the number of young that invade and occupy the gaps should not be consistently greater for parent populations that produce more offspring, otherwise the most productive species would come to monopolise the space even in a continuously disturbed environment.

... involving reef fish ...

Some tropical reef communities of fish may conform to this model (Sale, 1977 , 1979). They are extremely diverse. For example, the number of species of fish on the Great Barrier Reef, off eastern Australia, ranges from 900 in the south to 1500 in the north, and more than 50 resident species may be recorded on a single patch of reef 3 m in diameter. Only a proportion of this diversity is likely to be attributable to resource partitioning of food and space – indeed, the diets of many of the coexisting species are very similar. In this community, vacant living space seems to be a crucial limiting factor, and it is generated unpredictably in space and time when a resident dies or is killed. The lifestyles of the species match this state of affairs. They breed often, sometimes year-round, and produce numerous clutches of dispersive eggs or larvae. It can be argued that the species compete in a lottery for living space in which larvae are the tickets, and the first arrival at the vacant space wins the site, matures quickly and holds the space for its lifetime.

Three species of herbivorous pomacentrid fish co-occur on the upper slope of Heron Reef, part of the Great Barrier Reef. Within rubble patches, the available space is occupied by a series of contiguous and usually non-overlapping territories, each up to 2 m^2 in area, held by individuals of *Eupomacentrus apicalis*, *Plectroglyphidodon lacrymatus* and *Pomacentrus wardi*. Individuals hold territories throughout their juvenile and adult life and defend them against a broad range of chiefly herbivorous species, including conspecifics. There seems to be no particular tendency for space initially held by one species to be taken up, following mortality, by the same species. Nor is any successional sequence of ownership evident (Table 18.4). In a similar vein, Geange and Stier (2009) found that success on reef patches in French Polynesia of ecologically similar wrasse species (genus *Thalassoma*) was determined by their order of arrival.

Table 18.4 Reoccupation patterns of reef fish after a resident dies. Numbers of individuals of each species observed occupying sites, or parts of sites, that had been vacated during the immediately prior interperiod between censuses through the loss of residents of each species. The sites vacated through loss of 120 residents have been reoccupied by 131 fish; the species of the new occupant is not dependent on the species of the previous resident.

	Reoccupied by		
Resident lost	**E. apicalis**	**P. lacrymatus**	**P. wardi**
Eupomacentrus apicalis	9	3	19
Plectroglyphidodon lacrymatus	12	5	9
Pomacentrus wardi	27	18	29

... or plants in grassland or forest

Thus, the maintenance of high reef diversity depends, at least in part, on the unpredictability of the supply of living space; and as long as all species win some of the time and in some places, they will continue to put larvae into the plankton, and hence, into the lottery for new sites. An analogous situation has been postulated for the highly diverse chalk grasslands of Great Britain (Grubb, 1977) and even for trees in temperate and tropical forest gaps (Busing & Brokaw, 2002). Any small gap that appears is rapidly exploited, by a seed in grassland and very often by a sapling in a forest gap. In these cases, the tickets in the lottery are saplings or seeds (either in the act of dispersal or as components of a persistent seed bank in the soil). Which seeds or saplings develop to established plants, and therefore which species comes to occupy the gap, may depend on a strong random element since many species overlap in their requirements for successful growth. The successful plant rapidly establishes itself and retains the patch for its lifetime, in a similar way to the reef fish described earlier.

APPLICATION 18.3 Managing successional mosaics for conservation

Successional mosaic theory is relevant to the fate of endangered species if they require a particular successional stage for success.

It was long thought that Canada lynx (*Lynx canadensis*) inhabited primary forest unoccupied by people. Indeed, a federal judge ruled in 1997 that such mature forest is a prerequisite for them (Hoving *et al.*, 2004). However, lynx specialise in hunting snowshoe hares (*Lepus americanus*) whose habitat is in early successional dense shrubland or immature forest. Hoving *et al.* (2004) addressed this paradox by comparing landscapes in Maine, USA, where lynx were present or absent (determined by surveying tracks in the snow). Snowshoe hare densities were high in densely regenerating forest where trees had previously been completely removed by clear-cutting. Lynx were also most likely to occur in landscape units (each 100 km^2) containing a large proportion of regenerating forest. Both lynx and hares were least likely to occur where there has been recent clear-cutting or partial forest felling, and were neither positively nor negatively associated with mature forest.

Clear-cutting is beneficial to lynx in the long term (but not for the first several years) because it produces dense forest regeneration with abundant snowshoe hares, a pattern that may mimic natural disturbances such as fire or insect outbreaks that kill trees over large areas. Future forest harvesting will need to be planned so that suitably sized patches of regenerating forest are available in the landscape.

Just as lynx need snowshoe hares, the caterpillars of butterflies need suitable host plants. During the last century, livestock grazing and periodic droughts reduced herbaceous ground cover in piñon–juniper woodland (*Pinus edulis*, *Juniperus monosperma*) in New Mexico and Arizona, USA, interrupting the natural fire regime and producing dense forest lacking the original mosaic that included open patches with grassy understories and rich soils. Kleintjes *et al.* (2004) assessed a management regime intended to restore the former habitat mosaic and benefit host plants and butterflies. Overstory reduction and slash mulching (ORSM) involves removing small trees (<20 cm diameter) and the main branches of larger trees and applying the resulting wood and foliage as a fine surface layer of organic matter (mulch), particularly to areas of eroded soil. Both abundance and species richness of butterflies were significantly higher four years after ORSM treatment than in untreated woodland (Figure 18.20). The ground cover of grasses and forbs, including five of the 10 most common host plants, was greater in the open patches in the restored successional mosaic. For example, the success of the legume *Lotus wrightii* benefited the orange sulphur butterfly (*Colias eurytheme*) and the western green hairstreak (*Callophrys affinis*).

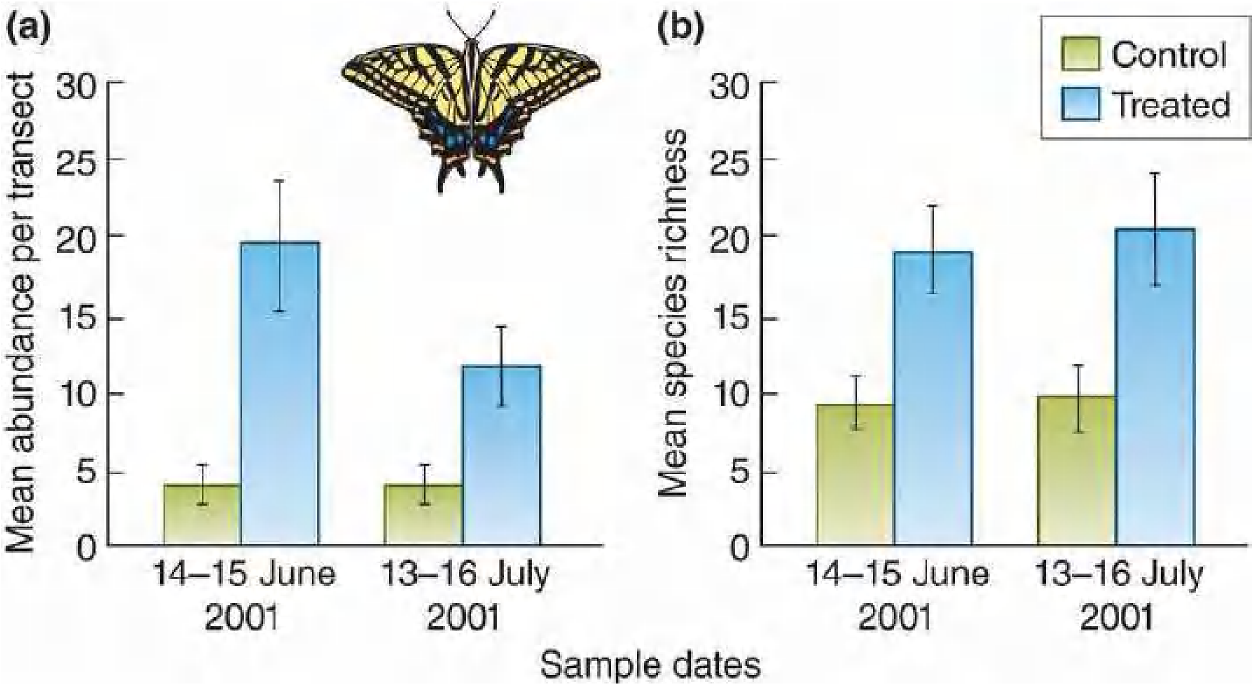


Figure 18.20 Butterflies profit from overstory reduction and slash mulching. (a) Abundance (± 95% confidence limits) and (b) species richness of butterflies are increased four years after overstory reduction and slash mulching (ORSM = 'treated', in comparison to 'control') to create open habitat patches in piñon–juniper woodland in New Mexico. The surveys were carried out on two occasions in mid-summer.

Source: After Kleintjes *et al*. (2004).

18.7 The metacommunity concept

It is clear from the sections above that to fully understand the nature of the community, it is vital to consider not only local factors, such as environmental conditions, disturbances and species interactions, but also regional factors such as the number of species available in the regional pool, the size and disposition of habitat patches, and dispersal patterns. The *metacommunity concept* brings together these local and regional factors. A metacommunity is defined as *a set of communities linked by dispersal of individuals of one or more potentially interacting species* (Wilson, 1992 ; Leibold *et al*., 2004).

local and regional factors jointly determine community composition

Four metacommunity models have been proposed that differ according to whether species are functionally equivalent or vary in their niche requirements, whether underlying rates of dispersal are high or low, and whether the regional environment is homogeneous or heterogeneous (Leibold *et al*., 2004) (Figure 18.21). Inevitably, these models are simplifications of nature, and none is entirely realistic or satisfying, but each confirms something important about the processes underlying community patterns.

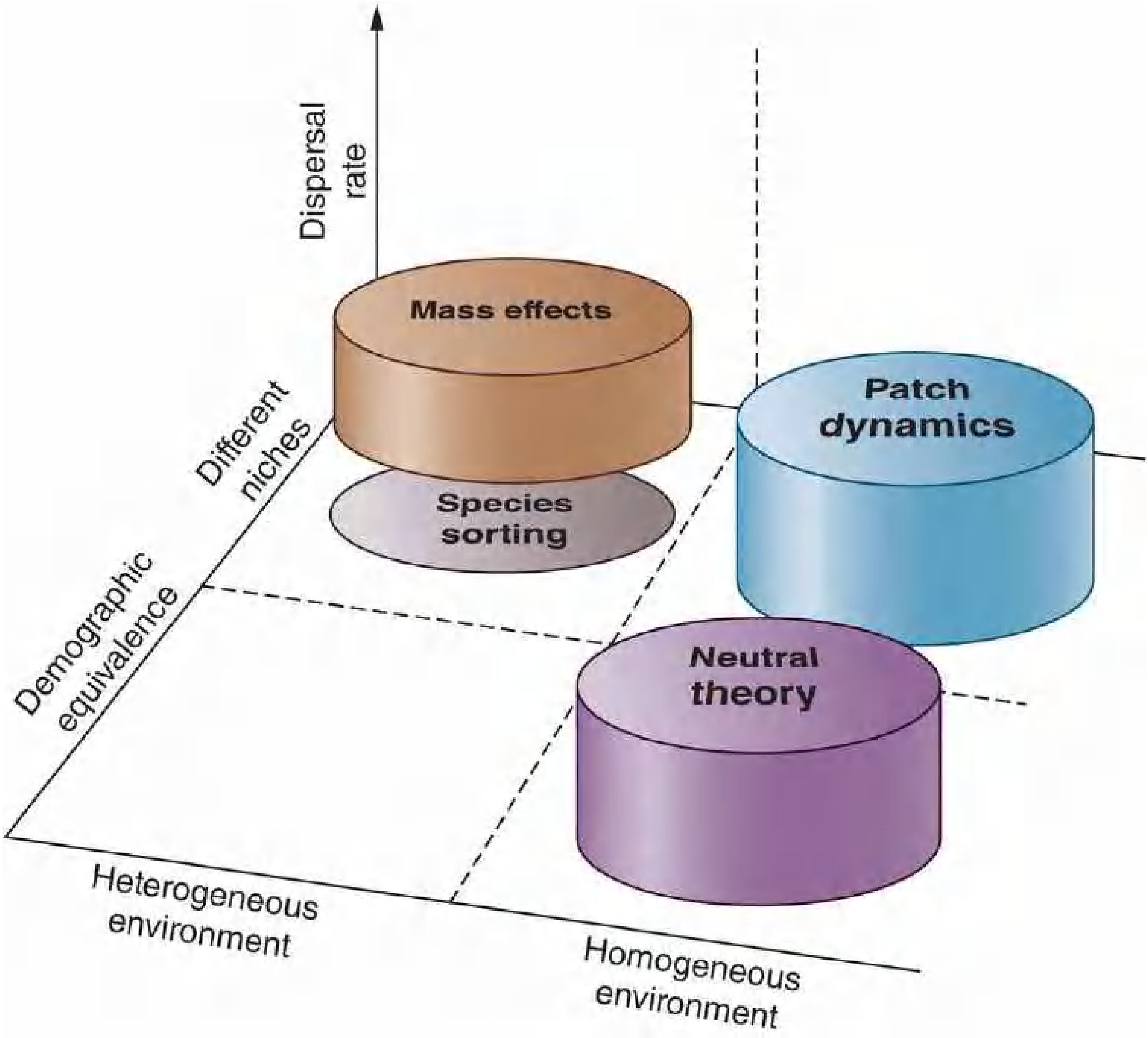


Figure 18.21 **Three-dimensional classification of four paradigms of metacommunity functioning.** Dispersal rates can vary substantially in each type of model except for the species-sorting metacommunity, where dispersal rate is assumed to be at the lower end of the scale.

18.7.1 The patch dynamics metacommunity model

This first metacommunity model is an extension to more than two species of the metapopulation model (see Section 6.7). The landscape is assumed to consist of a spatially homogeneous set of identical patches (Figure 18.21), each of which can be occupied or unoccupied. The model does not include short-term species dynamics, but simply tracks for each species the proportion of patches occupied, for which they compete. Rates of colonisation and extinction are assumed to be correlated, so that a species with a high rate of extinction can compensate by a high rate of colonisation. Modelling shows that this simple state of affairs can theoretically maintain many species indefinitely in the metacommunity if their niches differ (Figure 18.21), the most competitive wins in a patch, and a trade-off exists between competitive ability in the patch and the ability to colonise a patch (the classic competition–colonisation trade-off introduced earlier) (Leibold *et al.*, 2004).

examining the role of competition–colonisation trade-offs

While there is evidence of competition–colonisation trade-offs in nature and good reason to believe that they are not uncommon (see Section 18.5.2), many field studies have failed to provide good evidence for them, perhaps because patterns in nature may be overwhelmed by spatial heterogeneity. Cadotte *et al*. (2006) tested for a trade-off in a microcosm experiment designed to remove the problem of heterogeneity. Their aquatic microcosm consisted of several identical patches (bottles of nutrient solution and bacteria) connected serially, into which protozoan and rotifer species (13 in all) were introduced individually and their colonisation ability estimated as the number of weeks taken to colonise all the patches. Then, in single-bottle experiments, each species was introduced with one of the other 12 species (78 combinations) and competitive ranks were determined after eight weeks (in terms both of the number of trials where a species was still present, and the number of extinctions it caused). Figure 18.22 shows that in this artificial experiment involving species known to coexist in nature, there was indeed a competition–colonisation trade-off.

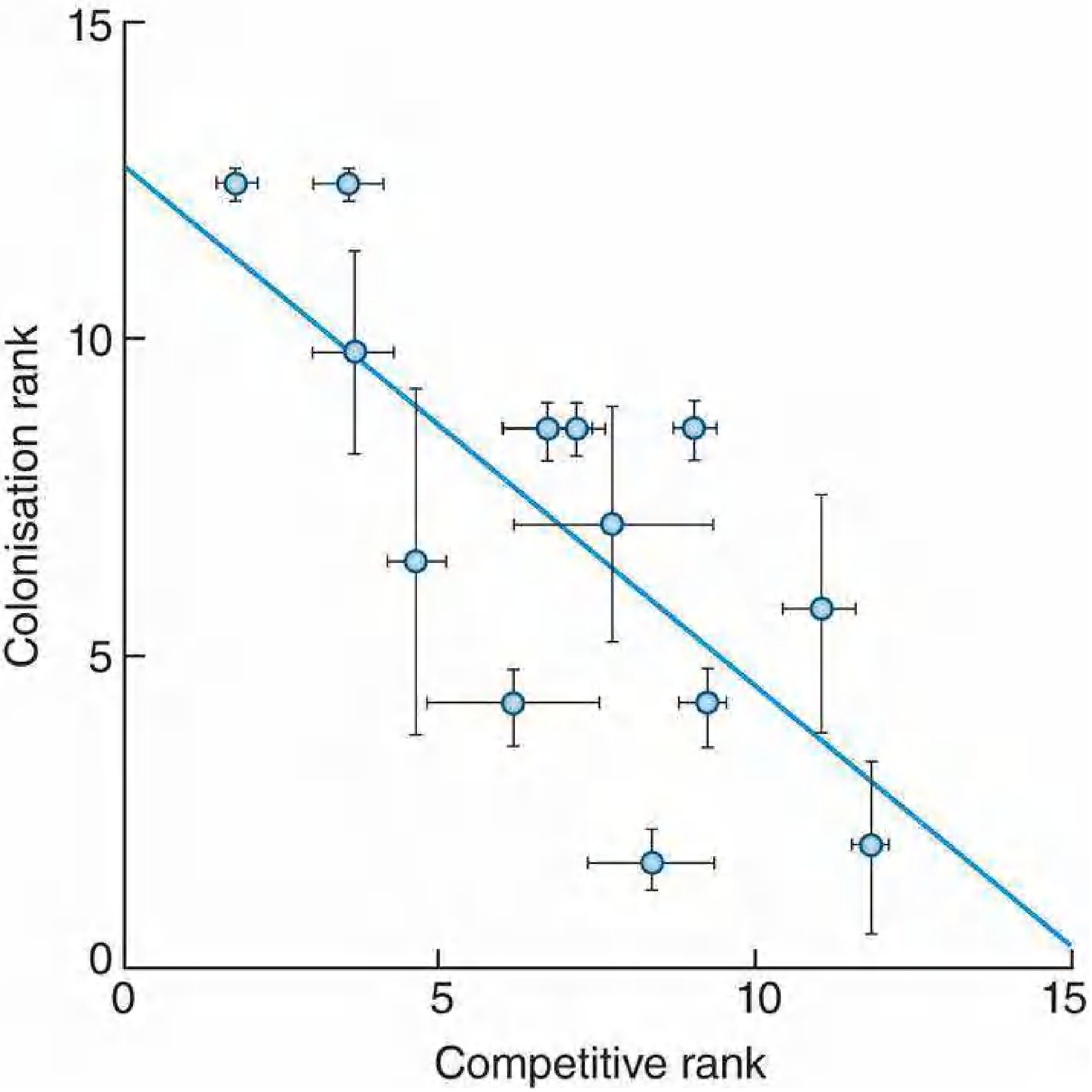


Figure 18.22 **Illustration of a competition–colonisation trade-off.** Protozoan and rotifer species that often coexist in nature and in laboratory mesocosms demonstrate a competition–colonisation trade-off. Means ± 95% CIs are shown.

Source: After Cadotte *et al*. (2006).

two overly simplistic perspectives – patch dynamics ...

The patch dynamics metacommunity model is highly simplistic because it does not consider the differences in conditions and resources among patches that commonly occur in nature. However, the value of the model is to establish that a simple competition–colonisation trade-off may contribute to maintaining diversity in a metacommunity whether or not other influential processes are at work.

18.7.2 The neutral metacommunity model

... and neutral models – ...

Hubbell's (2001) neutral model, that we met previously in Section 16.2.4, appears even more unrealistic because, in addition to assuming environmental homogeneity, it explicitly assumes that species niches do not differ in any way that is demographically significant or, more specifically, that all species are functionally identical in terms of their per capita rates of birth, death and dispersal (see Figure 18.21). This runs counter to most of the empirical examples already discussed in this chapter. However, it is worth recalling that a focus of Hubbell's early work was the extreme diversity of trees (more than 300 species) in a small area of tropical forest (50 ha). Could there really be so many distinct niches? Or might such diversity be maintained in a metacommunity by some means that does not require niche differences among the species?

The neutral model assumes a constant number of individuals in the metacommunity, and that all space is occupied, so that gains by one species are balanced by losses by another. When a site becomes vacated by death (which occurs at a constant rate), the new occupant is drawn at random from the species present (with probabilities of selection determined by relative abundances of species in the metacommunity). Species abundances vary through time because of random birth and death rates, and community dynamics follow a stochastic process of drift (see Section 18.1) that leads inexorably to the extinction of all but one species in the metacommunity. The process can take such a very long time, however, that new species arising via speciation may counteract the downward drift.

... that neverthess provide important insights

However unrealistic, Hubbell's model has the value of raising the profile of drift and speciation, both of which tend to have been ignored by many community ecologists. Furthermore, the neutral model can be viewed as a null hypothesis against which to judge the importance of the various factors that actually determine community composition and biodiversity. In the case of 'competitive lotteries', such as that involving reef fish discussed in Section 18.6.3, neutral processes may be particularly prominent.

18.7.3 The species-sorting metacommunity model

two models that are more realistic – species sorting ...

Another alternative, the species-sorting model, is much closer to what we commonly observe in nature: the environment is highly heterogeneous and species differ in their niche requirements and abilities to utilise different patch types. Dispersal between patches is assumed to occur at a low rate (see Figure 18.21) such that it does not fundamentally affect abundance in a given patch or the outcome of species interactions there. However, the dispersal rate is high enough that each species is capable of reaching all patches where abiotic conditions allow it to persist. Thus, species sort themselves across the region according to their niche requirements, with the best-adapted species in any patch tending to exclude others. Populations in a patch are assumed to reach equilibrium (whether a stable point or more complex oscillating dynamics) in between colonisation events or extinction-causing disturbances (Leibold *et al.*, 2004). The perspective is effectively the classic idea of niche differentiation and coexistence, but scaled up to a metacommunity in which dispersal allows species to alter their distributions in response to temporal shifts in environmental conditions across the landscape. The perspective can be extended to include the idea of a successional mosaic (see Section 18.6), where disturbances open up patches and the success of a dominant species in a patch may be facilitated by a subdominant (termed a 'patch disturbance–succession metacommunity' by Gonzalez (2009)).

18.7.4 The mass-effects metacommunity model

... and mass effects

Like the species-sorting case, the final, mass-effects metacommunity model also assumes niche differences among species in a heterogeneous environment of different patch types, but now dispersal rates are assumed to be higher. Thus, dispersal between patches can affect abundance (immigration supplements local birth rates while emigration affects local loss rates) and therefore influence species interactions within a patch. One consequence is that emigrants from high-quality patches, where for a particular species births exceed deaths, can sustain a population of that species in a low-quality patch where it would otherwise go extinct. This is known as a ‘mass effect’. The model, based on assumptions that match much of what is commonly observed in nature, predicts high levels of regional diversity but with many rare species (Gonzalez, 2009).

18.7.5 Patterns in abundance and diversity predicted by metacommunity models

Mouquet and Loreau (2003) provide a good example of the insights that metacommunity modelling can yield about abundance and diversity. Their ‘mass-effects’ model involves 20 (or 40) species that differ in niche requirements (each being the best competitor in one community, the second best in another, and so on) competing for space in 20 (or 40) patches that differ in local conditions. They assumed a constant proportion of dispersal between patches and investigated the consequences of different rates of dispersal for patterns of abundance and diversity.

a revealing mass-effects model ...

Figure 18.23a shows how relative abundance patterns within a patch responded to different rates of dispersal. When dispersal was zero, dominance by the best local competitor in the particular patch (species A) was complete, but at intermediate dispersal rates some species were rescued from competitive exclusion. At high rates of dispersal the species that was the best competitor at the regional scale became dominant (species B). Rank–abundance patterns (see Section 18.2.2) within a patch also responded, with the distribution shifting from geometric at very low values of dispersal to lognormal at intermediate values and back to geometric at high values of dispersal (Figure 18.23b). Perhaps most interesting of all, patterns of species richness responded strongly to dispersal rate within the metacommunity (Figure 18.23c). In the absence of dispersal, α-richness (within patch diversity) was minimal, with a single species in each patch, while β-richness (between patch) and γ-richness (regional) were at their maxima. As dispersal was increased, with emigration rescuing more species from competitive exclusion, α-richness increased and β-richness declined. Beyond the peak in α-richness, further dispersal led to progressively greater dominance by the species that proved most competitive at the regional level until, with very high dispersal, the community consists of this single species, so that γ-richness (and α-richness) were at a minimum and β-richness at zero.

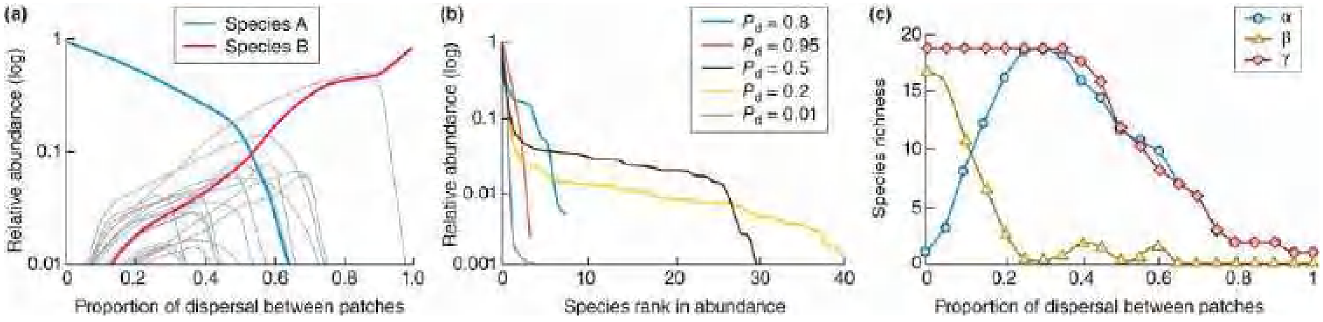


Figure 18.23 **Dispersal rate affects species richness in a metacommunity.** (a) Relative abundance of 20 species within a patch in relation to dispersal rate in the mass-effects metacommunity model of Mouquet and Loreau (2003). Species A is the most competitive in the patch illustrated, while species B is the most competitive in the metacommunity as a whole. (b) Rank–abundance diagrams within a patch (in simulations with 40 species) in relation to proportion of dispersal between patches (P_d values). (c) Patterns in α- (within patch), β- (between patch) and γ-richness (regional) in the metacommunity in relation to dispersal rate (in simulations with 20 species).

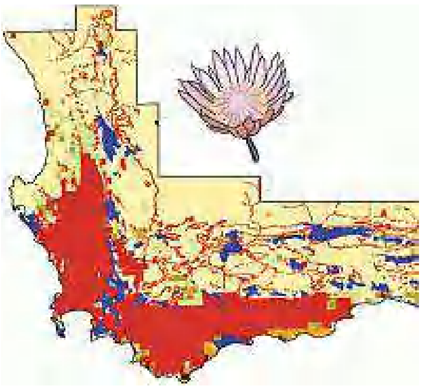
18.7.6 The value and shortcomings of metacommunity models

> ... that highlights why community ecologists must consider metacommunity ideas

The key message from the metacommunity perspective is that the set of species and their relative abundances in a patch (community) studied by a community ecologist are not determined solely by the properties of the species and of the patch in question.

The models certainly do not do justice to the reality of variation in patch qualities, sizes, shapes and interpatch distances or the different classes of disturbances with their own particular scales, timing and phasing. Nor do the models realistically account for interspecific differences in the nature and timing of responses to this variation in nature. Thus, it seems likely that real metacommunities may be located along a continuum from neutral to niche-based dynamics, with some species' dynamics conforming to the former and some to the latter (e.g. Thompson & Townsend, 2006). Moreover, some species in a real metacommunity can be expected to experience low dispersal between patches such that it does not fundamentally affect abundance and species interactions within a patch, as in the species-sorting model, while other species disperse at higher rates that conform to the mass-effects model.

Nevertheless, the models make clear that the dispersal abilities of the species, the arrangement of patches and the factors that physically affect dispersal between patches in the landscape (e.g. barriers, habitat corridors) can all play important roles (e.g. Jones *et al.*, 2015). Furthermore, in our response to ever-growing concerns about biodiversity loss, strategies to conserve and restore biodiversity will often be doomed to failure unless we take into account the metacommunity structure that exists.



Chapter 19
Patterns in Biodiversity and their Conservation

19.1 Introduction

hot spots of species richness

Why do the numbers of species vary from place to place, and from time to time? These are questions that present themselves not only to ecologists but to anybody who observes and ponders the natural world. They are interesting questions in their own right – but they are also questions of practical importance. As we saw in Figure 1.13, a remarkable 44% of the world's plant species and 35% of vertebrate species (other than fish) are endemic to just 25 separate 'hot spots' occupying a small proportion of the earth's surface (Myers *et al.*, 2000). Devising survival plans for single species may be the best way to cope with individual species that are in deep trouble and of special importance (see Section 15.4), but there is no possibility that all endangered species could be dealt with one at a time. Funds for conservation are simply too limited for this. We can, though, aim to conserve the greatest biodiversity by setting aside whole communities in protected areas. It is crucial that the choice of areas is based on sound knowledge of the factors that determine species richness, as laid out in the sections of this chapter. So far, the success of protected areas has been limited, as we show in Application 19.1.

APPLICATION 19.1 Setting aside protected areas to conserve biodiversity

Protected areas of various kinds (national parks, nature reserves, sites of special scientific interest, etc.) grew in both number and area during the 20th century and beyond. One recent estimate found 13% of the world's land area to be protected, though only about 1% of the sea area was (Mora & Sale, 2011). Despite setting aside increased areas for protection, the living planet index, a measure of biodiversity based on the generally reducing population sizes of 1686 vertebrate species worldwide, has continued to decline (Figure 19.1).

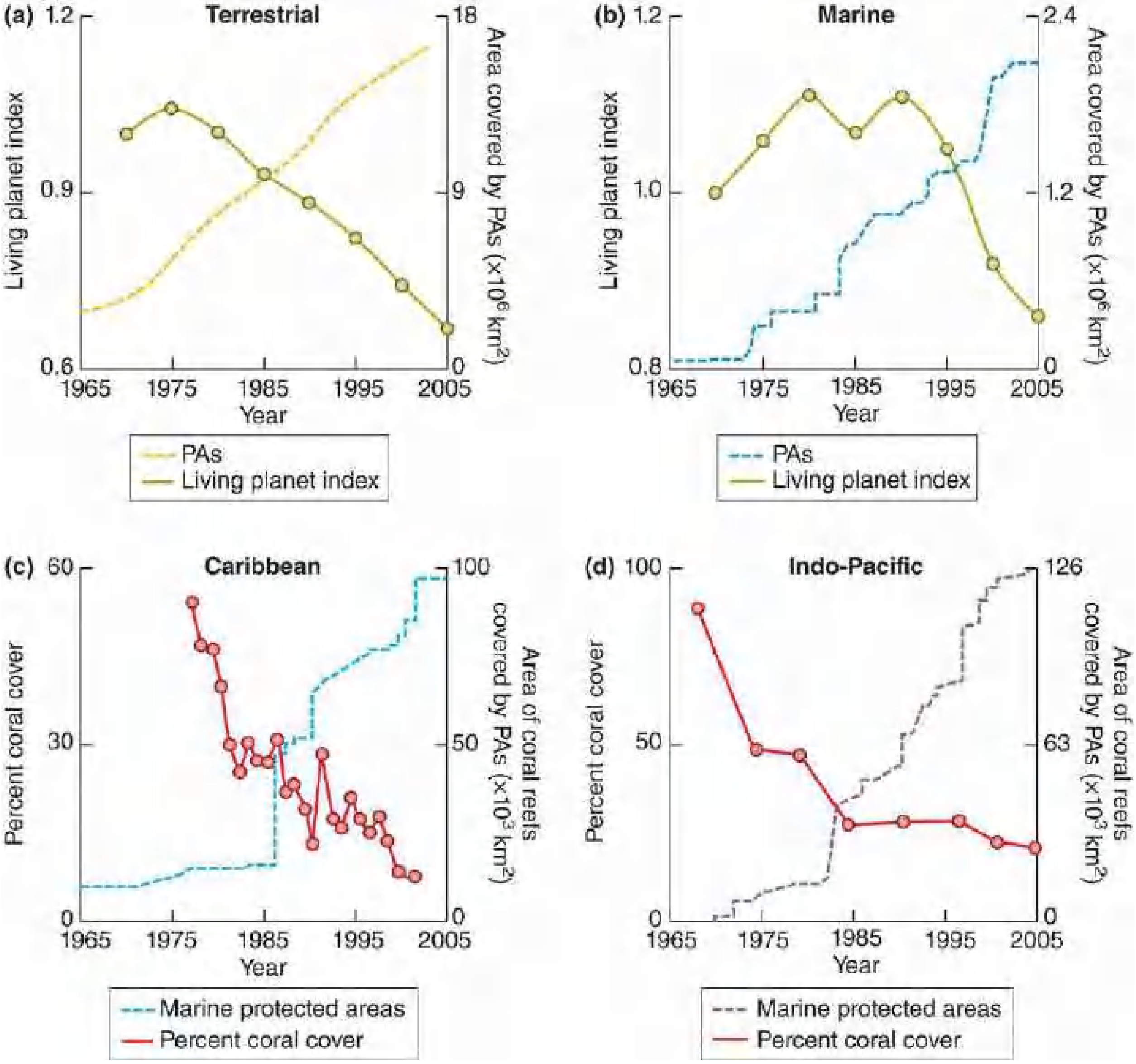


Figure 19.1 Biodiversity has declined despite an increase in protected areas, and coral health has declined despite an increase in coral area covered by marine protected areas. (a) Trends since 1965 in the global extent of area covered by terrestrial protected areas (PAs) and the living planet index, a measure of biodiversity based on the changing population size of 1686 vertebrate species worldwide. (b) Equivalent plots for marine protected areas. (c) Trends since 1965 in the extent of area for coral reefs in the Caribbean Sea included in marine protected areas and the percent coral cover (a measure of coral health). (d) Equivalent plots for the Indo-Pacific region.

Source: After Mora & Sale (2011).

the IUCN classification

A protected area has been defined by the International Union for the Conservation of Nature (IUCN) (Dudley, 2008) as ‘a clearly defined geographical space, recognized, dedicated, and managed through legal or other effective means, to achieve the long-term conservation of nature with associated ecosystem services and cultural values’. Indeed, there are several separate categories of protected areas, as follows:

- Strict *nature reserves* are ‘strictly protected areas set aside to protect biodiversity and also possibly geological/geomorphical features, where human visitation, use and impacts are strictly controlled and limited to ensure protection of the conservation values’.
- *Wilderness areas* are ‘large unmodified or slightly modified areas, retaining their natural character and influence without permanent or significant human habitation, which are protected and managed so as to preserve their natural condition’.
- *National parks* are similar to wilderness areas but also add a more human perspective since they ‘provide a foundation for environmentally and culturally compatible, spiritual, scientific, educational, recreational, and visitor opportunities’.
- *Natural monuments* or *features* are also more visitor-orientated and generally small, being areas ‘set aside to protect a specific natural monument, which can be a landform, sea mount, submarine cavern, geological feature such as a cave or even a living feature such as an ancient grove’.
- *Habitat/species management areas* are areas where the focus is much more on particular species or habitats. These areas therefore typically ‘need regular, active interventions to address the requirements of particular species or to maintain habitats’.
- *Protected landscapes/seascapes*. In this case the focus is not only human but also aesthetic, since these are areas where the interaction of people and nature over time has produced ‘an area of distinct character with significant, ecological, biological, cultural and scenic value: and where safeguarding the integrity of this interaction is vital to protecting and sustaining the area and its associated nature conservation and other values’.
- *Protected areas with sustainable use of natural resources*. Finally, here the human focus is much more practical. These areas ‘conserve ecosystems and habitats together with associated cultural values and traditional natural resource management systems. They are generally large, with most of the area in a natural condition, where a proportion is under sustainable natural resource management and where low-level non-industrial use of natural resources compatible with nature conservation is seen as one of the main aims of the area’.

Of course, areas may be protected without becoming part of the official IUCN portfolio, but we can see from these categories that any protected area must seek to satisfy a range of criteria that may often conflict with one another, and that different types of areas balance these criteria out in different ways. Notable among the criteria are conservation itself, access for humans for educational or recreational purposes, and the ability of humans to exploit the natural resources of the area, either directly for food or more broadly to provide for their subsistence and well-being.

biodiversity and species richness

Before proceeding further, it is important to distinguish between species richness (the number of species present in a defined geographic unit), which we have discussed many times in previous chapters, and biodiversity. The term biodiversity makes frequent appearances in both the popular media and the scientific literature – but it often does so without any clear definition. At its simplest, biodiversity is synonymous with species richness, and certainly species richness is a key aspect of biodiversity. Biodiversity, though, can also be viewed at levels of biological organisation finer and coarser than the species. For example, we may include genetic diversity within species, recognising the value of conserving genetically distinct subpopulations and subspecies. Above the species level, we may wish to ensure that species with no close relatives are afforded special protection, so that the overall evolutionary variety of the world's biota is maintained as large as possible. At a larger scale still, we may include in biodiversity the variety of community types present in a region – swamps, deserts, early and late stages in a woodland succession, and so on. Thus, 'biodiversity' may itself, quite reasonably, have a diversity of meanings. Those using the term should always try to ensure their audience knows what they mean.

Nonetheless, in this chapter we focus our attention on species richness, partly because of its fundamental nature, but mainly because there are so many more data available for this than for any other aspect of biodiversity. We will identify patterns in species richness, and seek to understand what factors drive those patterns – habitat isolation, productivity, and so on. This is clearly related to, but different from, discussions of, for example, the effects *of* species richness *on* community productivity (see Section 20.3.6), not least because we will often be dealing with correlations where the pattern of causation may not be certain.

the question of scale: macroecology

As with other areas of ecology, scale is a paramount feature in discussions of species richness. Thus, the number of species living on a boulder in a river will reflect local influences such as the range of microhabitats provided and the consequences of species interactions taking place. But larger scale influences of both a spatial and temporal nature will also be important. Species richness may be large on our boulder because the regional pool of species is itself large (in the river as a whole or, at a still larger scale, in the geographic region) or because there has been a long interlude since the boulder was last turned over by a flood (or since the region was last glaciated). Comparatively more emphasis has been placed on local as opposed to regional questions in ecology, prompting Brown and Maurer (1989) to designate a subdiscipline of ecology as *macroecology* to deal explicitly with understanding distribution and abundance at large spatial and temporal scales. Geographic patterns in species richness are a principal focus of macroecology (e.g. Gaston & Blackburn, 2000 ; Smith *et al*., 2014).

four types of factor affect species richness

There are various types of factor to which the species richness of a community can be related. First, there are factors that can be referred to as geographic gradients, notably latitude, elevation and, in aquatic environments, depth. These have often been correlated with species richness, as we shall discuss below, but presumably they cannot be causal agents in their own right. If species richness changes with latitude, then there must be some other factor changing with latitude, exerting a direct effect on the communities.

A second group of factors does indeed show a tendency to be correlated with these gradients, but they are not perfectly correlated. To the extent that they are correlated at all, they may play a part in explaining latitudinal and other trends. But because they are not perfectly correlated, they serve also to blur these relationships. Such factors include climatic variability, the input of energy, the productivity of the environment and possibly the 'age' of the environment and the 'harshness' of the environment.

A further group of factors vary geographically but quite independently of latitudinal or other gradients. This is true of the amount of physical disturbance a habitat experiences, the isolation of the habitat and the extent to which it is physically and chemically heterogeneous.

Finally, there is a group of factors that are biological attributes of a community, but are also important influences on the structure of the community of which they are part. Notable amongst these are the amount of predation, parasitism or competition in a community (discussed in Chapter 16), the spatial or architectural heterogeneity generated by the organisms themselves, and the successional status of a community (discussed in Chapter 18). These are second-order factors in that they are themselves the consequences of influences outside the community. Nevertheless, they can all play powerful roles in the final shaping of community structure.

Here, therefore, we begin by constructing a simple theoretical framework (following MacArthur (1972)) to help us think about variations in species richness (Section 19.2). We then consider factors whose variation is primarily spatial (productivity, energy, spatial heterogeneity, environmental harshness – Section 19.3), and those whose variation is primarily temporal (climatic variation and environmental age – Section 19.4). Next, we turn to patterns in species richness related to habitat area and remoteness (island patterns – Section 19.5), before moving to the trends in species richness related to latitude, elevation, depth and succession (Section 19.6). First of all, however, we must return briefly to a question we touched on in Section 18.2, namely, how do we arrive at reliable estimates of species richness?

19.1.1 Estimating richness: rarefaction and extrapolation

In order to relate variations in species richness either to forces that may determine richness (like productivity) or features that may have a pattern of association with richness (like latitude), we must be confident that those variations are real. As we noted in Section 18.2, this can be difficult, because estimates of richness always come from incomplete samples of the community (not the whole community) and are therefore unlikely to have detected all the species present. Moreover, different samples are likely to reflect contrasting levels of completeness, and this may undermine the validity of comparisons between them.

There are, in fact, several types of data in which these problems may arise (Gotelli & Colwell, 2011 ; Colwell *et al.*, 2012). In the first place, the data may be either *individual-based*, where individuals are taken, ideally at random and independently, from the community in question and their species identity noted, or they may be *sample-based*, where representative samples of a standard size (quadrats, etc.) are taken from the community and the species identity of all the sampled individuals noted. In addition, we may simply note the presence of a species in a sample – in which case we have *incidence data*; or we may count the numbers of individuals in each species, generating *abundance data*. Obviously, abundance data can be converted to incidence data, but not vice versa.

Irrespective of data type, however, there are two broad approaches to help overcome the problems of richness estimate comparability. The first, and longest established, is to use *rarefaction* – down-sampling the larger samples until they contain the same number of observations as the smallest sample, and then comparing the richnesses of these reduced samples. To do this, it is necessary to have a *species accumulation curve*, plotting the cumulative number of species observed with increases in the number of individuals sampled (individual-based data) or the number of samples (sample-based data) (Figure 19.2a). From this, a smooth rarefaction curve can be fitted to the data, based on realistic assumptions about the underlying sampling process, which will of course differ depending on whether the sampling is individual-based or sample-based (Colwell *et al.*, 2012).

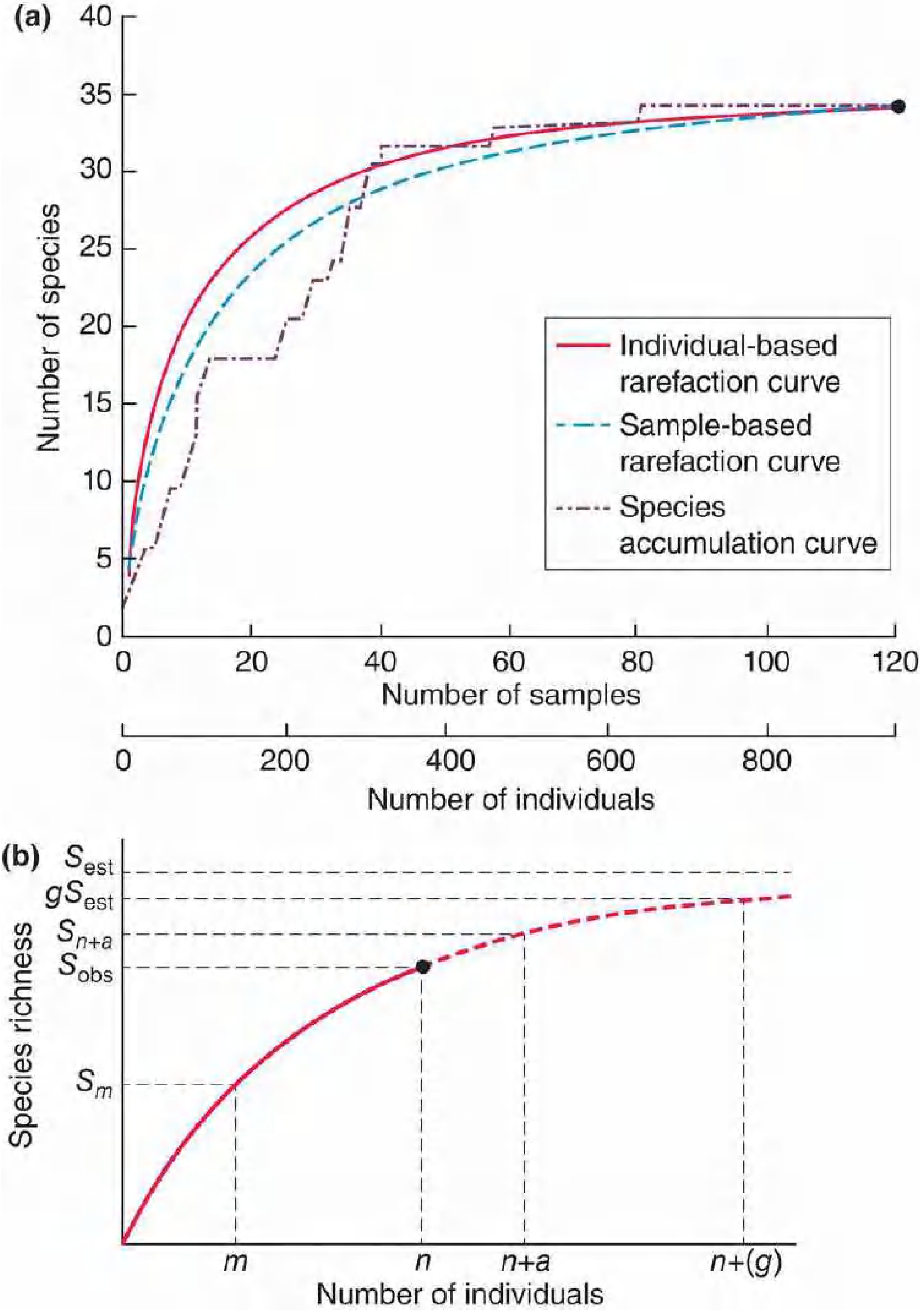

(a)
Number of species
40
35
30
25
20
15
10
5
0
0
20
40
60
80
100
120
Number of samples
0
200
400
600
800
Number of individuals
Individual-based rarefaction curve
Sample-based rarefaction curve
Species accumulation curve
(b)
Species richness
S_{est}
gS_{est}
S_{n+a}
S_{obs}
S_m
m
n
n+a
n+(g)
Number of individuals

Figure 19.2 Rarefaction and extrapolation as ways of estimating species richness from raw data. (a) A hypothetical species accumulation curve (raw data) and the estimated rarefaction curves, based on whether those raw data were obtained from increasing numbers of samples or increasing numbers of individuals. (b) A typical rarefaction and extrapolation curve (solid and dashed lines, respectively) for increasing numbers of individuals, though a comparable curve can be drawn for increasing numbers of samples. Ignoring all methodological details, the raw data have a maximum number of species observed, S_{obs}, from a sample of n individuals, generating a curve with an estimated asymptotic species richness of S_{est}. Rarefaction to a lowest sample size, m, generates an estimate for this community at this sample size of S_m. Extrapolation to an augmented sample size, $n + a$, generates an estimate for this community at this sample size of S_{n+a}. In order to reach a proportion of the asymptotic richness, gS_{est}, the estimated augmented sample size is $n + (g)$.

Source: (a) After Gotelli & Colwell (2011). (b) After Colwell *et al.* (2012).

An obvious disadvantage of rarefaction, however, is that data (often hard-earned data!) are discarded. An alternative, therefore, is *extrapolation*, where the rarefaction curve is extrapolated so that the number of species is estimated that would be observed in a sample that generates a specified proportion of the estimated asymptotic richness at the top of the species accumulation/rarefaction curve. These and other estimates are illustrated in Figure 19.2b. Details of how these estimates can be made are provided for standard richness measures by Gotelli and Colwell (2011), and for Hill numbers (see Section 18.2) by Chao *et al.* (2014). Inevitably, there will be statistical advances that alter these details. The fundamental message, though, will remain the same – namely, that studies of species richness should always ensure that the comparisons being made, or associations being sought, are based not on sampling artefacts but on genuine differences in richness.

19.2 A simple model of species richness

Assume, for simplicity, that the resources available to a community can be depicted as a one-dimensional continuum, R units long (Figure 19.3). Each species uses only a portion of this resource continuum, and these portions define the *niche breadths* (n) of the various species; the average niche breadth within the community is $\bar{n}$. Some of these niches overlap, and the overlap between adjacent species can be measured by a value o. The average niche overlap within the community is then $\bar{o}$. With this simple background, it is possible to consider why some communities should contain more species than others.

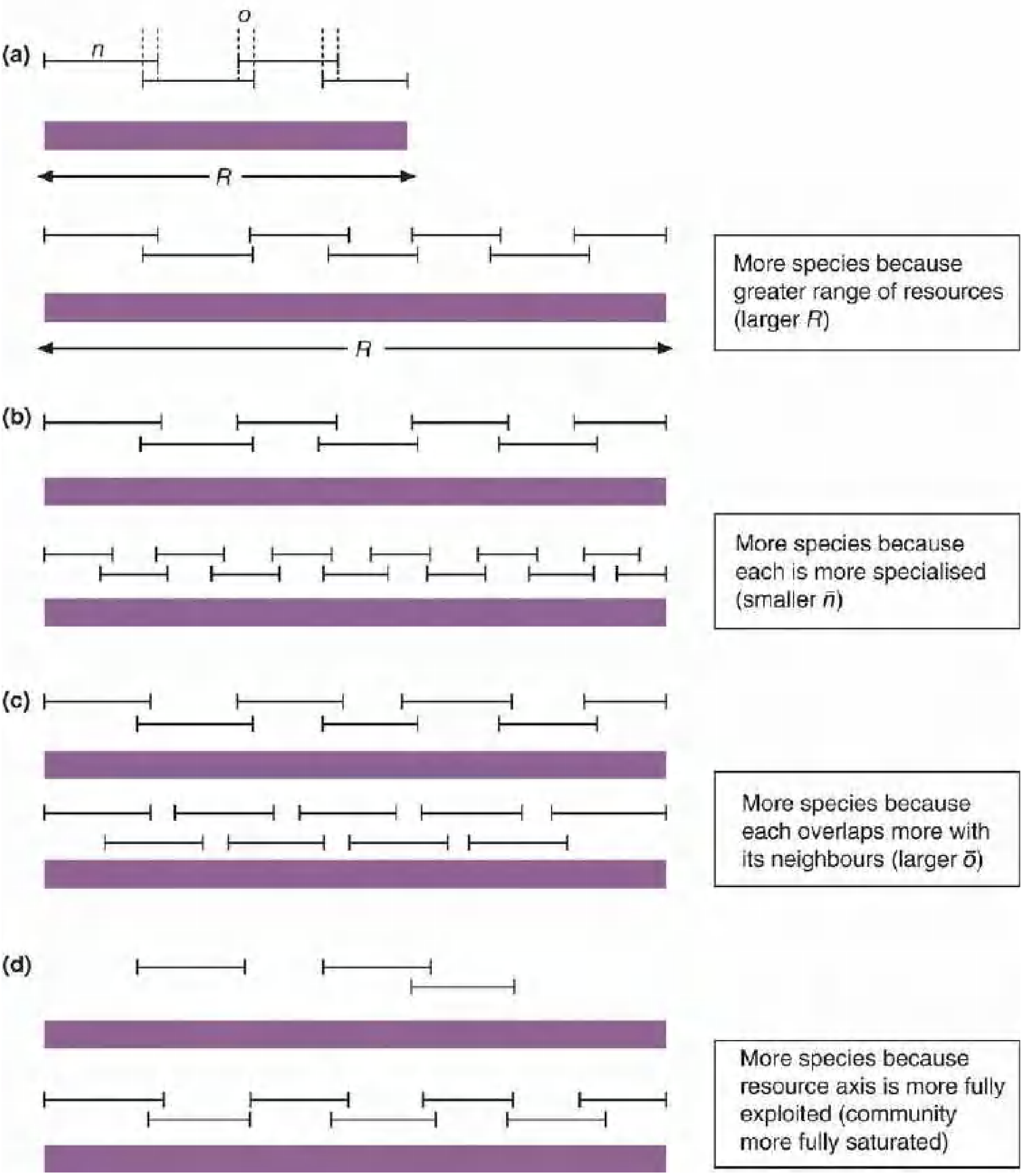


Figure 19.3 A simple model of species richness. Each species utilises a portion n of the available resources (R), overlapping with adjacent species by an amount o. More species may occur in one community than in another (a) because a greater range of resources is present (larger R), (b) because each species is more specialised (smaller average n), (c) because each species overlaps more with its neighbours (larger average o), or (d) because the resource dimension is more fully exploited.

Source: After MacArthur (1972).

a model incorporating niche breadth, niche overlap and resource range

First, for given values of $\bar{n}$ and $\bar{o}$, a community will contain more species the larger the value of *R*, i.e. the greater the range of resources (Figure 19.3a). Second, for a given range of resources, more species will be accommodated if $\bar{n}$ is smaller, i.e. if the species are more specialised in their use of resources (Figure 19.3b). Further, if species overlap to a greater extent in their use of resources (greater $\bar{o}$), then more may coexist along the same resource continuum (Figure 19.3c). Finally, a community will contain fewer species when more of the resource continuum is unexploited (Figure 19.3d).

species interactions

Immediately, this model allows us to organise the effects of species interactions on richness that we discussed in Chapter 16. If a community is dominated by interspecific competition, the resources are likely to be fully exploited. Species richness will then depend on the range of available resources, the extent to which species are specialists and the permitted extent of niche overlap (Figures 19.3a–c), but intense competition, by excluding species, is likely to reduce richness.

Predation, on the other hand, is capable of exerting contrasting effects. First, predators can exclude certain prey species. In the absence of these species, the community may then be less than fully saturated, reducing species richness. Second, predation may tend to keep species below their carrying capacities for much of the time, reducing the intensity and importance of direct interspecific competition for resources. This may then permit much more niche overlap and a greater richness of species than in a community dominated by competition (Figure 19.3c). Lastly, predation may generate richness patterns similar to those produced by competition when prey species compete for ‘enemy-free space’ (see Chapter 8).

Facilitation, too, may allow species to persist with narrow niches (Figure 19.2b) or greater niche overlap with other species (Figure 19.2c) than would otherwise be possible, increasing species richness.

Finally, disturbances, if frequent and severe, may remove species from a community, leaving it less than fully saturated and with reduced species richness (Figure 19.3d). However, at somewhat lower frequencies and severities, disturbance may prevent competitive exclusion in much the same way as predation (Figure 19.3c), leading to peak species richnesses at intermediate levels of disturbance (see Section 18.6.2).

19.3 Spatially varying factors that influence species richness

We can subdivide into two separate hypotheses the general idea that spatial (geographic) variations in climate play a critical role in determining species richness. The first is the productivity hypothesis (Section 19.3.1). This emphasises the importance of climate in determining productivity at the lowest trophic level and the resources these then provide further up the food chain. The alternative, the energy hypothesis, emphasises the direct role of energy (often measured by environmental temperature) on organisms throughout the community (Section 19.3.2).

19.3.1 Productivity and resource richness

For plants, the productivity of the environment can depend on whichever nutrient or condition is most limiting to growth (dealt with in detail in Sections 20.3 and 20.4). Broadly speaking, the productivity of the environment for animals follows the same trends as for plants, both as a result of the changes in resource levels at the base of the food chain, and as a result of the changes in critical conditions such as temperature.

If higher productivity is correlated with a wider range of available resources, then this is likely to lead to an increase in species richness (see Figure 19.3a). However, a more productive environment may have a higher rate of supply of resources but not a greater variety of resources. This might lead to more individuals per species rather than more species. Alternatively again, it is possible, even if the overall variety of resources is unaffected, that rare resources in an unproductive environment may become abundant enough in a productive environment for extra species to be added, because more specialised species can be accommodated (see Figure 19.3b). In general, then, there are rather straightforward grounds for expecting that species richness should increase with productivity.

increased productivity might lead to increased richness ...

Certainly there have been examples appearing to support this contention. The species richness of fish in North American lakes, for example, increases with an increase in productivity of the lake's phytoplankton (Figure 19.4a). There are also strong positive correlations between species richness and precipitation for both seed-eating ants and seed-eating rodents in the south-western deserts of the USA (Figure 19.4b), where it is well established that mean annual precipitation is closely related to plant productivity and thus to the amount of seed resource available. In this latter case, it is particularly noteworthy that in species-rich sites, the communities contain more species of very large ants (which consume large seeds) and more species of very small ants (which take small seeds). It seems that either the range of seed sizes is greater in the more productive environments or the abundance of seeds becomes sufficient to support extra consumer species with narrower niches. A strong positive relationship also exists between total species richness of plants in a set of large terrestrial ecoregions (defined as having distinctive biotas) and net primary productivity estimated for each ecoregion (Figure 19.4c).

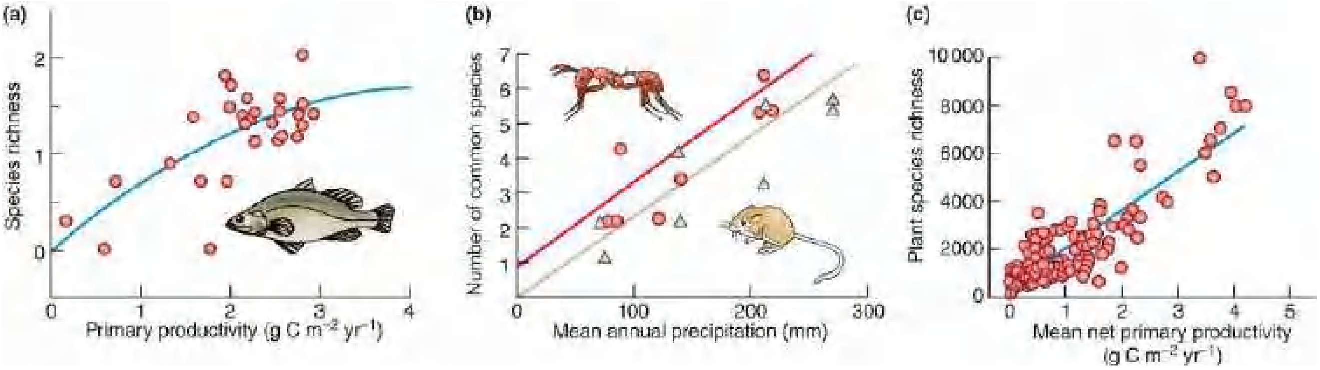


Figure 19.4 Species richness increases with productivity in fish, ants and rodents, and plants. The best-fit lines are statistically significant. (a) Species richness of fish increased with primary productivity of phytoplankton in a series of North American lakes. (b) The species richness of seed-eating rodents (yellow triangles) and ants (red circles) inhabiting sandy soils in south-west USA increased along a geographic gradient of increasing precipitation and, therefore, of increasing productivity. (c) The species richness of plants in 115 large ecoregions (>300 000 km^2) distributed across the globe increased in relation to modelled net primary productivity (based on remotely sensed estimates of photosynthetically active radiation at the vegetation surface).

Source: (a) After Dodson *et al.* (2000). (b) After Brown & Davidson (1977). (c) After McBride *et al.* (2014).

... or decreased richness ...

On the other hand, an increase in diversity with productivity is by no means universal. We saw in Figure 18.4, for example, in the long-term experiment at Rothamsted, that fertilised areas showed a progressive decline in species richness (and diversity). Similarly, a survey of plants in rich-fen sites in England and Wales (wetland sites where the water table was at or slightly above the substratum and with a high proportion of plants that thrive in lime-rich soils) showed clear evidence that species richness was lowest where productivity was greatest (Figure 19.5).

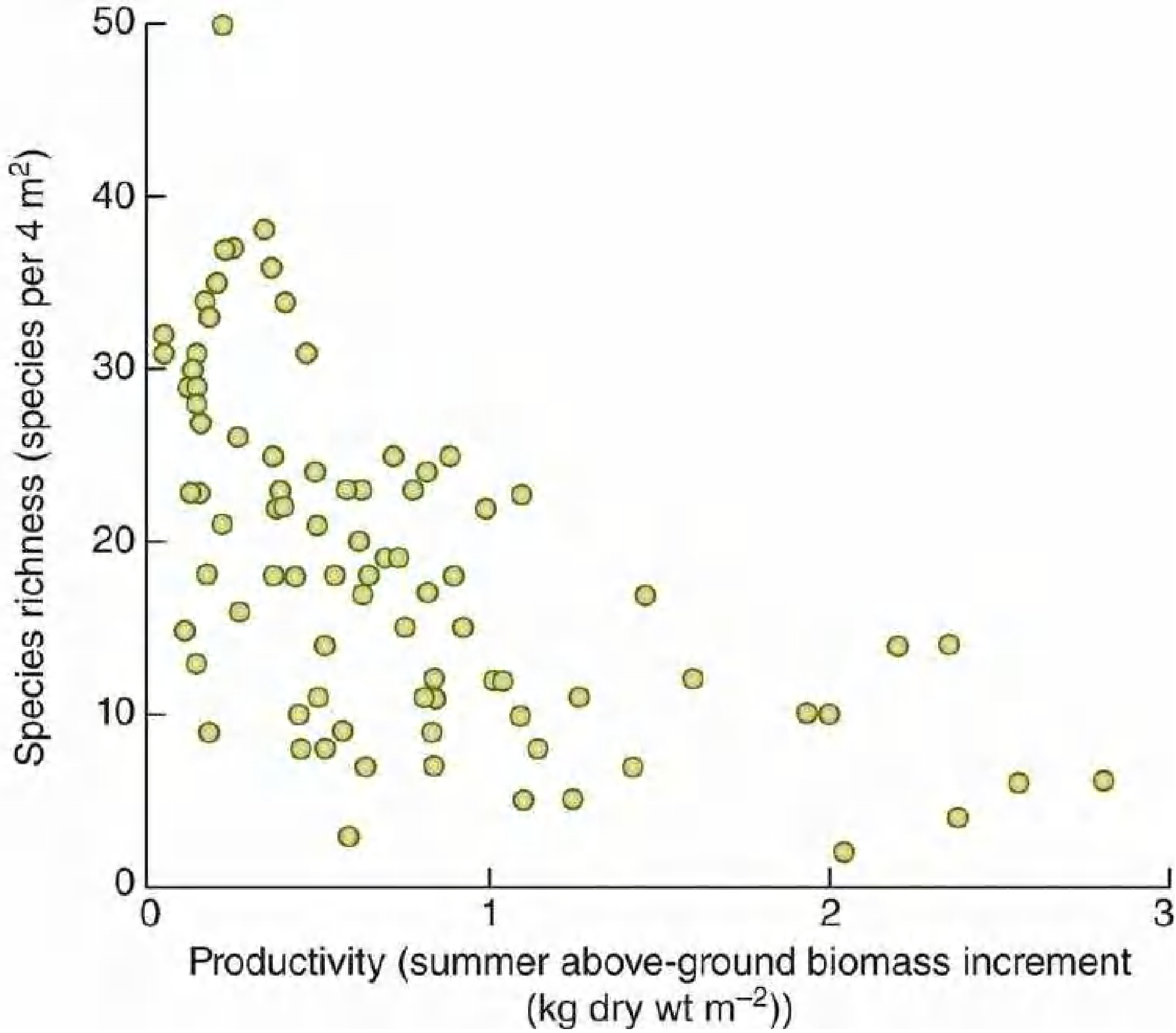


Figure 19.5 Species richness decreases with productivity in British plants. The relationship between species richness and productivity for rich-fen communities of plants in England and Wales, as evidenced by samples from 4 m^2 quadrats.

Source: After Wheeler & Shaw (1991).

Indeed, an association between high productivity and low species richness has often been found in plant communities (reviewed by Cornwell & Grubb, 2003). It can be seen, too, where human activities lead to an increased input of plant resources like nitrates and phosphates into lakes, rivers, estuaries and coastal marine regions. When such ‘cultural eutrophication’ is severe, we consistently see a decline in the species richness of phytoplankton (despite an increase in their productivity). Rosenzweig (1971) referred to such declines as illustrating ‘the paradox of enrichment’. One possible resolution of the paradox is that high productivity leads to high rates of population growth, bringing about the extinction of some of the species present because of a speedy conclusion to any potential competitive exclusion (see Chapter 8). With lower productivity, the environment is more likely to have changed before competitive exclusion is achieved.

APPLICATION 19.2 Resolving conflicting requirements of agriculture versus conservation

Species richness is generally much lower in agricultural land than in nearby nature reserves, despite its high productivity, in part because of the targeted use of pesticides, but also because of the effect of nutrient enrichment associated with intensive farming. There is thus a conflict between the objectives of maximising agricultural productivity and minimising biodiversity loss. Deciding how best to resolve these conflicting requirements is itself an important issue in conservation. One example comes from the need to conserve butterflies in England but also for farmers to produce food from the same areas of land. Food can be produced conventionally, with the aid of artificial pesticides and fertilisers, or organically, without them. Yields tend to be lower on organic farms (although not by as much as is sometimes suggested, see Section 15.2.5), so more land is needed to produce the same crop yield. However, organic farms provide better habitat for butterflies, though they are not so good for butterfly conservation as nature reserves themselves (Figure 19.6a).

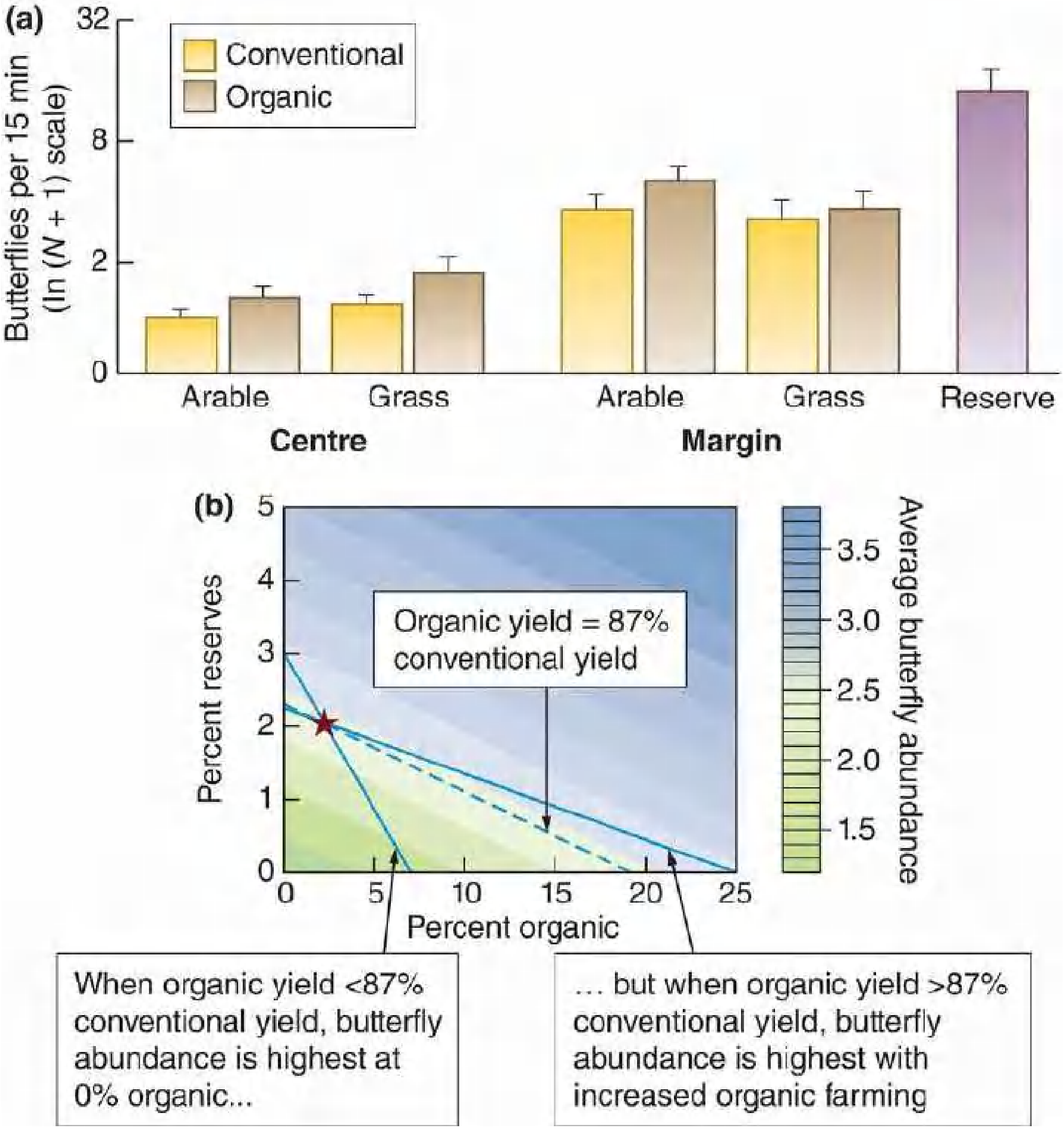


Figure 19.6 Butterfly densities are highest on nature reserves, but higher on organic than on conventional farms, but balancing conservation and farming needs depends on the relative yields of organic and conventional farms. (a) The effect of land use on butterfly density at a range of sites in England. Sites were classified as either arable or grass, and at either the centre or margins of fields, in either conventional or organic farms, or as nature reserves. Bars are SEs. (b). Land sharing or land sparing? The shading represents contours in the density of butterflies (individuals seen per 15 minutes) to be expected at different combinations of percentages of nature reserves and organic farmland in the habitat. The star represents the estimated current situation. The three blue lines represent combinations giving the same farmland yield, assuming different relationships between the yields on organic and conventional farms: namely, organic yield less than, equal to or more than 87% of conventional yield. The best solution for each of these solutions is found at the point on its line that is highest on the density contours.

Source: After Hodgson *et al*. (2010).

‘land sharing’ or ‘land sparing’

A choice therefore exists between ‘land sharing’ – farming organically and so producing food and protecting butterflies on the same land – and ‘land sparing’: farming intensively to produce food in some areas while conserving intensively in others such as nature reserves. The right choice depends not only on the relative values for conservation of different land uses, shown in Figure 19.6a, but also on the relative yields of organic and conventional farms.

The search for the best solution is shown in Figure 19.6b, which combines the two conflicting requirements. First, contours represent the density of butterflies to be expected at different combinations of nature reserves and organic farmland in the habitat. Second, there are lines representing combinations of equal overall farmland yield for different assumptions about relative yield on organic and conventional farms. Our starting point is the current situation in terms of the percentages of reserve, organic and conventional farmland, also shown in the figure. The question is: ‘How can we conserve the most butterflies while maintaining current yields?’ Our aim, therefore, is to choose the line that most accurately represents the organic/conventional balance of yield, and then find the point on that line that takes us as high up the density contours as possible, that is, that conserves the most butterflies.

The analysis in Figure 19.6b indicates that if organic yields are lower than 87% of those achieved on conventional farms, then conventional farming and sparing land for nature reserves is the better conservation option. However, if we can achieve organic yields that are more than 87% of those on conventional farms, then sharing production and conservation on the same organic land is best. Clearly, decisions about whether and where to establish protected areas depend on more than simply what those areas are able to protect. And indeed, decisions about organic or conventional farming depend on more factors than yields and conservation.

... or an increase then a decrease (hump-shaped relationships)

It is perhaps not surprising, then, that several studies have demonstrated both an increase and a decrease in richness with increasing productivity – that is, that species richness may be highest at intermediate levels of productivity. Species richness declines at the lowest productivities because of a shortage of resources, but also declines at the highest productivities where competitive exclusions speed rapidly to their conclusion. Such humped curves are seen, for instance, when the number of lake phytoplankton species is plotted against overall phytoplankton productivity (Figure 19.7a), for the same lakes that we have already seen in Figure 19.4a; it is seen when the species richness of desert rodents is plotted against precipitation (and thus productivity) along a geographic gradient in Israel (Figure 19.7b), in contrast to the relationship in Figure 19.4b; and it is seen when the species richnesses of both gastropod and bivalve mollusc species, living on the sea floor in deep-sea communities, are plotted against the level of particulate organic carbon (POC), the rain of chemical energy falling as dead organic matter from the sea surface.

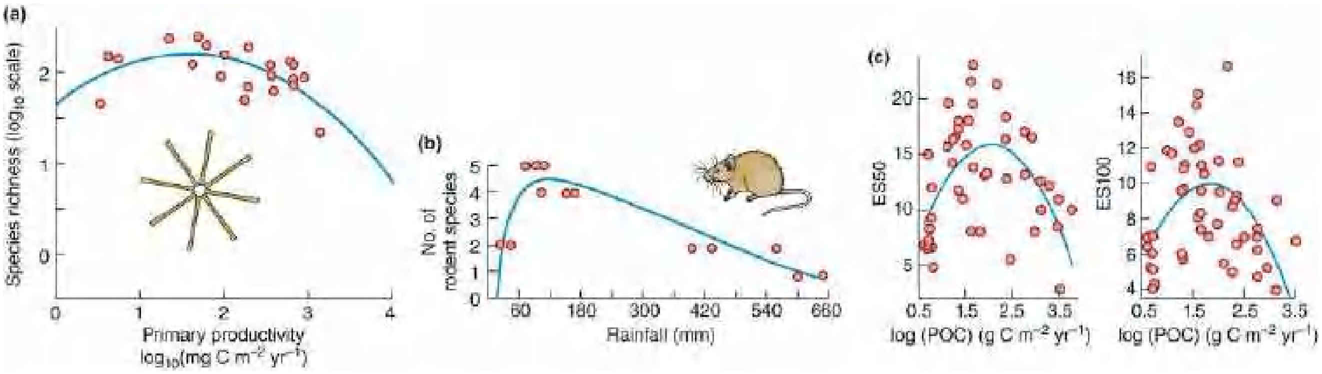


Figure 19.7 Humped relationships between species richness and productivity. The best-fit lines are statistically significant. (a) Species richness of phytoplankton at a range of productivities for the same lakes as for the fish data in Figure 19.4a. (b) Species richness of desert rodents in Israeli deserts plotted against annual rainfall. (c) Species diversity of gastropod molluscs, left, and bivalve molluscs, right (ES50 and ES100, the expected number of species normalised to a sample size of 50 or 100 individuals, respectively) in deep-sea communities, plotted against the flux of particulate organic carbon (POC) falling from above.

Source: (a) After Dodson *et al.* (2000). (b) After Abramsky & Rosenzweig (1983). (c) After Tittensor *et al.* (2011).

Indeed, a number of meta-analyses have been carried out of the wide range of studies of the richness–productivity relationship, from both terrestrial and aquatic systems, in a search for some general rules. These have been heavily criticised for the arguably uncritical way in which the original datasets have been accepted at face value and combined (Whittaker, 2010) – a problem that besets all such analyses, despite their attractions. Nonetheless, one conclusion that survives such criticism is the observation that a very wide range of relationships has been found: most are positive, but some are negative, some humped (unimodal), some have no detectable relationship at all, and some are even U-shaped, cause unknown (Figure 19.8). Clearly, increased productivity can and does lead to increased or decreased species richness, or both, or neither.

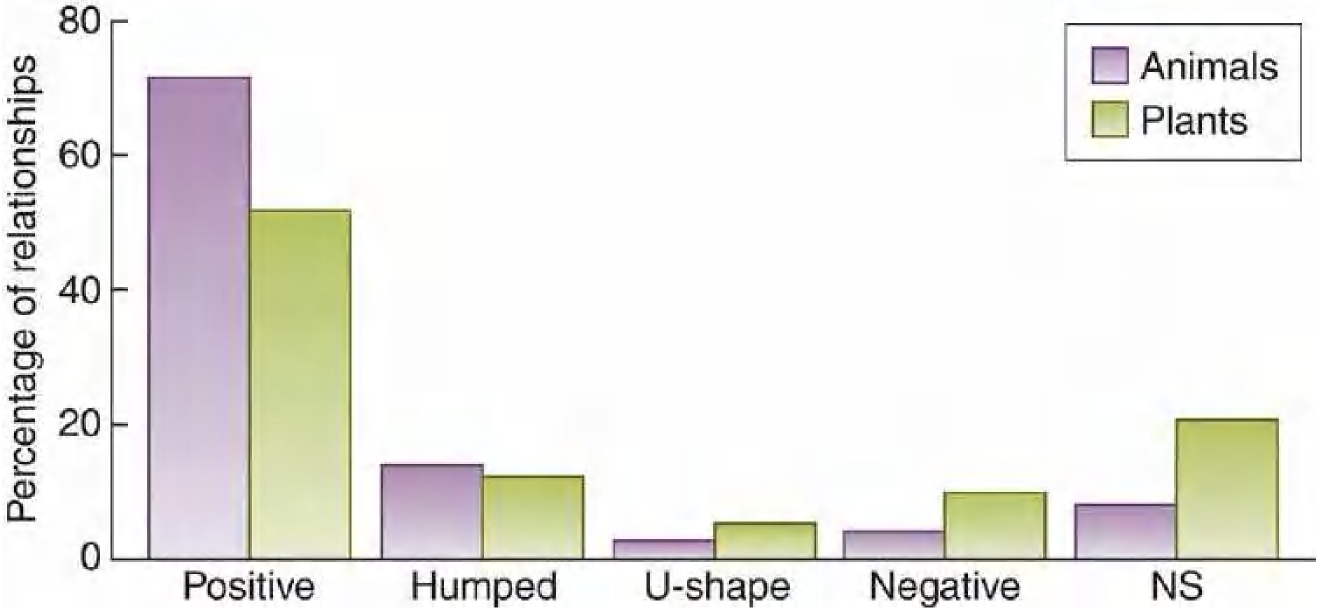


Figure 19.8 Studies of richness–productivity relationships show a range of outcomes. The percentage of published studies on animals (purple bars, n = 115) and plants (green bars, n = 58) showing various patterns in the relationship between species richness and productivity. NS, not significant.

Source: After Gillman *et al.* (2015).

Some of this variation amongst studies can be accounted for simply by assuming that in particular cases, only a proportion of the productivity spectrum was sampled, so that, for example, only the ascending or descending limb of a humped-shaped curve was observed. And where no relationship was observed, this may of course be because of shortcomings in the data collection. But a good case can also be made for the nature of the relationship changing, depending on the spatial scale at which it is observed. To make this overall pattern clearer, we can turn to the concepts of α-, β- and γ-diversity, discussed in Section 18.2. In particular, it has often been suggested that at the local scale, the relationship between α-diversity either declines with productivity or there is a hump-shaped relationship, whereas at the regional scale γ-diversity tends to increase with productivity (e.g. Whittaker, 2010). This suggestion is borne out by a study of ponds in south-western Michigan and north-eastern Pennsylvania, USA, where α-diversity focused on the relationship within ponds, while γ-diversity focused on whole catchment areas (i.e. regions) each of which was a collection of ponds. For both plants and animals, there was a hump-shaped relationship between α-diversity and pond productivity, but γ-diversity increased with the productivity of whole catchment areas (Figure 19.9a, b). This in turn suggests that the differences between communities within a region (β-diversity) must increase with productivity, and this, too, was borne out by the data (Figure 19.9c). The implication, clearly, is that the drivers of the richness–productivity relationship are different at different scales, and this, too, is a conclusion that it would be safe to accept.

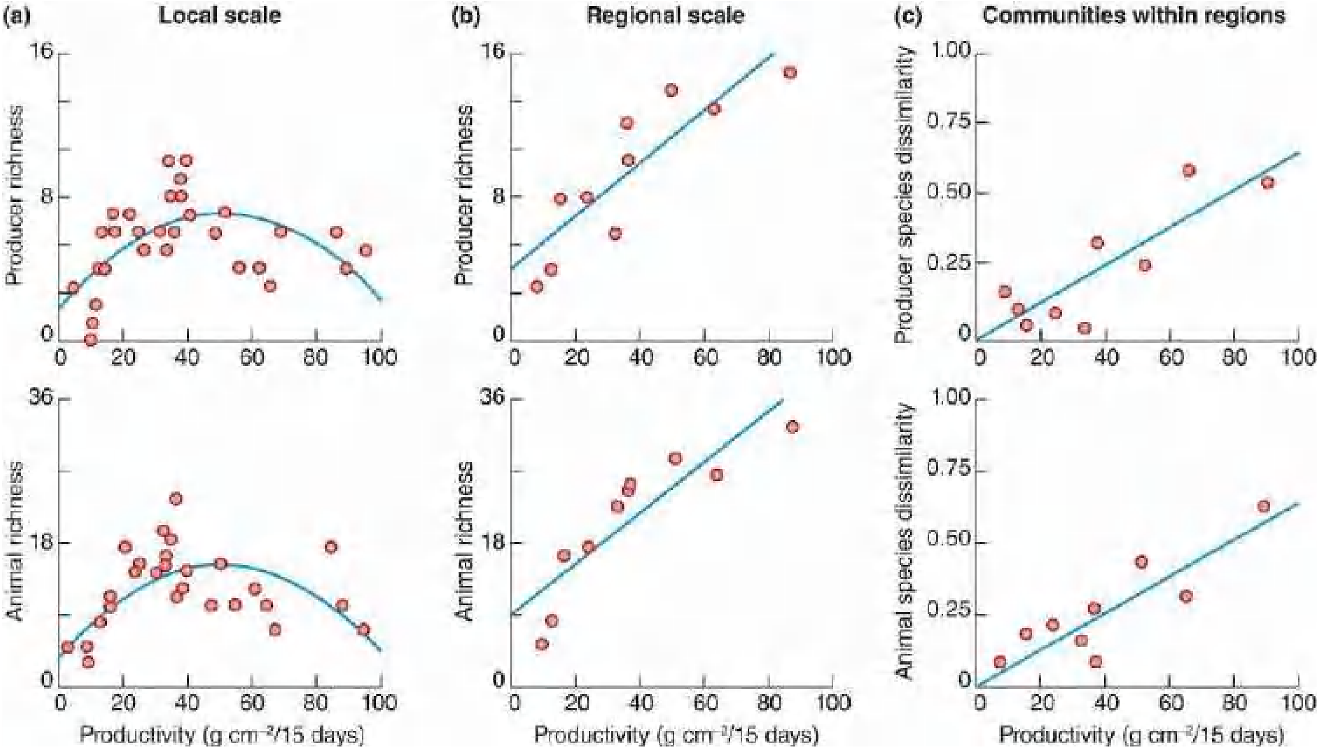


Figure 19.9 Diversity–productivity relationships for aquatic communities change with scale. Diversity–productivity relationships for ponds in south-western Michigan and north-eastern Pennsylvania, USA. Producers are shown in the top graphs (vascular plants and macroalgae); animals are shown in the graphs below (insects, crustaceans, amphibians, etc.). Productivity was estimated as rate of algal biomass accrual on artificial substrates. All lines are statistically significant. (a) α-diversity – the diversity within individual ponds. (b) γ-diversity – the diversity within regions: catchment areas containing several ponds. (c) β-diversity – the dissimilarity amongst ponds within catchment areas. Each point is the average dissimilarity (1 – ‘Jaccard’s’ similarity index) amongst ponds in one catchment area.

Source: After Chase & Leibold (2002).

One plausible, more detailed suggestion would be that at the very smallest scales, increased productivity leads to a decline in richness as competition excludes rare species for which there is no space (as for the small quadrats in Figure 19.5). At somewhat larger scales (e.g. the ponds in Figure 19.9), there is sufficient space for initial rises in productivity to enhance richness prior to the subsequent decline, and hence the whole hump-shaped relationship is apparent. But at larger scales still (e.g. the catchment areas in Figure 19.9 and the large ecoregions in Figure 19.4c), species that are excluded from one community in a region are likely to survive in other, somewhat different communities, such that no decline in richness at higher productivities is observed, and richness simply continues to rise as more exploitable niches become available for the regional species pool. Work in the future will hopefully tell us whether this suggestion is not only plausible but correct – or indeed, the whole story. Whatever the truth of the matter, the distinction between small, medium and large scales is certain to differ between different communities – smaller in a grassland, for example, than in a forest. As ever in ecology, we must remember to see the world from the organisms' point of view.

19.3.2 Energy

When we turn from productivity to energy, the argument for a direct effect of energy on species richness is essentially a metabolic argument – higher temperatures for longer periods mean more time for species to be active, higher rates of metabolism (more growth, more reproduction, more individuals, etc.) and less chance of cold and frost damage (Hawkins *et al.*, 2003). Certainly there are examples in which species richness for whole groups is correlated positively either with indicators of energy such as annual potential evapotranspiration (PET – the amount of water that would evaporate or be transpired from a saturated surface, and hence a measure of atmospheric energy; see the American birds and butterflies in Figure 19.10a, b), or simply with annual mean temperature itself (the Chinese vertebrates in Figure 19.10d).

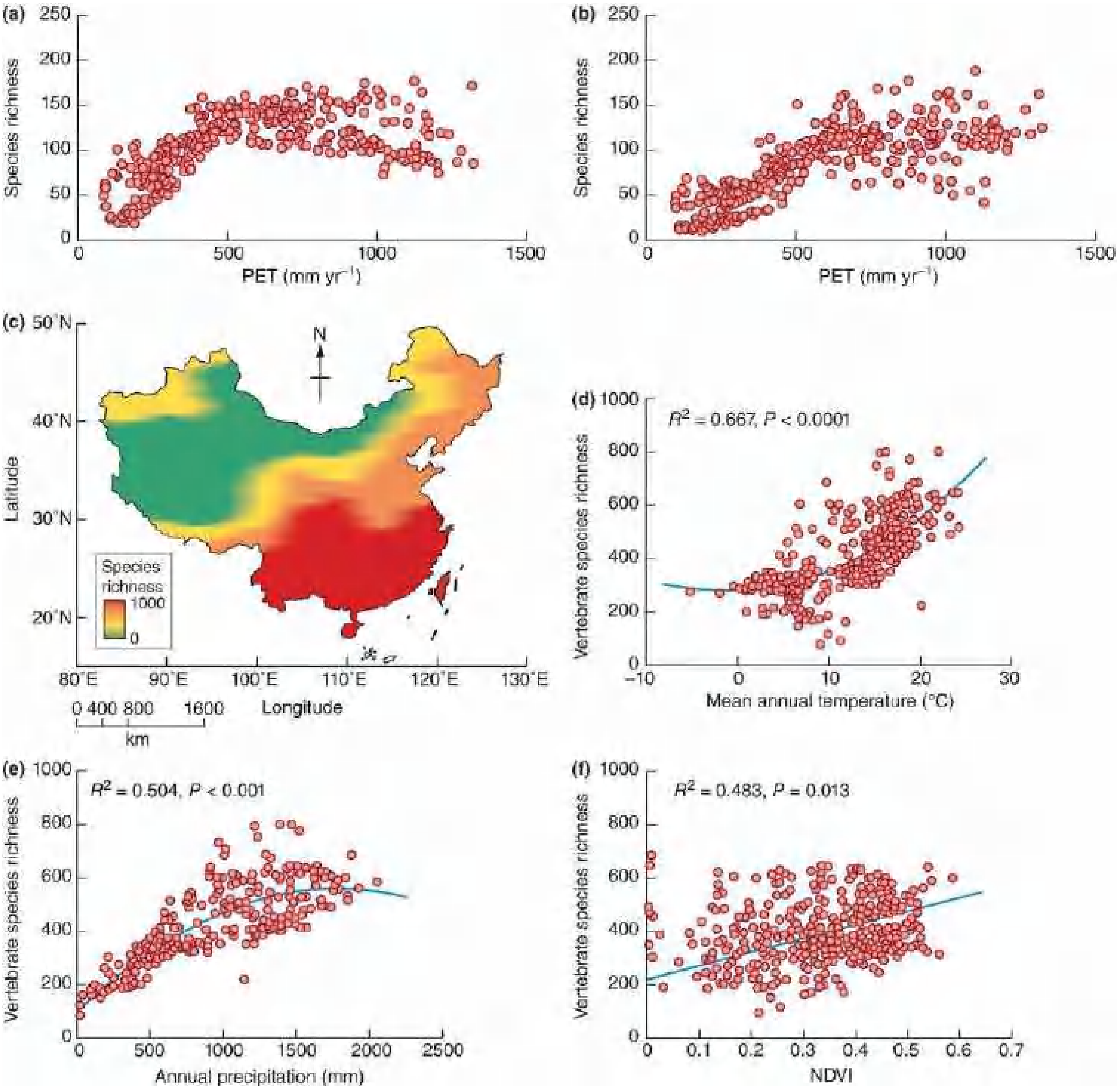


Figure 19.10 At broad scales, species richness increases with environmental energy but also with correlates of environmental energy. (a) The relationship between the species richness of birds in North America and potential evapotranspiration (PET), a measure of environmental energy, organised at the scale of 48 400 km^2 grid cells. (b) A similar relationship for North American butterflies. (c) Variation in the species richness of vertebrates over the whole of China organised at the scale of 10 000 km^2 square grid cells. (d) The relationship between the species richness of vertebrates in China, as in (c), and mean annual temperature. (e) The relationship between the species richness of vertebrates in China and mean annual precipitation. (f) The relationship between the species richness of vertebrates in China and NDVI (normalised difference vegetation index), a measure of vegetation productivity.

Source: (a, b) After Hawkins *et al.* (2003). (c–f) After Luo *et al.* (2012).

On the other hand, the apparently humped relationships in Figure 19.10a and b indicate that energy, as measured in this case by PET, is unlikely to be the whole story. O’Brien (2006), for example, argued that the combined effects of water and energy availability on plants determined their photosynthetic rates and productivity, with consequent effects at higher trophic levels. Thus, while productivity and richness are likely to increase in the left-hand limbs of Figure 19.10a and b, as energy increases and water is less often frozen (and hence more often available), they may also decline to the right in more arid environments where the highest energy levels lead to rapid and frequent evaporation of water. Field *et al.* (2005) provided empirical support for such patterns.

We must be careful, too, as always, not to mistake association for causation, especially given that energy is itself bound to be one of the drivers of the productivity associations discussed. We can see, for example, that the species richness of vertebrates in China is also correlated with annual rainfall and with an index of vegetation productivity (Figure 19.10e, f). Is only one of these the 'true' driver of variations in species richness? Or do all three have some part to play?

To try to disentangle the separate effects of energy and productivity, each in their own right, Belmaker and Jetz (2011) compiled large datasets on amphibians, birds and mammals. They took the statistical approach of determining the effect of net primary productivity (NPP; see Chapter 20) after having taken account of variations in energy (mean annual temperature) (Figure 19.11a), and conversely determining the effect of energy after having taken account of variations in NPP (Figure 19.11b). These relationships were also studied at a range of scales. The largest involved comparing assemblages encompassed within circles 2000 km in diameter. At the smallest scale, local assemblages were compared (median area 488 km^2, equivalent to a circle 22 km in diameter). Between these were assemblages encompassed within circles 200 km in diameter.

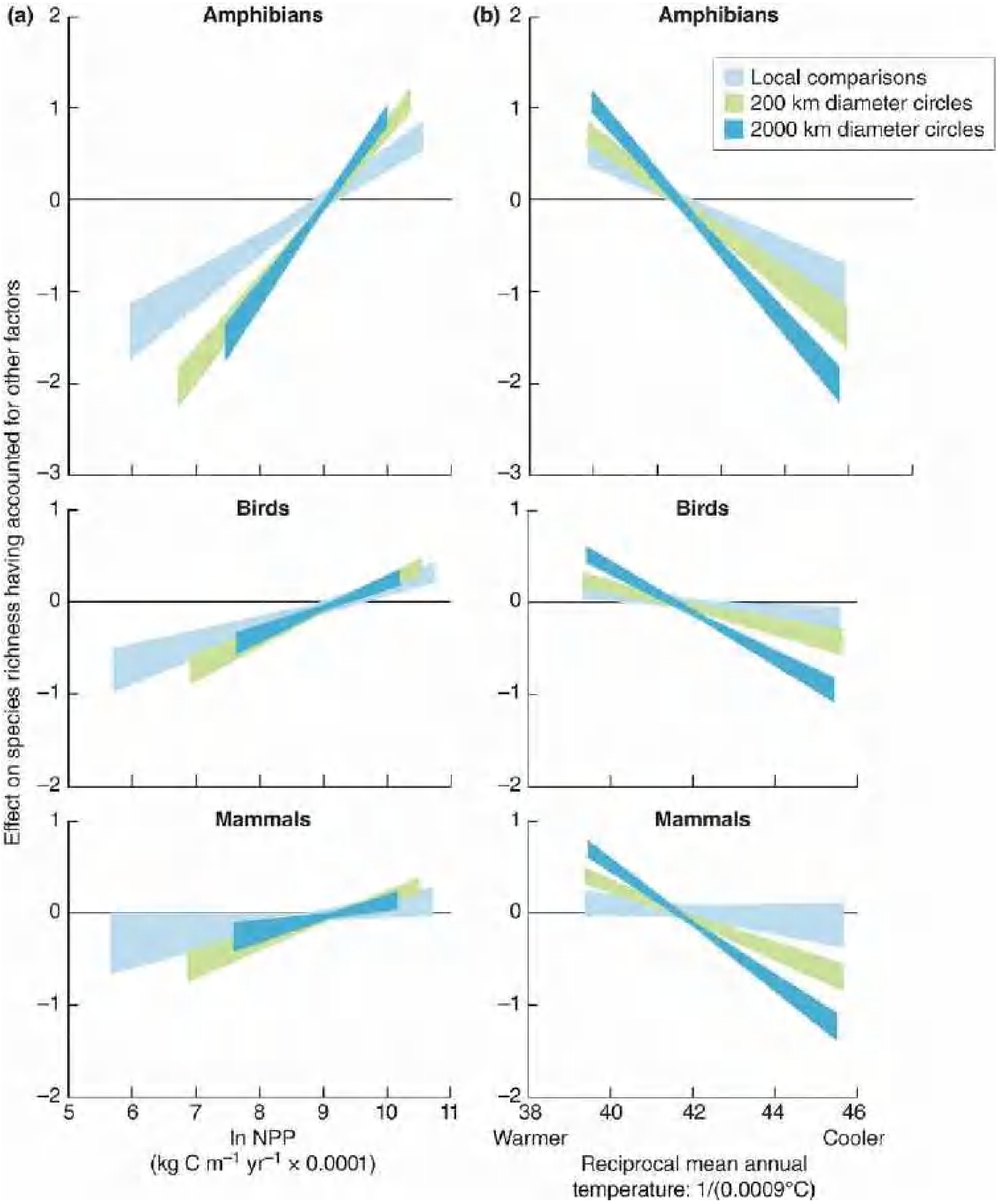

(a)
(b)
Amphibians
Birds
Mammals
Local comparisons
200 km diameter circles
2000 km diameter circles
Effect on species richness having accounted for other factors
2
1
0
−1
−2
−3
5
6
7
8
9
10
11
ln NPP
(kg C m^{-1} yr^{-1} × 0.0001)
38
40
42
44
46
Warmer
Cooler
Reciprocal mean annual temperature: 1/(0.0009°C)

Figure 19.11 Species richness increases with both productivity and energy input, but the strengths of these relationships vary with spatial scale. (a) Relationships and 95% CIs for the regressions of species richness of amphibians, birds and mammals on net primary productivity (NPP), having taken effects of temperature on species richness into account in the statistical analyses. Data are collected from throughout the world and are calculated on the basis of comparisons at a range of scales, as indicated. (b) Similar, but for energy, measured as a reciprocal mean annual temperature so scales are comparable, having taken effects of NPP on species richness into account in the statistical analyses.

Source: After Belmaker & Jetz (2011).

It is apparent, first, that both energy and productivity retain their associations with species richness even after the other has been taken into account. It is also apparent, though, that the strengths of these relationships are strongly dependent on the scales at which they are studied. In particular, the associations for energy were strongest of all at the largest scale, where energy seemed, if anything, to be more important than productivity, whereas at smaller scales associations with energy were generally much weaker and in some cases non-existent (Figure 19.11b). By contrast, the relationships with productivity were much less sensitive to changes of scale. While they were less strong than the energy relationships at the largest scale, they were not much altered at the local scale (Figure 19.11b).

Thus energy, beyond its effects on productivity, seems to have an important role to play in helping us understand patterns of species richness at the global scale. Perhaps it sets an upper limit on richness, irrespective of the variations in other factors discussed in this chapter. But the role of productivity, as we discussed earlier, reflecting many things other than energy, seems also to be important at a whole range of scales. We should remember, for example, that there are highly species-rich communities at the bottom of the deep oceans where there is no light and temperatures are low. In such situations, energy input and productivity must be very largely decoupled from one another.

19.3.3 Spatial heterogeneity

We have already seen how the patchy nature of an environment, coupled with aggregative behaviour, can lead to coexistence of competing species (see Section 8.7.4). In addition, environments that are more spatially heterogeneous can be expected to accommodate extra species because they provide a greater variety of microhabitats, a greater range of microclimates, more types of places to hide from predators and so on. In effect, the extent of the resource spectrum is increased (see Figure 19.3a). A meta-analysis of 192 studies confirmed this expectation, for a variety of taxa throughout the world (Figure 19.12a). It was true, too, across five categories of heterogeneity, two of them biotic – land cover variation (the patchiness of environments) and vegetational variation (structural, taxonomic or functional diversity) – and three of them abiotic: climatic, topographic and soil type variation (Figure 19.12b). The effects were particularly strong for vegetational variation, the majority of which related to physical rather than taxonomic heterogeneity, and for topographic variation (bigger effect sizes in Figure 19.12b). Also, unsurprisingly, effects of heterogeneity were more likely to be detected in studies carried out with larger plot sizes, which tend to offer more opportunity for heterogeneity.

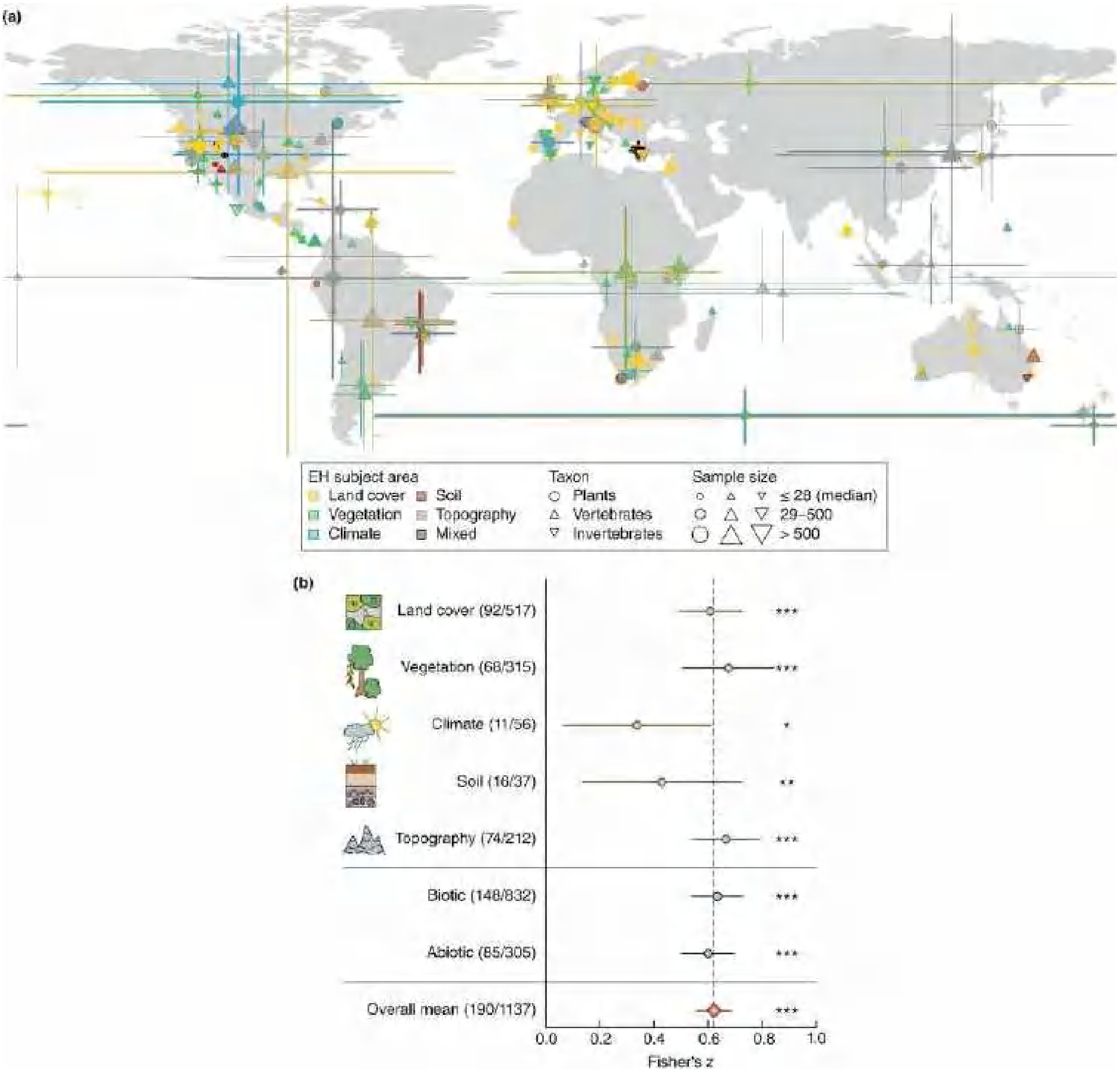


Figure 19.12 Species richness increases with the spatial or structural heterogeneity of the environment. (a) Locations of studies of the relationship between species richness and spatial environmental heterogeneity (EH), classified by subject area (colour), taxonomic group (symbol) and sample size (symbol size), as indicated. The lines indicate the latitudinal and longitudinal extents of the studies. Overlapping symbols are offset for clarity. (b) A meta-analysis of the effect sizes found in these studies, measured by Fisher's *z*-statistic, decomposed according to the type of heterogeneity (brown circles) and whether this was biotic or abiotic (blue circles). The values in parentheses are the number of studies and data points in each category. Lines are 95% CIs. The red diamond and dashed black line indicate the overall mean. *, ** and *** indicate $P < 0.05$, $P < 0.01$ and $P < 0.001$, respectively.

Source: After Stein *et al*. (2013).

19.3.4 Environmental harshness

what is harsh?

Environments dominated by an extreme abiotic factor – often called harsh environments – are more difficult to recognise than might be immediately apparent. From our own point of view, we might describe as extreme both very cold and very hot habitats, unusually alkaline lakes and grossly polluted rivers. However, species have evolved and live in all such environments, and what is very cold and extreme for us must seem benign and unremarkable to a penguin in the Antarctic.

We might try to get around the problem of defining environmental harshness by letting the organisms decide. An environment may then be classified as extreme if organisms, by not living there, show it to be so. But if the claim is to be made – as it often is – that species richness is lower in extreme environments, then this definition is circular, designed to prove the very claim we wish to test.

Perhaps the most reasonable definition of an extreme condition is one that requires, of any organism tolerating it, a morphological structure or biochemical mechanism that is not found in most related species, and is costly, either in energetic terms or in terms of the compensatory changes in the organism's biological processes that are needed to accommodate it. For example, plants living in highly acidic soils (low pH) may be affected directly through injury by hydrogen ions or indirectly via deficiencies in the availability and uptake of important resources such as phosphorus, magnesium and calcium. In addition, aluminium, manganese and heavy metals may have their solubility increased to toxic levels, and mycorrhizal activity and nitrogen fixation may be impaired. Plants can only tolerate a low pH if they have specific structures or mechanisms allowing them to avoid or counteract these effects.

are harsh environments the cause of low species richness?

Environments that experience a low pH can thus be considered harsh, and the mean number of plant species recorded per sampling unit in a study in the Alaskan Arctic tundra was indeed lowest in soils of low pH (Figure 19.13a). Similarly, the species richness of benthic stream invertebrates in the Ashdown Forest (southern UK) was markedly lower in the more acidic streams (Figure 19.13b). Further examples of extreme environments that are associated with low species richness include hot springs, caves and highly saline water bodies such as the Dead Sea. The problem with these examples, however, is that they are also characterized by other features associated with low species richness such as low productivity and low spatial heterogeneity. In addition, many occupy small areas (caves, hot springs) or areas that are rare compared with other types of habitat (only a small proportion of the streams in southern England are acidic). Hence extreme environments can often be seen as small and isolated islands. We see elsewhere in this chapter that these features, too, are usually associated with low species richness. It appears reasonable that intrinsically extreme environments should as a consequence support few species, but this has proved a difficult proposition to establish.

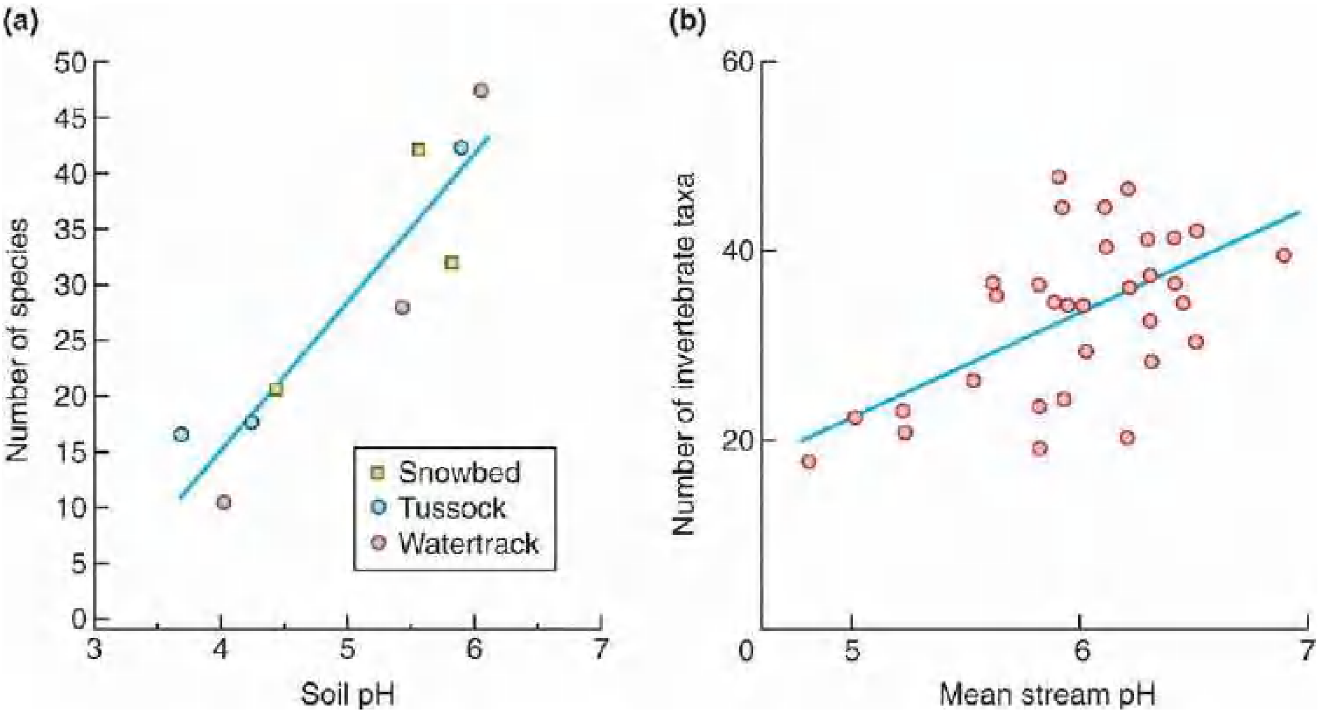


Figure 19.13 Species richness is lower in 'harsher' (lower pH) environments. (a) The number of plant species per 72 m^2 sampling unit in the Alaskan Arctic tundra increases with pH ($P < 0.001$). (b) The number of taxa of invertebrates in streams in Ashdown Forest, southern England, increases with the pH of the streamwater ($P < 0.005$).

Source: (a) After Gough *et al.* (2000). (b) After Townsend *et al.* (1983).

19.4 Temporally varying factors that influence species richness

Temporal variation in conditions and resources may be predictable or unpredictable. The distinction is an important one. Predictable variation often occurs on a timescale similar to the generation times of the species themselves. This is the case for seasonal climatic variation, discussed next, and is thus a pattern to which the organisms can adapt, usually by adopting schedules in their own metabolism that parallel the seasonal cycle. On the other hand, the predictable absence of seasonal variation may allow specialised adaptation, without the threat of local extinction that a marked change in conditions would provoke. By contrast, unpredictable changes in the habitat are most naturally viewed as a disturbance to a community that changes its composition, following which the community may gradually revert to its predisturbance state. Such disturbances were discussed in Section 18.7.

19.4.1 Climatic variation

temporal niche differentiation in seasonal environments

In a predictable, seasonally changing environment, different species may be suited to conditions at different times of the year. More species might therefore be expected to coexist in a seasonal environment than in a completely constant one (see Figure 19.3b). Different annual plants in temperate regions, for instance, germinate, grow, flower and produce seeds at different times during a seasonal cycle. Similarly, phytoplankton and zooplankton pass through a seasonal succession in large, temperate lakes with a variety of species dominating in turn as changing conditions and resources become suitable for each.

specialisation in non-seasonal environments

On the other hand, there are opportunities for specialisation in non-seasonal environments that do not exist in seasonal environments. For example, it would be difficult for a long-lived obligate fruit-eater to exist in a seasonal environment when fruit is available for only a very limited portion of the year. But such specialisation is found repeatedly in non-seasonal, tropical environments where fruit of one type or another is available continuously.

Broadly, studies tend to support the second of these two suggestions – that species richness increases as climatic variation *de*creases. For example, as we move along the west coast of North America from Panama in the south to Alaska in the north, there is a significant decrease in the species richness of birds, mammals and gastropods as the range of monthly mean temperatures steadily increases (MacArthur, 1975). However, this correlation does not of course prove causation, since there are many other things that change between Panama and Alaska. Thus, there is no established relationship between climatic constancy, as such, and species richness.

More supportive, perhaps, is evidence from species-rich biomes in especially unvarying environments, where neither high productivity, nor high energy inputs, nor any of the other factors usually proposed provide a ready explanation. For example, communities on the abyssal seafloor, 3000-6000 m below the sea surface, are characterised by an absence of light and low temperatures (–0.5 to +3.0°C), and are dependent for a food resource on detritus falling from the surface communities above. Nonetheless, various studies have reported abyssal sediments containing around 50 different species of polychaete worm for every 150 individuals sampled, 100 non-microbial species per 0.25 m^2, and frequently very high levels of β-diversity – different species found in different local communities (Smith *et al.*, 2008). The unvarying nature of these communities, buffered from the fluctuations in the habitats high above them, are likely to have favoured both narrow specialisation and local speciation. Similarly, the subtropical oceanic gyres – large masses of semi-isolated water in the Pacific and Atlantic Oceans, surrounded by a circular current of water moving around them – have low productivity but high species richness; a single trawl may bring up 50 different species of fish from their open waters (Barnett, 1983). The biome, though, is characterised by remarkable spatial and temporal invariance.

19.4.2 Environmental age: evolutionary time

unchanging tropics and recovering temperate zones?

It has also often been suggested that communities that are ‘disturbed’ even on very extended timescales may nonetheless lack species because they have yet to reach an ecological or an evolutionary equilibrium. Thus communities may differ in species richness because some are closer to equilibrium and are therefore more saturated than others (see Figure 19.3d).

For example, many have argued that the tropics are richer in species than are more temperate regions at least in part because the tropics have existed over long and uninterrupted periods of evolutionary time, whereas the temperate regions are still recovering from the Pleistocene glaciations when the climatic and biotic zones of the temperate region moved toward the equator. A counter argument has been that a stark contrast between unchanging tropics and disturbed and recovering temperate regions is overly simplistic, and that during the glaciations, the tropical forest, particularly in Amazonia, may have contracted to a limited number of small refuges surrounded by grasslands. However, this itself has been disputed – see Hoorn *et al.* (2010) for a longer term perspective on the genesis of Amazonian biodiversity.

A comparison between the two polar regions may be more instructive. Both Arctic and Antarctic marine environments are cold, seasonal and strongly influenced by ice but their histories are quite different. The Arctic basin lost its fauna when covered by thick permanent ice at the height of the last glaciation and recolonisation is underway; whereas a shallow water fauna has existed around the Antarctic since the mid-Paleozoic (Clarke & Crame, 2003). Today the two polar faunas contrast markedly, the Arctic being depauperate and the Antarctic rich, most likely reflecting the importance of their histories.

19.5 Habitat area and remoteness: island biogeography

larger islands contain more species

It is well established that the number of species on islands decreases as island area decreases. Such a *species–area relationship* is shown, for example, in Figure 19.14a for plants on small islands east of Stockholm, Sweden. Islands, however, need not be islands of land in a sea of water. Lakes are islands in a 'sea' of land, mountaintops are high-elevation islands in a low-elevation ocean, gaps in a forest canopy where a tree has fallen are islands in a sea of trees, and there can be islands of particular geological types, soil types or vegetation types surrounded by dissimilar types of rock, soil or vegetation. Species–area relationships can be equally apparent for these types of islands (Figure 19.14b, c).

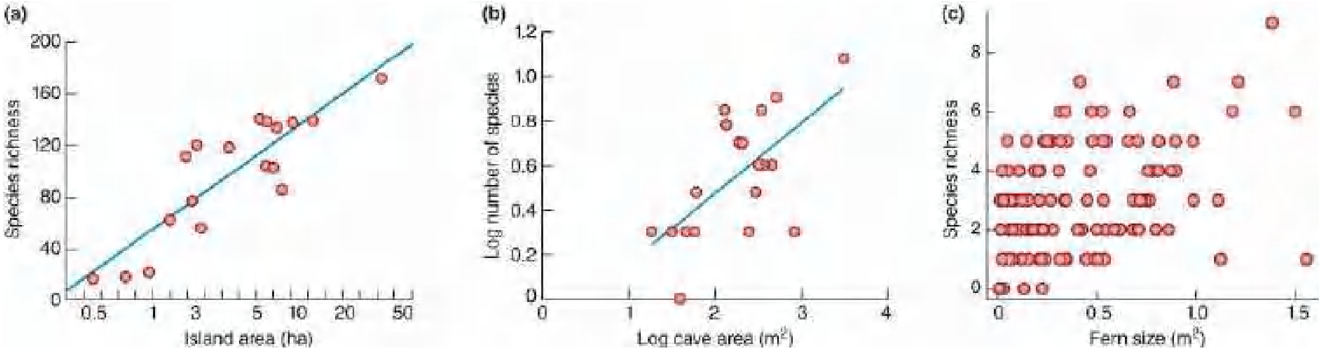


Figure 19.14 Species–area relationships showing species richness increasing with 'island' size. (a) Plants on islands east of Stockholm, Sweden ($P < 0.05$). (b) Bats inhabiting different-sized caves in Mexico ($P < 0.05$). (c) Plant communities living inside the epiphytic bird's nest fern, *Asplenium goudeyi*, on Lord Howe Island between Australia and New Zealand. (Generalised linear model, $P < 0.001$, having accounted for variations in fern isolation.)

Source: (a) After Lofgren & Jerling (2002). (b) After Brunet & Medellin (2001). (c) After Taylor & Burns (2015).

The relationship between species richness and habitat area is one of the most consistent of all ecological patterns. But is this impoverishment of species on islands more than we would expect in comparably small areas of mainland? Or to look at it the other way round, does the characteristic isolation of islands contribute to their impoverishment of species? These are important questions for an understanding of community structure, since there are many oceanic islands, many lakes, many mountaintops, many woodlands surrounded by fields, and so on. Moreover, protected areas are islands within an 'ocean' of unprotected habitat.

19.5.1 MacArthur and Wilson's 'equilibrium' theory

Probably the most obvious reasons why larger areas should contain more species are, first, that the smallest areas are simply not large enough to support viable populations of many species, and second, that larger areas typically encompass more different types of habitat. However, MacArthur and Wilson (1967) believed these explanations to be too simple. In their *Equilibrium Theory of Island Biogeography*, they argued: (i) that island size and isolation themselves played important roles – that the number of species on an island is determined by a balance between immigration and extinction; (ii) that this balance is dynamic, with species continually going extinct and being replaced (through immigration) by the same or by different species; and (iii) that immigration and extinction rates may vary with island size and isolation.

MacArthur and Wilson’s immigration curves

Taking immigration first, imagine an island that as yet contains no species at all. The rate of immigration of species will be high, because any colonising individual represents a species new to that island. However, as the number of resident species rises, the rate of immigration of new, unrepresented species diminishes. The immigration rate reaches zero when all species from the source pool (i.e. from the mainland or from other nearby islands) are present on the island in question (Figure 19.15a).

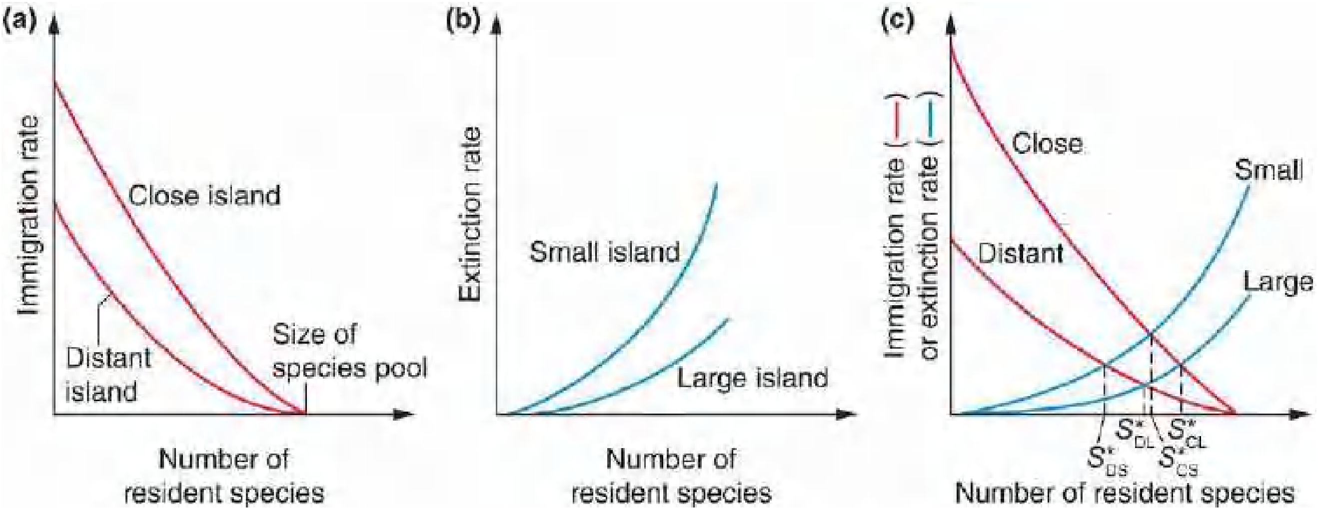


Figure 19.15 MacArthur and Wilson’s (1967) Equilibrium Theory of Island Biogeography. (a) The rate of species immigration on to an island, plotted against the number of resident species on the island, for close and distant islands. (b) The rate of species extinction on an island, plotted against the number of resident species on the island, for large and small islands. (c) The balance between immigration and extinction on small and large and on close and distant islands. In each case, S^* is the equilibrium species richness; C, close; D, distant; L, large; S, small.

The immigration graph is drawn as a curve, because the immigration rate is likely to be particularly high when there are low numbers of residents and many of the species with the greatest powers of dispersal are yet to arrive. In an analysis of immigration rates for breeding birds on 13 small islands of the British Isles, rates decreased consistently with increasing numbers of residents, and for seven islands the curves were indeed significantly concave (Manne *et al.*, 1998). In fact, the curves should arguably be blurs rather than lines, since the precise curve will depend on the exact sequence in which species arrive, and this will vary by chance. In this sense, the immigration curve can be thought of as the ‘most probable’ curve.

The exact immigration curve will depend on the degree of remoteness of the island from its pool of potential colonisers (Figure 19.15a). The curve will always reach zero at the same point (when all members of the pool are resident), but it will generally have higher values on islands close to the source of immigration than on more remote islands, since colonisers have a greater chance of reaching an island the closer it is to the source (Figure 19.15a). It is also plausible that immigration rates will generally be higher on a large island than on a small island, since the larger island represents a larger target for the colonisers.

extinction curves

The rate of species extinction on an island (Figure 19.15b) is bound to be zero when there are no species there, and it will generally be low when there are few species. However, as the number of resident species rises, the extinction rate is assumed by the theory to increase, probably at a more than proportionate rate. This is thought to occur because with more species, competitive exclusion becomes more likely, and the population size of each species is on average smaller, making it more vulnerable to chance extinction. The study of breeding birds in small British islands referred to above (Manne *et al.*, 1998) found that extinction rates did indeed increase consistently with increasing numbers of residents, and that for nine islands the curves were significantly concave. Similar reasoning suggests that extinction rates should be higher on small than on large islands as population sizes will typically be smaller on small islands (Figure 19.15b). As with immigration, the extinction curves are best seen as 'most probable' curves.

the balance between immigration and extinction

In order to see the net effect of immigration and extinction, their two curves can be superimposed (Figure 19.15c). The number of species where the curves cross (S^*) is a dynamic equilibrium and should be the characteristic species richness for the island in question. Below S^*, richness increases (immigration rate exceeds extinction rate); above S^*, richness decreases (extinction exceeds immigration). The theory, then, makes a number of predictions. First, the number of species on an island should eventually become roughly constant through time. Second, this should be a result of a continual *turnover* of species, with some becoming extinct and others immigrating. Third, large islands should support more species than small islands. And last, species number should decline with the increasing remoteness of an island.

the predictions of equilibrium theory are not all exclusive to this theory

Note, though, that several of these predictions could also be made without any reference to the equilibrium theory. An approximate constancy of species number would be expected if richness were determined simply by the intrinsic characteristics of an island. Similarly, a higher richness on larger islands would be expected as a consequence of larger islands having more habitat types. One test of the equilibrium theory, therefore, would be to ask whether richness increases with area at a rate greater than could be accounted for by increases in habitat diversity alone (see Section 19.5.2).

The effect of island remoteness can also be considered quite separately from the equilibrium theory, as MacArthur and Wilson (1967) recognised. Merely recognising that many species are limited in their dispersal ability, and have not yet colonised all islands, leads to the prediction that more remote islands are less likely to be saturated with potential colonisers (see Section 19.5.3). However, the final prediction arising from the equilibrium theory – constancy as a result of turnover – is truly characteristic of the equilibrium theory (see Section 19.5.4).

19.5.2 Habitat diversity alone – or a separate effect of area?

partitioning variation between habitat diversity and island area itself

The most fundamental question in island biogeography, then, is whether there is an 'island effect' as such, or whether islands simply support few species because they are small areas containing few habitats or a narrow range of resources. Does richness increase with area at a rate greater than could be accounted for by increases in habitat diversity alone? Some studies have attempted to partition species–area variation on islands into that which can be entirely accounted for in terms of habitat or resource diversity, and that which remains and must be accounted for by island area in its own right. For beetles on the Canary Islands, the relationship between species richness and resource diversity (as measured by plant species richness) is much stronger than that with island area, and this is particularly marked for the herbivorous beetles, presumably because of their particular food plant requirements (Figure 19.16a).

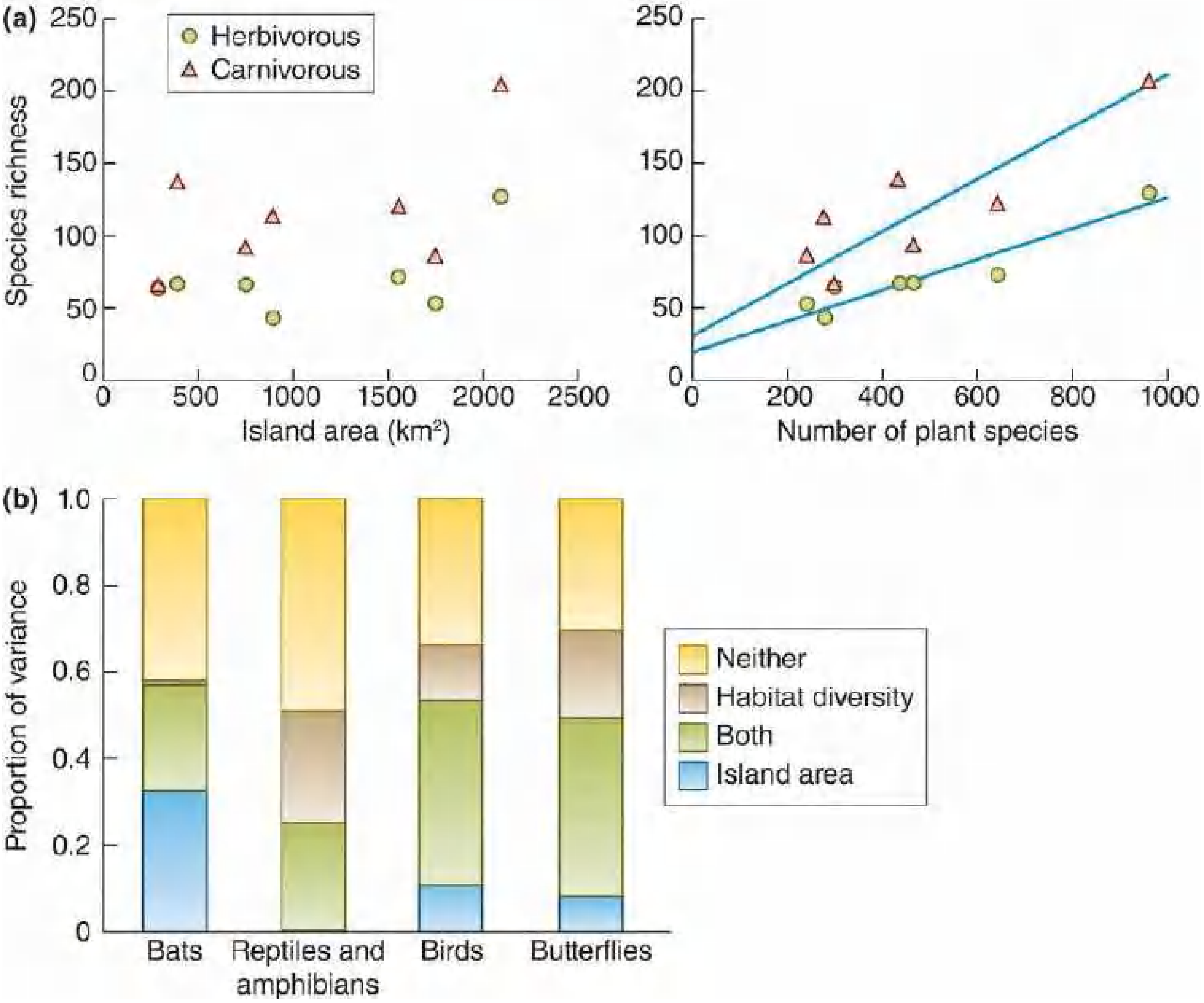


Figure 19.16 Increases of species richness with area are sometimes related to area itself and sometimes to habitat diversity. (a) The species richness of herbivorous and carnivorous beetles of the Canary Islands increases with plant species richness but not with island area. (b) The proportion of variance in species richness among islands in the Lesser Antilles, for four animal groups, showing varying proportions related uniquely to island area, uniquely to habitat diversity, to correlated variation between area and habitat diversity, and unexplained by either.

Source: (a) After Becker (1992). (b) After Ricklefs & Lovette (1999).

On the other hand, in a study of a variety of animal groups living on the Lesser Antilles in the West Indies, the variation in species richness from island to island was partitioned, statistically, into that attributable to island area alone, that attributable to habitat diversity alone, that attributable to correlated variation between area and habitat diversity (and hence not attributable to either alone), and that attributable to neither. For reptiles and amphibians (Figure 19.16b), like the beetles of the Canary Islands, habitat diversity was far more important than island area. But for bats, the reverse was the case, and for birds and butterflies, both area itself and habitat diversity had important parts to play.

experimental reductions in mangrove island area

A classic experiment was carried out to try to separate the effects of habitat diversity and area on some small mangrove islands in the Bay of Florida, USA (Simberloff, 1976). These islands consist of pure stands of the mangrove species *Rhizophora mangle*, which support communities of insects, spiders, scorpions and isopods. After a preliminary faunal survey, some islands were reduced in size – by means of a power saw. Habitat diversity was not affected, but arthropod species richness on three islands nonetheless diminished over a period of two years (Figure 19.17). A control island, the size of which was unchanged, showed a slight increase in richness over the same period, presumably as a result of random events.

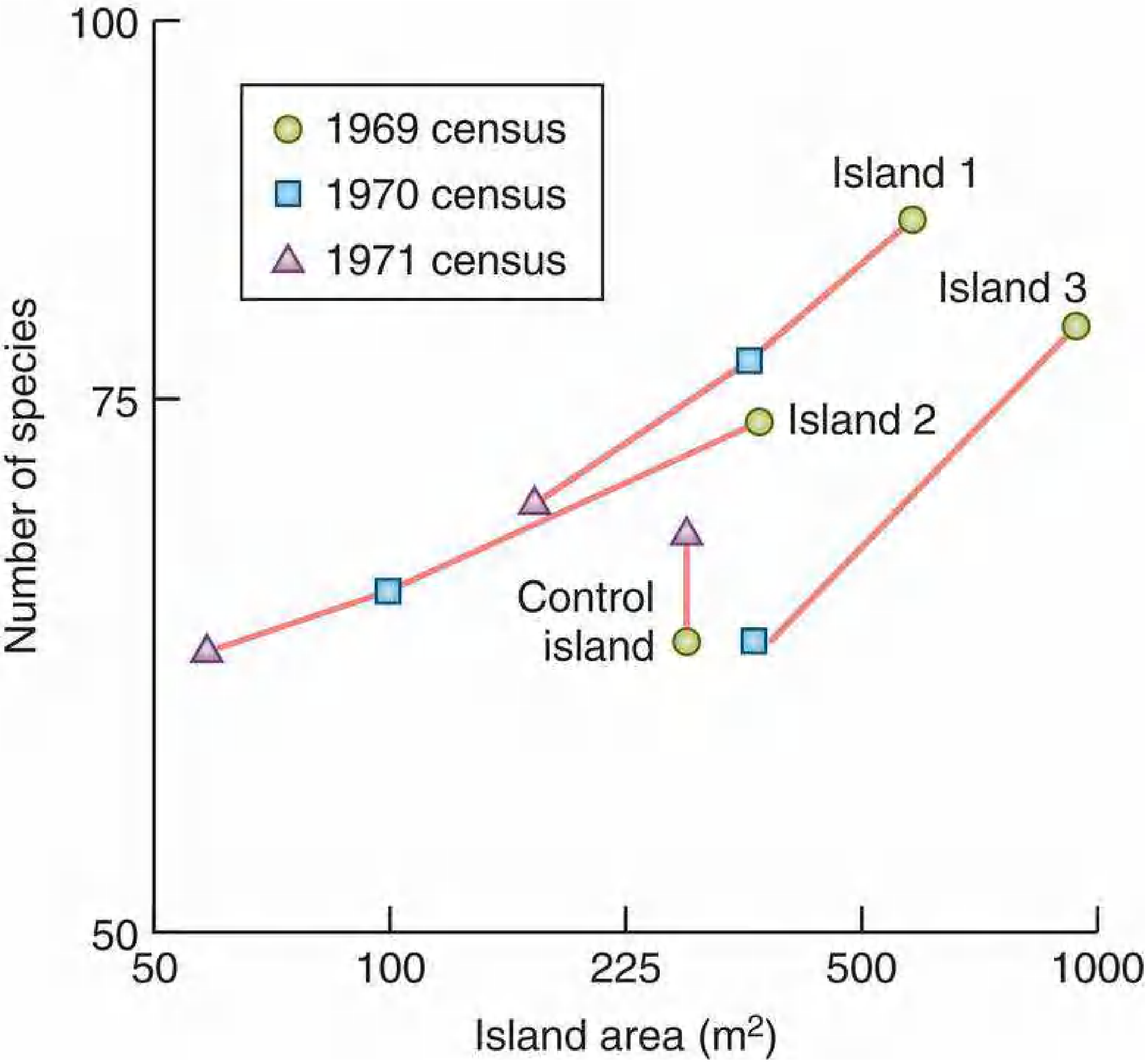


Figure 19.17 An artificial reduction in island area (but not habitat diversity) reduces species richness. The effect on the number of arthropod species of artificially reducing the size of mangrove islands in Florida. Islands 1 and 2 were reduced in size after both the 1969 and 1970 censuses. Island 3 was reduced only after the 1969 census. The control island was not reduced, and the change in its species richness was attributable to random fluctuations.

Source: After Simberloff (1976).

species–area graphs for islands and comparable mainland areas

Another way of trying to distinguish a separate effect of island area is to compare species–area graphs for islands with those for arbitrarily defined areas of mainland. The species–area relationships for mainland areas should be due almost entirely to habitat diversity (together with any ‘sampling’ effect involving increased probabilities of detecting rare species in larger areas). All species will be well able to ‘disperse’ between mainland areas, and the continual flow of individuals across the arbitrary boundaries will therefore mask local extinctions (i.e. what would be an extinction on an island is soon reversed by the exchange of individuals between local areas). An arbitrarily defined area of mainland should thus contain more species than an otherwise equivalent island, and this is usually interpreted as meaning that the slopes of the species–area graphs for islands should be steeper than those for mainland areas (since the effect of island isolation should be most marked on small islands, where extinctions are most likely). The difference between the two types of graph would then be attributable to the island effect in its own right. Table 19.1 shows that despite considerable variation, the island graphs do typically have steeper slopes.

Table 19.1 The slopes of species–area relationships are typically steeper for islands than for arbitrarily defined areas of land. Values of the slope z, of species–area curves ($\log S = \log C + z \log A$, where S is species richness, A is area and C is a constant giving the number of species when A has a value of 1), for arbitrary areas of mainland, oceanic islands and habitat islands. *Source*: After Preston (1962), May (1975b), Gorman (1979), Browne (1981), Matter *et al.* (2002), Barrett *et al.* (2003) and Storch *et al.* (2003).

Taxonomic group	**Location**	**z**
Arbitrary areas of mainland		
Birds	Central Europe	0.09
Flowering plants	England	0.10
Birds	Neoarctic	0.12
Savanna vegetation	Brazil	0.14
Land plants	Britain	0.16
Birds	Neotropics	0.16
Oceanic islands		
Birds	New Zealand islands	0.18
Lizards	Californian islands	0.20
Birds	West Indies	0.24
Birds	East Indies	0.28
Birds	East Central Pacific	0.30
Ants	Melanesia	0.30
Land plants	Galápagos	0.31
Beetles	West Indies	0.34
Mammals	Scandinavian islands	0.35
Habitat islands		
Zooplankton (lakes)	New York State	0.17
Snails (lakes)	New York State	0.23
Fish (lakes)	New York State	0.24
Birds (páramo vegetation)	Andes	0.29
Mammals (mountains)	Great Basin, USA	0.43
Terrestrial invertebrates (caves)	West Virginia	0.72

Note, too, that a reduced number of species per unit area on islands should also lead to a lower value for the intercept on the *S*-axis of the species–area graph. Figure 19.18a illustrates both an increased slope and a reduced value for the intercept for the species–area graph for ant species on isolated Pacific islands, compared with the graph for progressively smaller areas of the very large island of New Guinea. Figure 19.18b illustrates a similar relationship for reptiles on islands off the coast of South Australia.

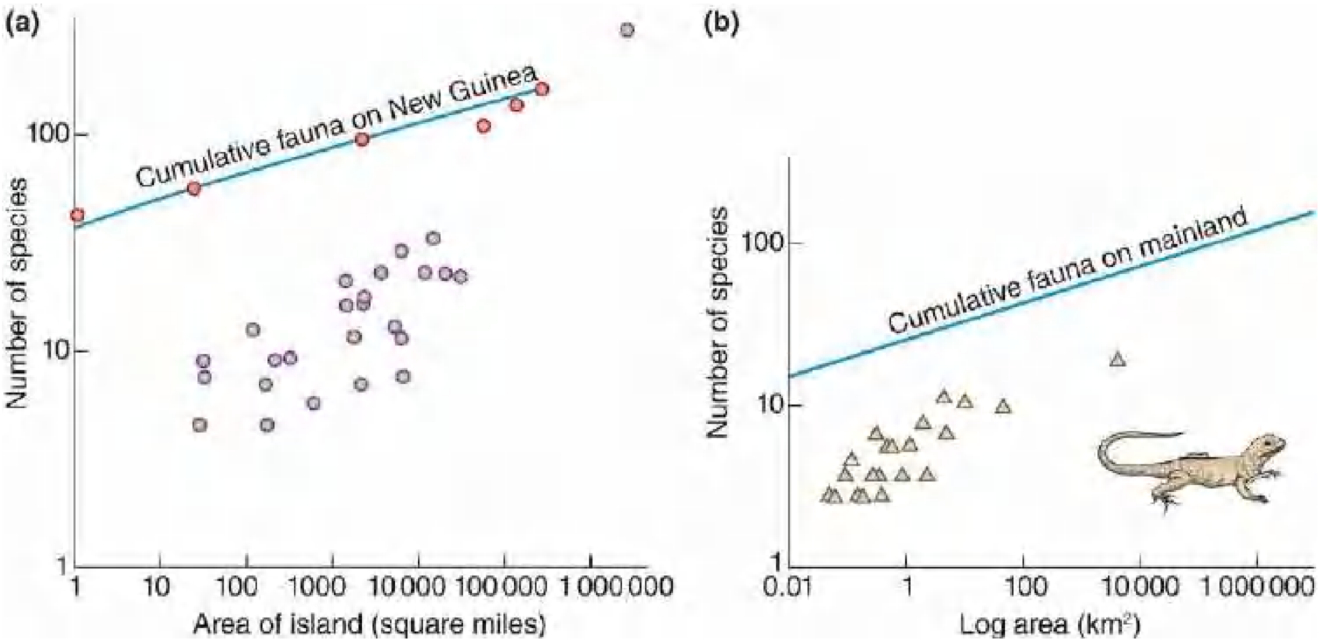


Figure 19.18 The intercepts of species–area relationships are typically lower for islands than for arbitrarily defined areas of land. (a) The species–area graph for ponerine ants on various Moluccan and Melanesian islands compared with a graph for different-sized sample areas on the very large island of New Guinea. (b) The species–area graph for reptiles on islands off the coast of South Australia compared with the mainland species–area relationship. In this case, the islands were formed within the last 10 000 years as a result of rising sea level.

Source: (a) After Wilson (1961). (b) After Richman *et al*. (1988).

plant extinction and immigration rates in relation to island size

Overall, therefore, studies like these suggest a separate area effect (larger islands are larger targets for colonisation; populations on larger islands have a lower risk of extinction) beyond a simple correlation between area and habitat diversity. Lofgren and Jerling (2002) were able to quantify plant extinction rates and immigration rates on islands of different sizes in the Stockholm Archipelago (see Figure 19.14a) by comparing species lists in their survey (1996–99) with those reported from the period 1884–1908. In the intervening time, 93 new species appeared while 20 species disappeared from the islands. Many of the newcomers were trees, bushes and shade-tolerant shrubs, reflecting succession after the cessation of cattle grazing and hay-making in the 1960s, but despite these confounding effects of succession, as predicted, the extinction rate was negatively correlated, and the immigration rate positively correlated, with island size.

19.5.3 Remoteness

It follows from the above argument that the island effect and the species impoverishment of an island should be greater for more remote islands. (Indeed, the comparison of islands with mainland areas is only an extreme example of a comparison of islands varying in remoteness, since local mainland areas can be thought of as having minimal remoteness.) Remoteness, however, can mean two things. It can simply refer to the degree of physical isolation. But a single island can also itself vary in remoteness, depending on the type of organism being considered: the same island may be remote from the point of view of land mammals but not from the point of view of birds.

species richness on islands often decreases with remoteness

The effects of remoteness can be demonstrated either by plotting species richness against remoteness itself, or by comparing the species–area graphs of groups of islands (or for groups of organisms) that differ in their remoteness (or powers of colonisation). In either case, there can be considerable difficulty in extricating the effects of remoteness from all the other characteristics by which two islands may differ. Nevertheless, the direct effect of remoteness can be seen, for example, in Figure 19.19a for ectomycorrhizal fungi on the roots of single trees in tree clumps (islands) isolated from contiguous forest at Point Reyes National Seashore, California, and in Figure 19.19b for fish in spring-fed ponds connected to the Shigenobu River in south-western Japan. In addition, Figure 19.19c contrasts the species–area graphs of two classes of organisms, birds and ferns, in two regions: the relatively remote Azores (in the Atlantic, far to the west of Portugal) and the Channel Islands (close to the north coast of France). The Azores are indeed far more remote than the Channel Islands from the point of view of birds. But the two island groups are apparently equally remote for ferns, which are particularly good dispersers because of their light, wind-blown spores.

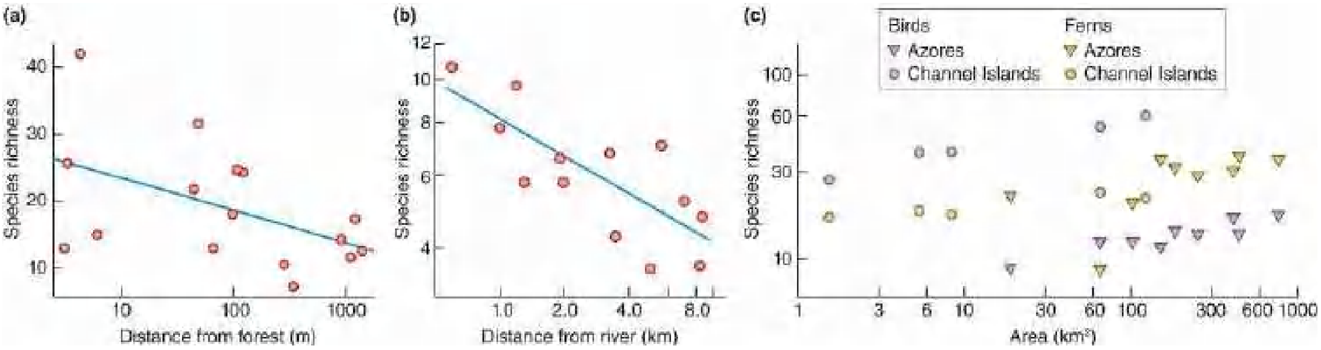


Figure 19.19 Species richness tends to decrease on islands the more isolated they are from a 'mainland' species pool. (a) Ectomycorrhizal fungi on the roots of single trees in tree clumps isolated from contiguous forest at Point Reyes National Seashore, California, USA (*P* = 0.048). (b) Fish in spring-fed ponds connected to the Shigenobu River in south-western Japan (*P* = 0.012). (c) Species–area plots in the Azores and the Channel Islands for land and freshwater breeding birds and for native ferns. The Azores are more remote for birds but not for ferns.

Source: (a) After Peay *et al.* (2010). (b) After Uchida & Inoue (2010). (c) After Williamson (1981).

the colonisation of Surtsey

A more transient but nonetheless important reason for the species impoverishment of islands, especially remote islands, is the fact that many lack species that they could potentially support, simply because there has been insufficient time for the species to colonise. An example is the island of Surtsey, which emerged in 1963 as a result of a volcanic eruption. The new island, 40 km south-west of Iceland, was reached by bacteria and fungi, some seabirds, a fly and the seeds of several beach plants within 6 months of the start of the eruption. Its first established vascular plant was recorded in 1965, and numbers have continued to grow ever since, though they may now be reaching an equilibrium (Figure 19.20). The figure also shows us that this process of colonisation has been accompanied, more or less throughout, by a succession of temporary extinctions, such that the number of species present is almost always less than the cumulative number that have colonised. The general importance of this example is that the communities of many islands can be understood *neither* in terms of simple habitat suitability *nor* as a characteristic equilibrium richness. Rather, they stress that many island communities have not reached equilibrium and are certainly not fully 'saturated' with species.

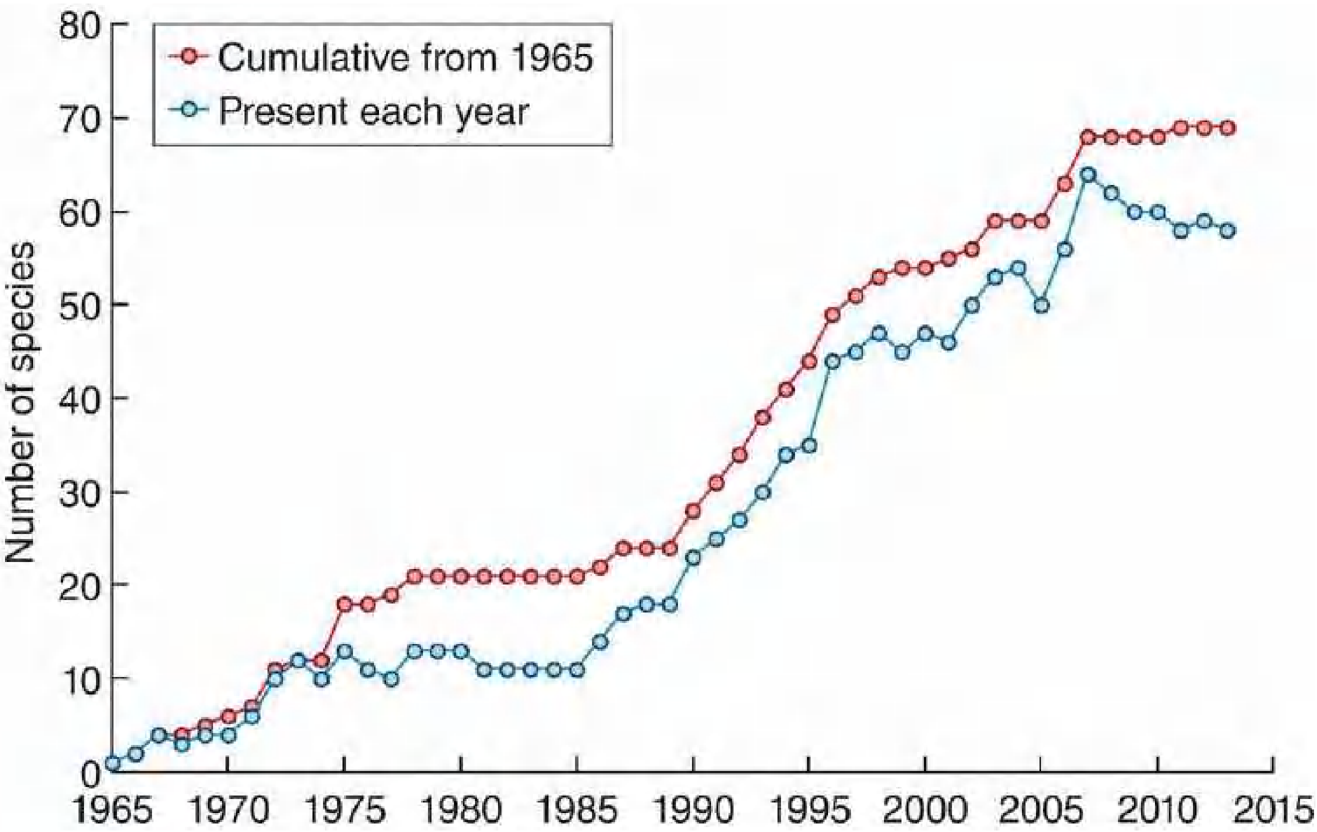


Figure 19.20 Islands may lack species because they have not had time to colonise. The number of species of vascular plants recorded on the new volcanic island of Surtsey, off Iceland, each year since the first arrival in 1965 up to 2013, and the cumulative number ever observed there.

Source: After Magnússon *et al*. (2014).

19.5.4 Which species? Turnover

The equilibrium theory predicts not only a characteristic species richness for an island, but also a *turnover* of species in which new species continually colonise whilst others become extinct. This implies a significant degree of chance regarding precisely which species are present at any one time. However, good studies of turnover have been rare, because communities have to be followed over a period of time, and, ideally, every species should be logged on every occasion so as to avoid ‘pseudo-immigrations’ and ‘pseudo-extinctions’. Unsurprisingly, perhaps, many of the studies have been of birds – mobile enough for colonisation to occur at detectable rates and large enough to be seen. One revealing study involved censuses from 1949 to 1975 of the breeding birds in a small oak wood (Eastern Wood) in southern England. In all, 44 species bred in the wood over this period, with the number breeding in any one year varying between 27 and 36. The immigration and extinction ‘curves’ are shown in Figure 19.21. Clearly, the points are scattered in contrast to the assumed simplicity of the MacArthur–Wilson model. Indeed, there is no significant relationship in the extinction graph. But the negative correlation in the immigration graph is significant, and the two lines do seem to cross at roughly 32 species (the richness typically observed), with three new immigrants and three extinctions each year (turnover).

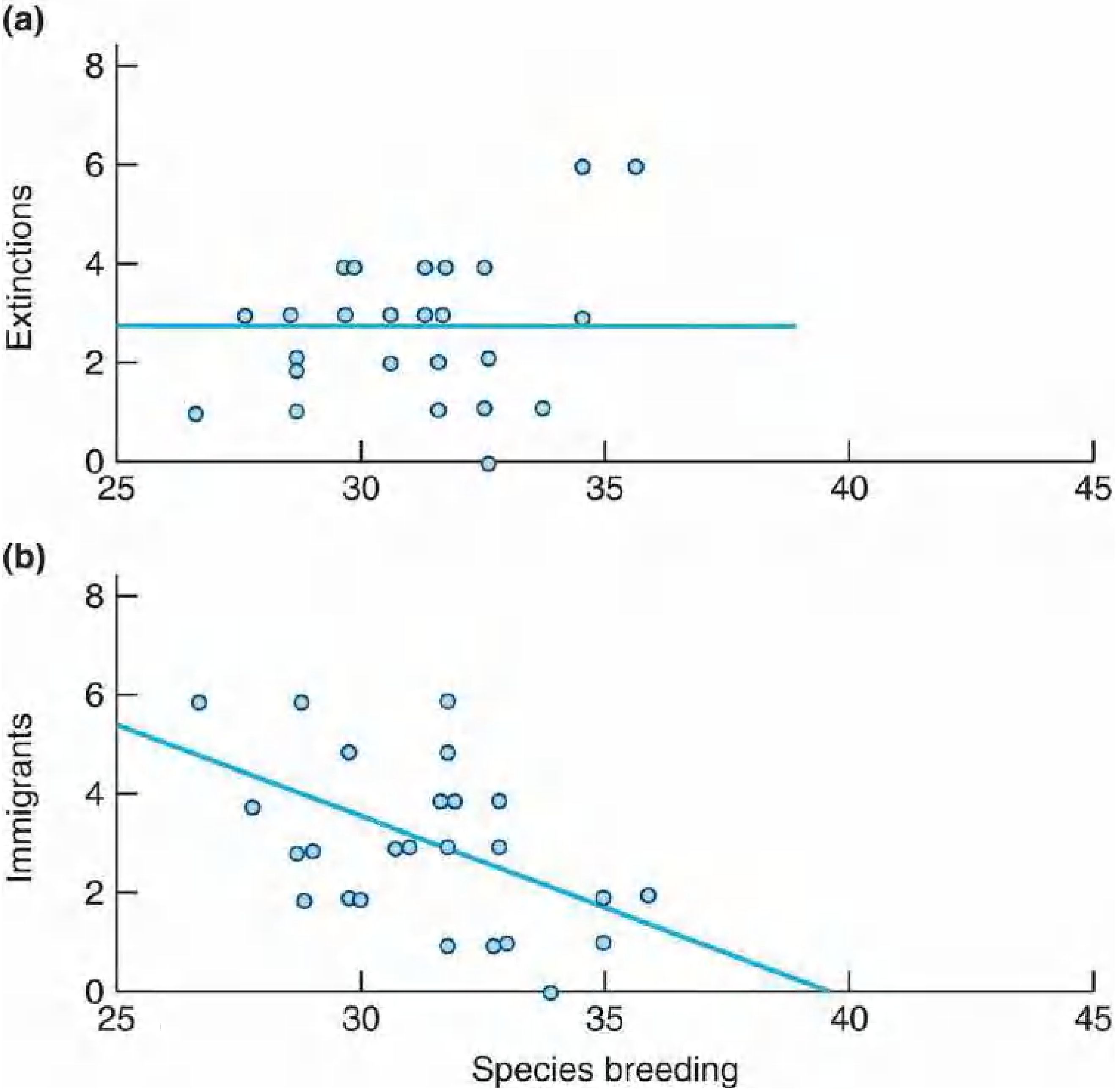


Figure 19.21 A constancy of species richness may hide a turnover of individual species. (a) Extinction and (b) immigration of breeding birds at Eastern Wood, UK. The line in (a) shows the average number of extinctions since there was no significant association with species richness. The line in (b) is the calculated regression line with a slope of −0.38; $P < 0.05$.

Source: After Beven (1976) and from Williamson (1981).

Experimental evidence of turnover and indeterminacy was provided by the work of Simberloff and Wilson (1969), who exterminated the invertebrate fauna on a series of small mangrove islands in the Florida Keys, USA and monitored recolonisation (see Figure 19.17). Within about 200 days, species richness had stabilised around the level prior to defaunation, but with many differences in species composition. Subsequently, the rate of turnover of species on the islands was estimated as 1.5 extinctions and colonisations per year (Simberloff, 1976).

Thus, the idea that there is a turnover of species leading to a characteristic equilibrium richness on islands, but an indeterminacy regarding particular species, certainly has support. On the other hand, several authors have noted that experimental studies, especially, tend to last only a few years, and that it tends to be the rarest species that contribute to species turnover. They have suggested therefore that the equilibrium theory is 'true but trivial' (Williamson, 1989) since the species to which it applies, because of their rarity, contribute least to the functioning of island ecosystems. A much longer term study, lasting 69 years, of the breeding birds on the island of Skokholm, 3.5 km from the coast of Wales, was able to address this question by noting the number of colonisations and extinctions across intervals of 1, 3, 6, 12 and 24 years (Figure 19.22). Certainly for the shortest intervals, and especially from year to year, turnover was dominated by the rarest species. However, it would not be true to say that there was no turnover of more abundant species. The most abundant species played little or no part in the year to year changes, but were important contributors to turnover over the longer intervals – which only long-term studies like this are able to detect.

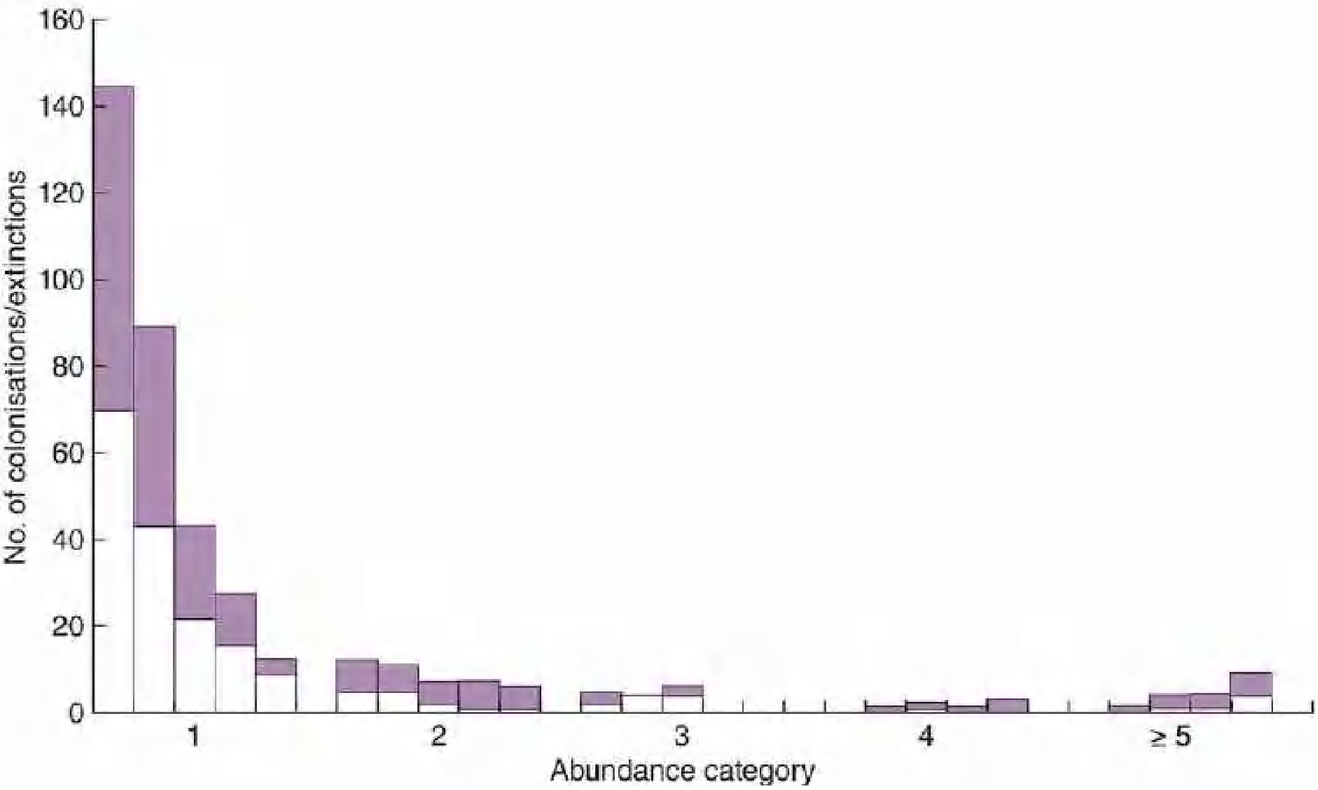


Figure 19.22 Island turnover is dominated by rare species at short timescales, but more abundant species turnover at longer timescales. Turnover of bird species on the island of Skokholm, UK, over 69 years (1928–47 and 1980–2002). There were five abundance categories, as indicated: maxima of 1, 2, 3, 4 and ≥5 breeding pairs. For each, colonisations (shaded) and extinctions (no shading) were noted over five intervals (1, 3, 6, 12 and 24 years) represented by the five bars in each abundance category.

Source: After McCollin (2017).

19.5.5 Which species? Disharmony

It has long been recognised – for example by Joseph Hooker in the mid-19th century – that one of the main characteristics of island biotas is 'disharmony', that is, the relative proportions of different taxa are not the same on islands as they are on the mainland. We have already seen that groups of organisms with good powers of dispersal like ferns are more likely to colonise remote islands than are groups with relatively poor powers of dispersal (Figure 19.19c). On the other hand, it would be wrong to assume that powers of dispersal override all other factors in determining which species are present or absent from particular islands. Carvajal-Endara *et al.* (2017), for example, analysed the distribution of plants on the Galápagos Islands by assuming they would arrive from a source pool on the South American mainland. They then related the presence or absence of plants on the Galápagos to their long-distance dispersal strategy, the 'niche dissimilarity' between the centroid of their climatic niche hypervolume on the mainland and that of the available niche space on the Galápagos, and the 'niche overlap' between points drawn at random from their climatic niche and that available on the Galápagos. Of these, niche dissimilarity was by far the best predictor. What the authors call 'habitat filtering' (determined by environmental constraints – see Section 18.1), rather than dispersal limitation, seems to be the main force shaping the Galápagos Islands' flora.

Species may also vary in their risk of extinction. For example, species that naturally have low densities per unit area, like many vertebrate predators, are bound to have only small populations on islands, where a chance fluctuation may eliminate them altogether, and are notable for their absence on many islands. Specialist predators, parasites and mutualists are also liable to be absent from islands, because their immigration can only lead to colonisation if their prey or partners have arrived first. Thus, disharmony may arise because some types of organism are more dependent than others on other species with which they interact being present.

incidence functions

In an influential attempt to understand island communities by combining ideas on dispersal and extinction differentials with those on sequences of arrival and habitat suitability, Diamond (1975) developed 'assembly rules' (what was required for a successful colonisation), and in particular *incidence functions* for the birds of the Bismark Archipelago off the coast of New Guinea. Incidence functions plot the proportion of islands occupied by a species against the species richness of the islands or, more commonly in practice, their size. (Richness and size are of course correlated.) This allowed Diamond to contrast, for example, 'supertramp' species (present only on the smallest islands because they have high rates of dispersal but a poorly developed ability to persist in communities with many other species) with 'high S' species (only able to persist on large islands with many other species and a wide array of habitat types).

Such contrasting incidence functions prompted Matthews *et al.* (2014) to plot, in turn, the proportions of specialist and generalist species against island size for four bird datasets (Figure 19.23). (Generalists tend to be found even on the smallest, most species-poor islands; specialists only on larger, species-rich islands, where abundances are greater and dependencies more likely to be met.) Absolute values differed between the datasets, but the pattern was the same in each case. The specialist : generalist ratio increased rapidly around a threshold island size but then tended to level off: the very smallest islands supported a very small proportion of specialist species. Again, we see disharmony, and that it takes far more than a count of the number of species present to characterise the community of an island.

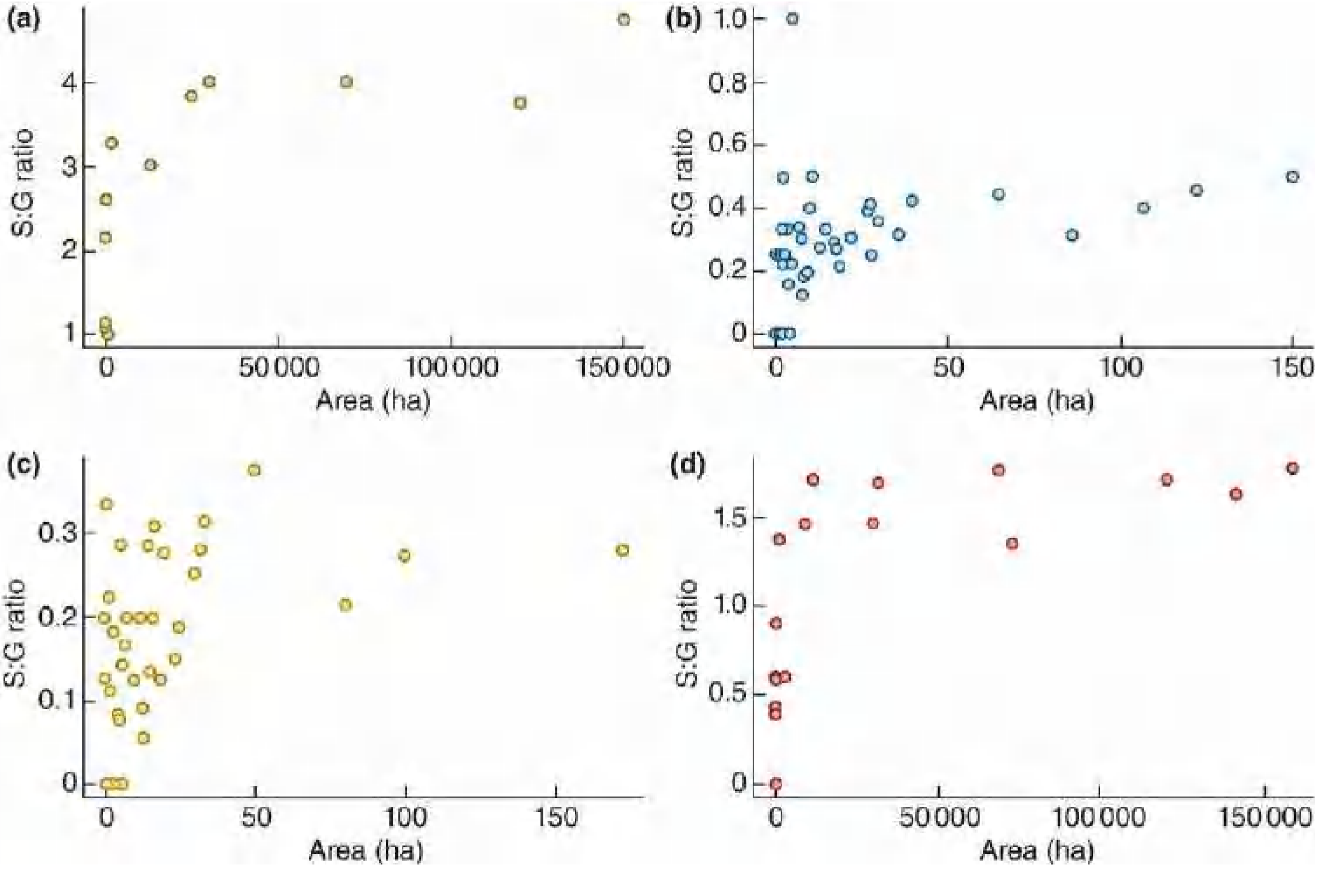


Figure 19.23 **The proportion of specialist species is low on small islands, rising to a plateau as island size increases.** The changing ratio of specialist to generalist species (S : G ratio) as island area increases for birds in four groups of islands, as follows: (a) agricultural fragments in Brazilian tropical forests, (b) agricultural fragments in Norway, (c) agricultural fragments in the UK, and (d) highland scrub habitats in Mexico.

Source: After Matthews *et al.* (2014).

19.5.6 Which species? Evolution

evolution rates on islands may be faster than colonisation rates

No aspect of ecology can be fully understood without reference to evolutionary processes (see Chapter 1), and this is particularly true for an understanding of island communities. On isolated islands, or for groups with poor dispersal, the rate at which new species evolve may be comparable with or even faster than the rate at which they arrive as new colonists. A classic illustration of the balance between colonisation and the evolution of endemic species is provided by the animals and plants of Norfolk Island (Figure 19.24). This small island (about 70 km^2) is approximately 700 km from New Caledonia and New Zealand, but about 1200 km from Australia. The ratio of Australian species to New Zealand and New Caledonian species within a group can therefore be used as a measure of that group's dispersal ability. As Figure 19.24 shows, the proportion of endemics on Norfolk Island (found only there) is highest in groups with poor dispersal ability and lowest in groups with good dispersal ability.

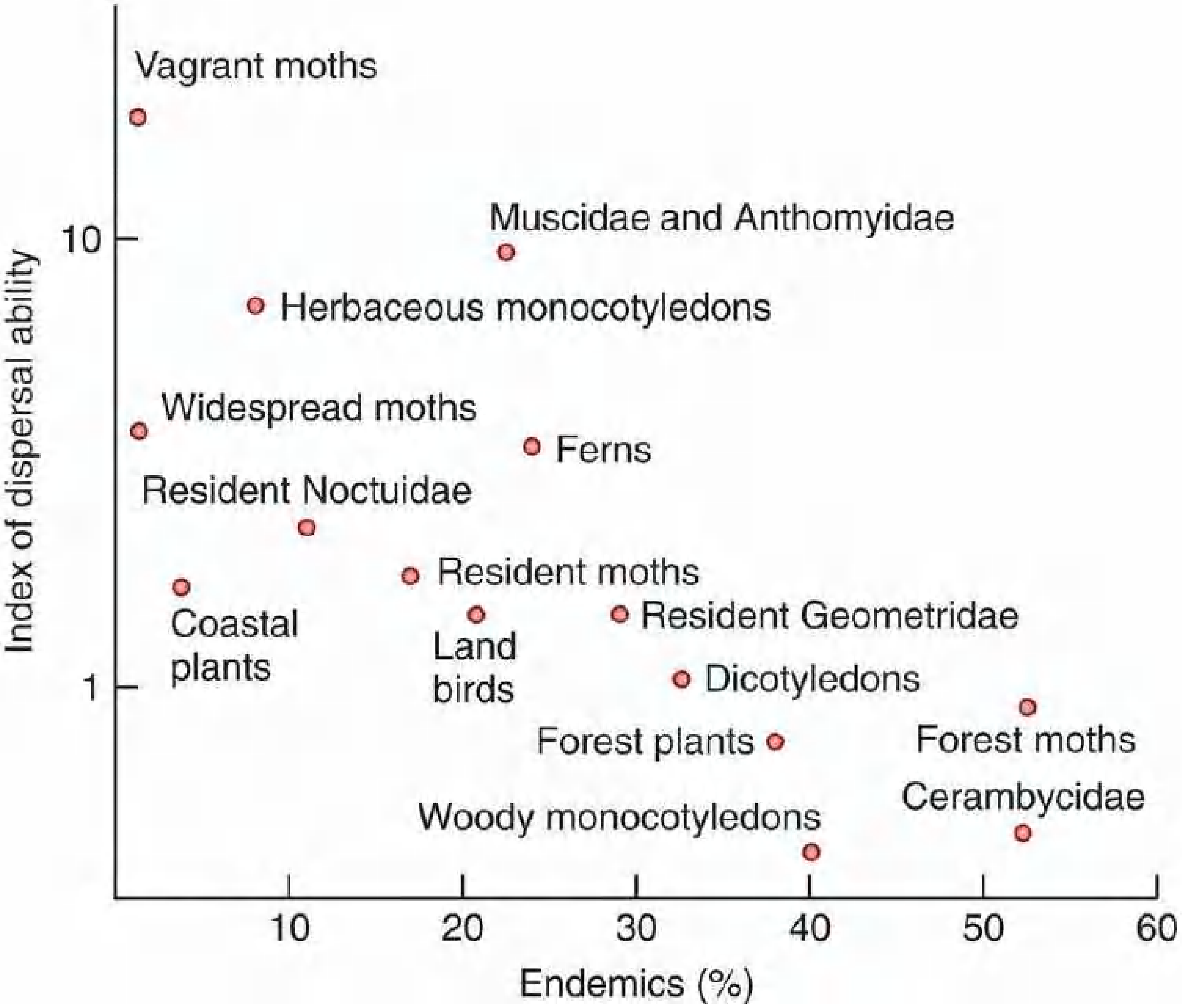
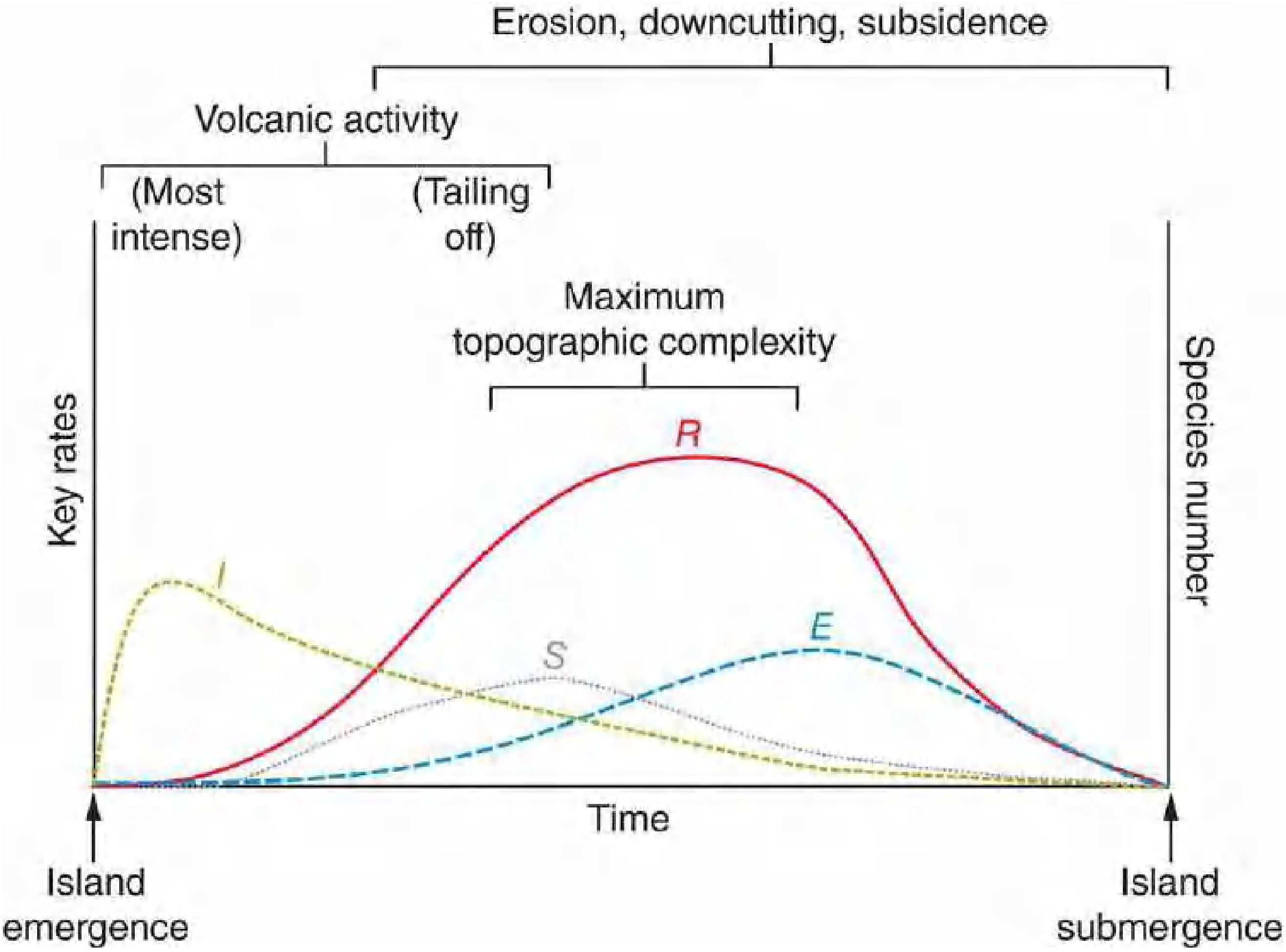


Figure 19.24 The proportion of endemics on an island increases with island isolation for the species concerned. Poorly dispersing groups on Norfolk Island have a higher proportion of endemic species, and are more likely to contain species that have reached Norfolk Island from either New Caledonia or New Zealand than species from Australia, which is further away. The converse holds for good dispersers.

Source: After Holloway (1977).

incorporating geology: a general dynamic model

These ecological and evolutionary processes are likely to interact, in turn, with geological processes, particularly volcanic eruptions, that often drive the ontogeny of remote oceanic islands especially, with initial emergence and expansion of an island giving way to subsequent erosion, subsidence and contraction. Indeed, a *general dynamic model* of oceanic island biogeography has been proposed to draw all these elements together (Whittaker *et al.*, 2008 ; Borregaard *et al.*, 2017). The key features are illustrated in Figure 19.25. Crucially, the model relates the rate of speciation on an island, and its contribution to the island's biota, to island age, proposing a maximum rate at an early, but not the earliest, stage of island ontogeny, when there is the greatest discrepancy between the range of opportunities for species to occupy available niches and the number of immigrant species that are doing so. Before this, species accumulation into the limited niche space is dominated by colonisation. Later, especially as an island shrinks, the available niche space may be close to saturation, competitive exclusion of existing species may increase, and opportunities for the evolution of new species will be much scarcer.

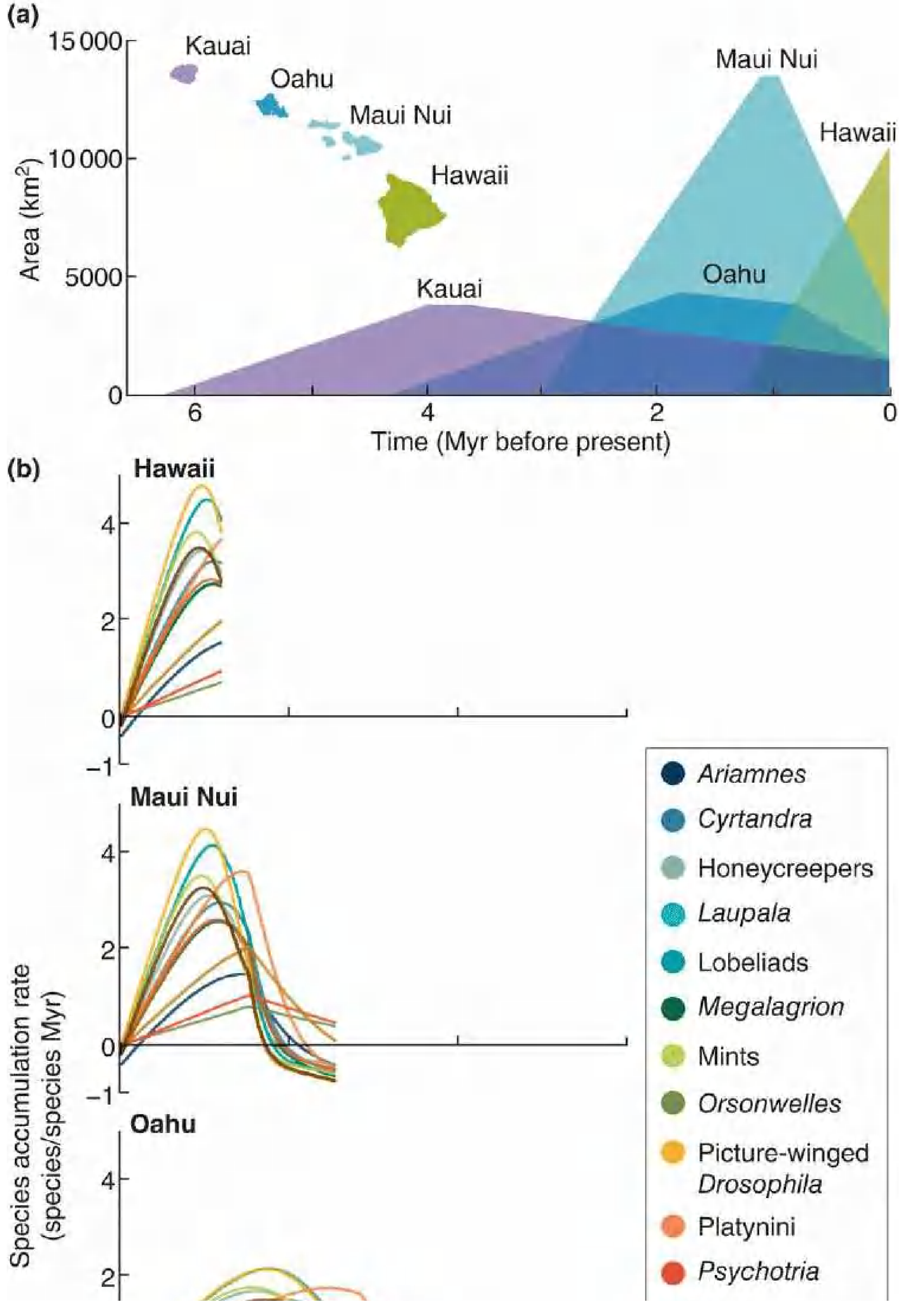


[Figure 19.25](#) Graphical summary of the general dynamic model of ocean island biogeography. The curves are as follows: *I*, immigration rate; *S*, speciation rate; *E*, extinction rate; *R*, species richness. The initial upturn in the immigration rate reflects a pattern in which newly emerged islands are often inhospitable initially, until the earliest colonisers make the conditions more 'attractive'.

Source: After Borregaard *et al.* (2017).

the changing role of evolution on the Hawaiian Islands

Support for the model is provided by a study of 14 endemic groups from the Hawaiian Islands; that is, groups where it is established that all members of the group derive from a single colonisation of the archipelago, and where that colonisation occurred when Kaui (the oldest of the current islands) was the only major island or even earlier than that time (Lim & Marshall, 2017). These groups included plants (for example, silverswords and lobeliads), birds (honeycreepers) and insects (for example, the famous picture-winged *Drosophila*). The analysis was focused on four main islands: Kauai, Oahu, Maui Nui (actually a collection of four islands connected for most of their history) and Hawaii, the youngest of the islands (Figure 19.26a). For each group, models were developed relating species accumulation (immigration and speciation, minus extinction) to the estimated changing size and age of the islands, and a best-fit model was found by comparing model outputs to the known richnesses of the various groups on the different islands. The models were then used to reconstruct the histories of species accumulation for each group on each island since its first appearance (Figure 19.26b).


(a)
15 000
10 000
5000
0
Area (km^2)
Kauai
Oahu
Maui Nui
Hawaii
Maui Nui
Hawaii
Oahu
Kauai
6
4
2
0
Time (Myr before present)
(b)
Hawaii
4
2
0
−1
Maui Nui
4
2
0
−1
Oahu
4
2
Species accumulation rate (species/species Myr)
● Ariamnes
● Cyrtandra
● Honeycreepers
● Laupala
● Lobeliads
● Megalagrion
● Mints
● Orsonwelles
● Picture-winged Drosophila
● Platynini
● Psychotria

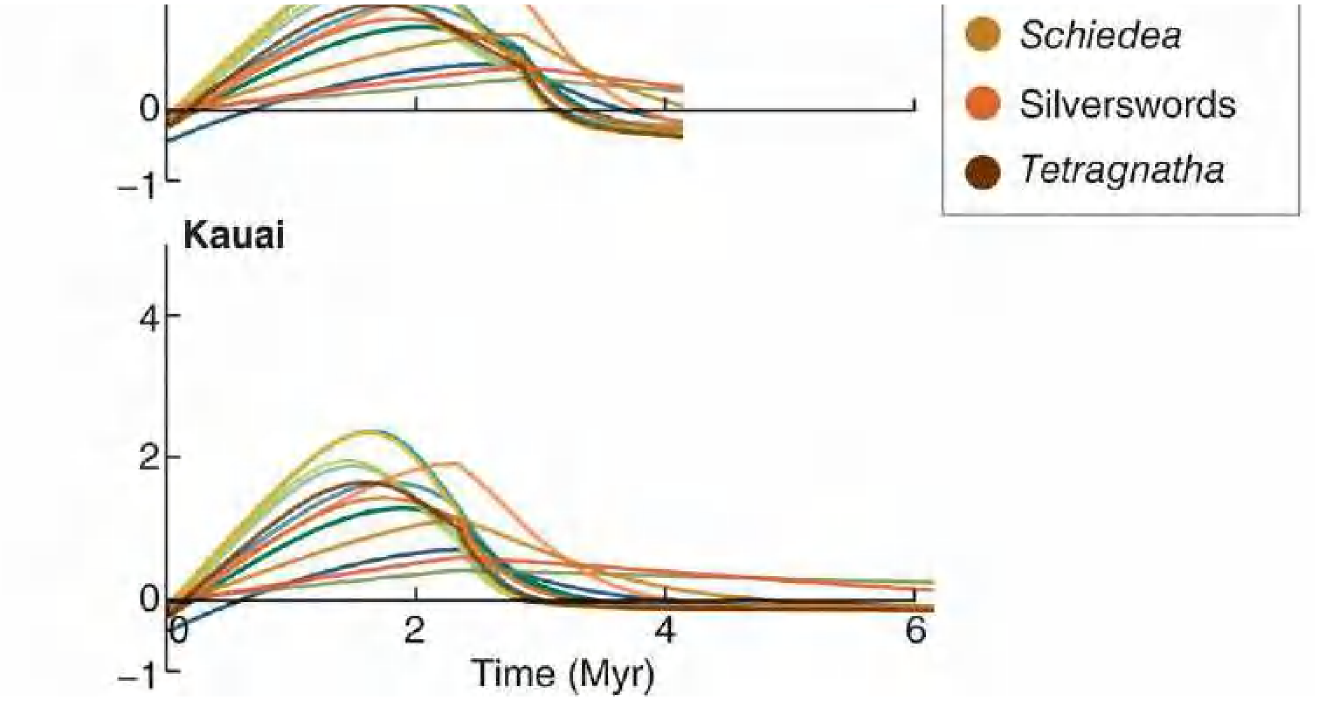


Figure 19.26 Speciation rates first rise then decline as islands emerge and then age in the Hawaiian Archipelago. (a) Map of the main islands and history of their changing size since they first emerged. (b) Estimated species accumulation rates over time for 14 endemic groups, as indicated, on the four islands.

Source: After Lim & Marshall (2017).

There is a clear pattern in every group on every island. There was an initial period in the history of the island (1–2 million years) when species accumulated rapidly – and since these were all endemic groups, the bulk of this must have been speciation. Indeed, the species accumulation rates during this phase increased with time. This, though, for most groups on most islands, was followed by a period (again of 1–2 million years) of decline in the species accumulation rate – the exception being Hawaii itself, because of its relative youth. In fact for many of the groups on the older islands, species accumulation rates have become negative. Interestingly, the groups in evolutionary decline tend to be those that initially had the highest rates of species accumulation (speciation) relative to the pace of geological change. Conversely, those still expanding tend to have lower speciation rates initially. This study, therefore, confirms the important role of evolution in the construction of island communities and the determination of species richness, but it also alerts us to patterns in the significance of that role, and that even here there may be important differences between different groups and hence further disharmonies behind the fine details of island biotas.

APPLICATION 19.3 Nature reserves as ecological ‘islands’

We have seen that species richness (biodiversity) is favoured by a whole range of factors. Nonetheless, it is possible to propose a very basic recipe for successful conservation (Figure 19.27), which has three core ingredients: habitat quality, habitat area and connectivity, the last of these being a combination of the spatial arrangement of suitable habitats and the provision of links between them. There are clear analogies here to aspects of island biogeography theory. Examples showing the importance of habitat quality, area and connectivity are provided in Figure 19.28. However, the argument has also been made that while, ideally, all should be considered when locating and designing nature reserves, in practice money and resources for conservation are likely to be limited. It may then be best to focus on habitat quality and area, where benefits are predictable and likely to apply to a wide range of species, than on aspects of connectivity, which may be less predictable and more species-specific (Hodgson *et al.*, 2011).

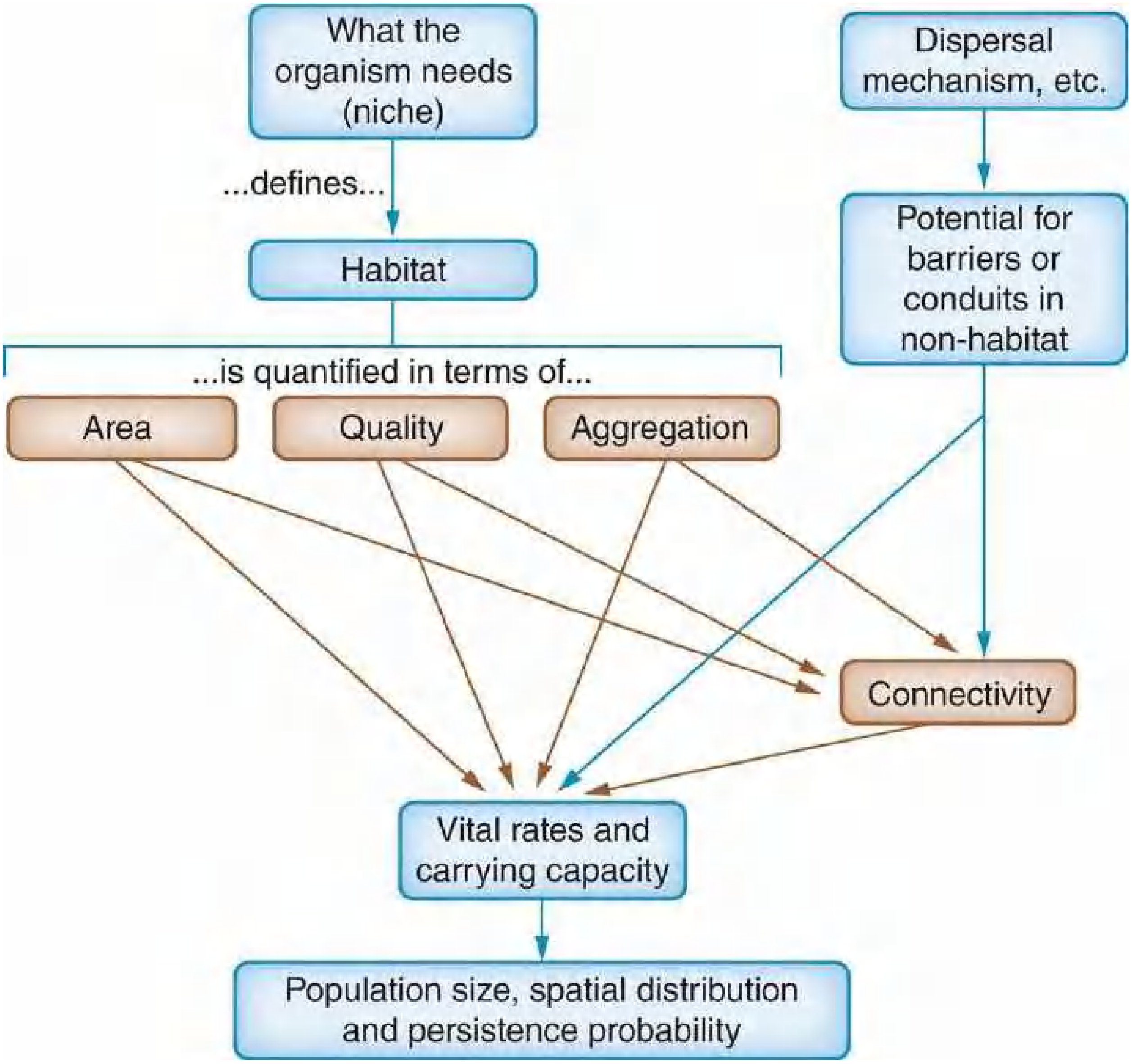


Figure 19.27 A simple recipe for species conservation. Arrows from A to B signify ‘A determines B’.

Source: After Hodgson *et al*. (2009).

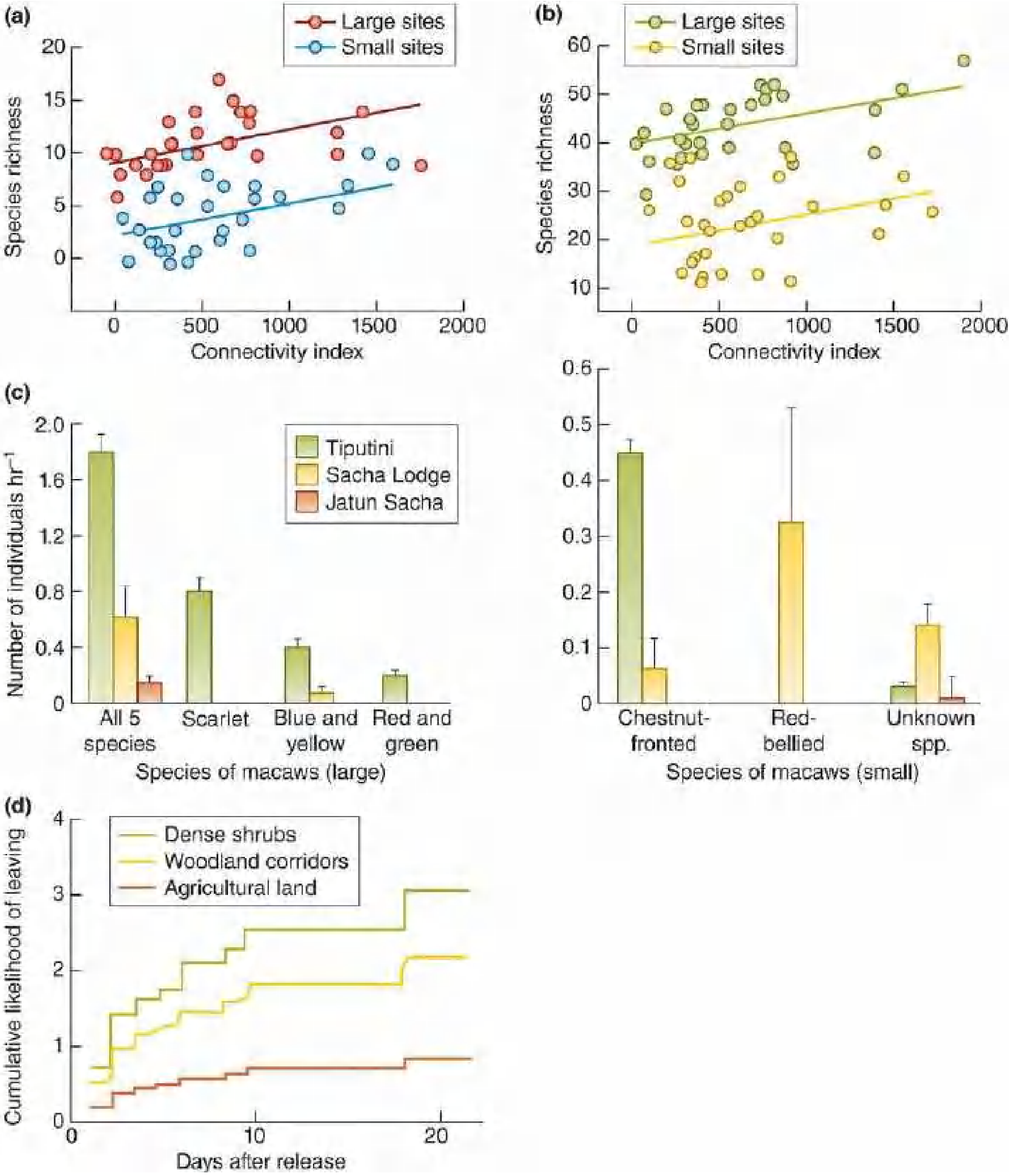

(a)
Large sites
Small sites
Species richness
Connectivity index
0
5
10
15
20
0
500
1000
1500
2000
(b)
Large sites
Small sites
Species richness
Connectivity index
10
20
30
40
50
60
(c)
Tiputini
Sacha Lodge
Jatun Sacha
Number of individuals hr^{-1}
0
0.4
0.8
1.2
1.6
2.0
All 5 species
Scarlet
Blue and yellow
Red and green
Species of macaws (large)
0
0.1
0.2
0.3
0.4
0.5
0.6
Chestnut-fronted
Red-bellied
Unknown spp.
Species of macaws (small)
(d)
Dense shrubs
Woodland corridors
Agricultural land
Cumulative likelihood of leaving
0
1
2
3
4
0
10
20
Days after release

Figure 19.28 Species richness increases with habitat connectivity, habitat area and habitat quality. (a) Habitat connectivity and area. The species richness of butterflies in different grassland patches in southern Germany in relation to a ‘connectivity index’ that combines edge-to-edge distances between a site and all other sites within a 2 km radius with the species’ dispersal ability. Separate relationships are shown for large sites (*c.* 2.4 ha) and small sites (*c.* 0.12 ha). Both relationships are significant ($P < 0.05$). (b) Similar relationships for plants in the same sites. (c) Habitat quality. The abundance of various species of macaw, large to the left, small to the right, in the Ecuadorian Amazon region. Bars are SEs. Of the three sites, Tiputini is largely unaffected by human interference, Sacha Lodge is a tourist site with a human visitation rate twice that at Tiputini, surrounded by military and petroleum camps, and Jatun Sacha is another tourist site, close to a major highway and with an even higher visitation rate. (d) Habitat quality and connectivity. Wild caught birds, Chucao Tapaculos (*Scelorchilus rubecula*), were released from small patches of their favoured woodland habitat in Chile and their dispersal from there monitored in landscapes where the release point was surrounded by agricultural land, woodland corridors or dense shrubs. Dispersal was significantly lower in agricultural land than in the other two ($P < 0.05$), which were not significantly different from one another.

Source: (a, b) After Bruckmann *et al*. (2010). (c) After Karubian *et al*. (2005). (d) After Castellon & Sieving (2006).

SLOSS

One particular question that may arise is whether to construct one large reserve or several small ones adding up to the same total area – sometimes referred to as the SLOSS (single large or several small) debate. If the region is homogeneous in terms of conditions and resources, it is quite likely that smaller areas will contain only a subset, and perhaps a similar subset, of the species present in a larger area. Species–area relationships would then suggest that it is preferable to construct the larger reserve in the expectation of conserving more species in total. On the other hand, if the region as a whole is heterogeneous, then each of the small reserves may support a different group of species and the total conserved might exceed that in one large reserve of the same size. In fact, collections of small islands tend to contain more species than a comparable area composed of one or a few large islands. The pattern is similar for habitat islands (Figure 19.29) and, most significantly, for national parks. Thus, several smaller parks contained more species than larger ones of the same area in studies of mammals and birds in East African parks, of mammals and lizards in Australian reserves and of large mammals in national parks in the USA (Quinn & Harrison, 1988).

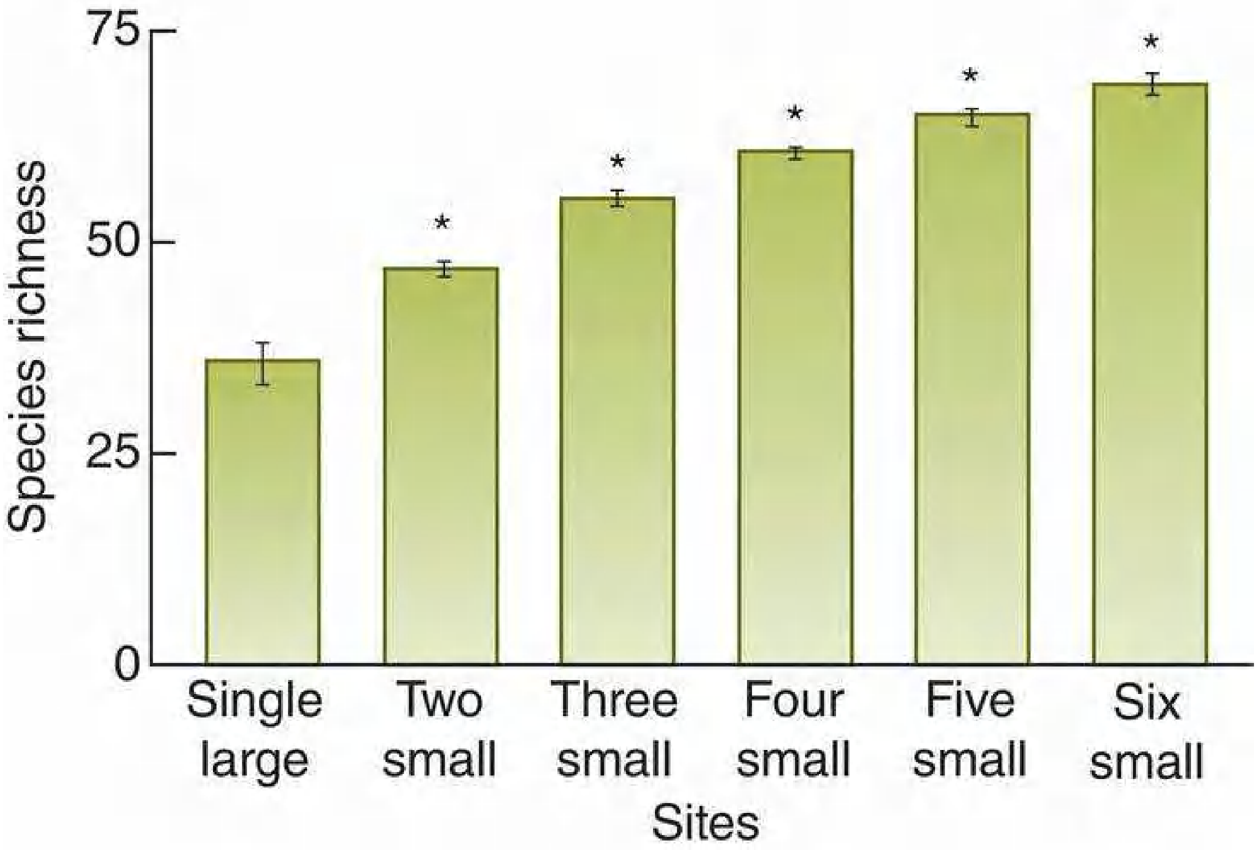


Figure 19.29 A single large site in Finland supports fewer plant species than groups of small sites of the same overall area. The species richness of plants in spruce (*Picea abies*) dominated mire (wetland) sites in Finland. With 24 sites in total, all combinations were examined in which a single large site could be compared with groups of between two and six sites of the same total area. Bars are SEs. Groups of small sites significantly different from the large one ($P < 0.05$) are marked *.

Source: After Virolainen *et al.* (1998).

A point of particular significance is that local extinctions are common events, and so recolonisation of habitat fragments is critical for the survival of fragmented populations. Thus, we need to pay particular attention to the spatial relationships amongst fragments, including the provision of dispersal corridors (see Section 15.4.5). There are potential disadvantages – for example, corridors could increase the correlation among fragments of catastrophic effects such as the spread of fire or disease – but the arguments in favour are persuasive. Indeed, high recolonisation rates (even if this means conservation managers themselves moving organisms around) may be indispensable to the success of conservation of endangered metapopulations. Note especially that human fragmentation of the landscape, producing subpopulations that are more and more isolated, is likely to have had the strongest effect on populations with naturally low rates of dispersal. Thus, the widespread declines of the world's amphibians may be due, at least in part, to their poor potential for dispersal (Blaustein *et al.*, 1994).

19.6 Gradients of species richness

In the preceding sections we have been looking at the possible effects of various factors on species richness and biodiversity generally. Now we turn to patterns in species richness, either in space or in time. We seek to use what we have learned about the effects of the various factors to account for the patterns we observe. Even in the strongest of these patterns, however, there is variation around any general trend. Of course, some of this can be attributed to either sampling or observer error; data collected from nature are never likely to reflect perfectly the underlying reality. But variation also arises simply because the observed species richness will have been co-determined by several of the factors we have discussed, and not all of these will vary in parallel with the gradient we are following. So, for example, when we turn next to latitudinal gradients, the question we will ask is ‘What accounts for the general trend?’ not ‘What accounts for each individual value within the trend?’

19.6.1 Latitudinal gradients

richness decreases with latitude

One of the most widely recognised patterns in species richness is the increase that occurs from the poles to the tropics. This can be seen in a wide variety of groups, including woody angiosperms (trees, shrubs, lianas), marine invertebrates, mammals and lizards (Figure 19.30). The pattern can be seen, moreover, in terrestrial, marine and freshwater habitats.

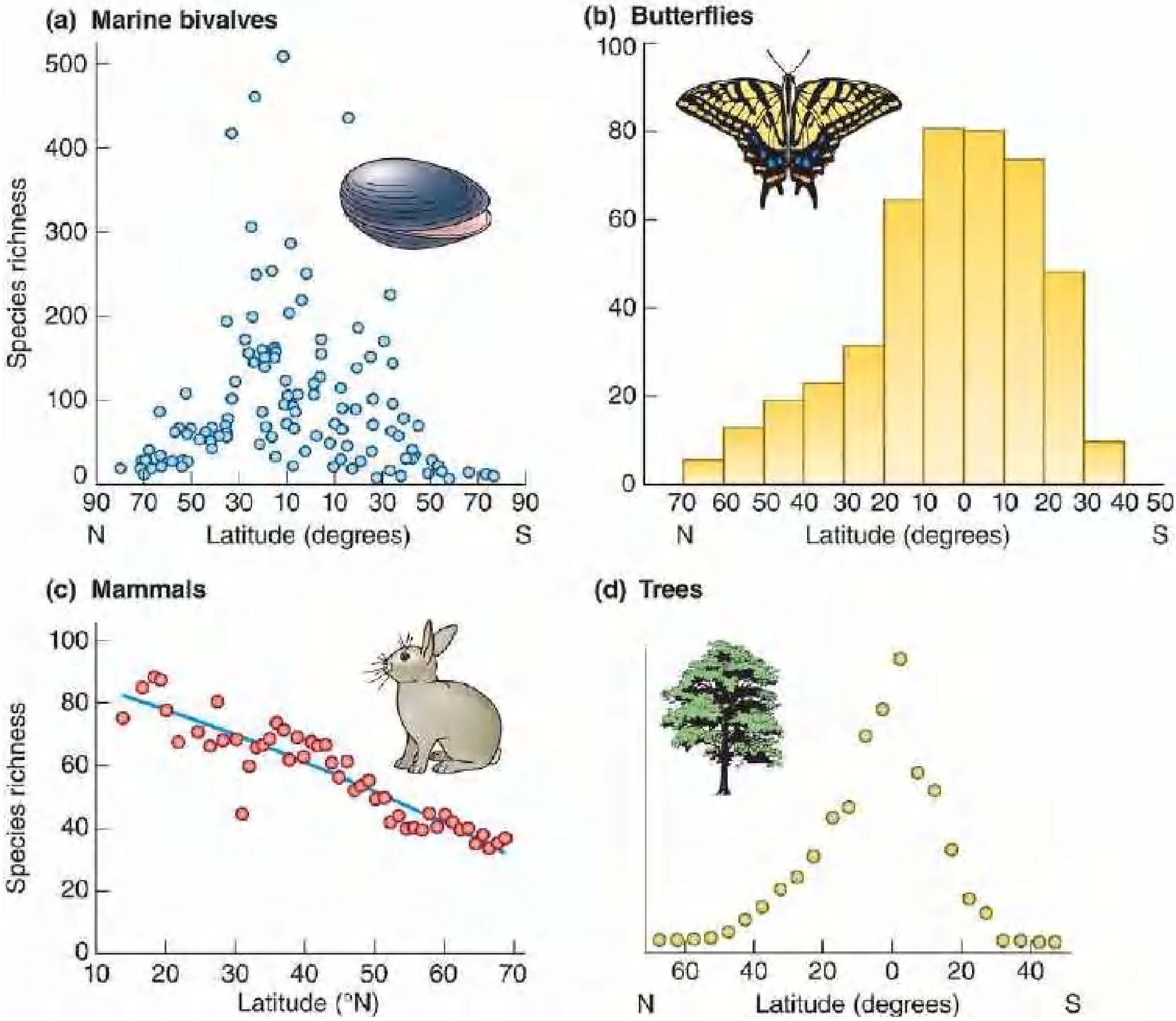


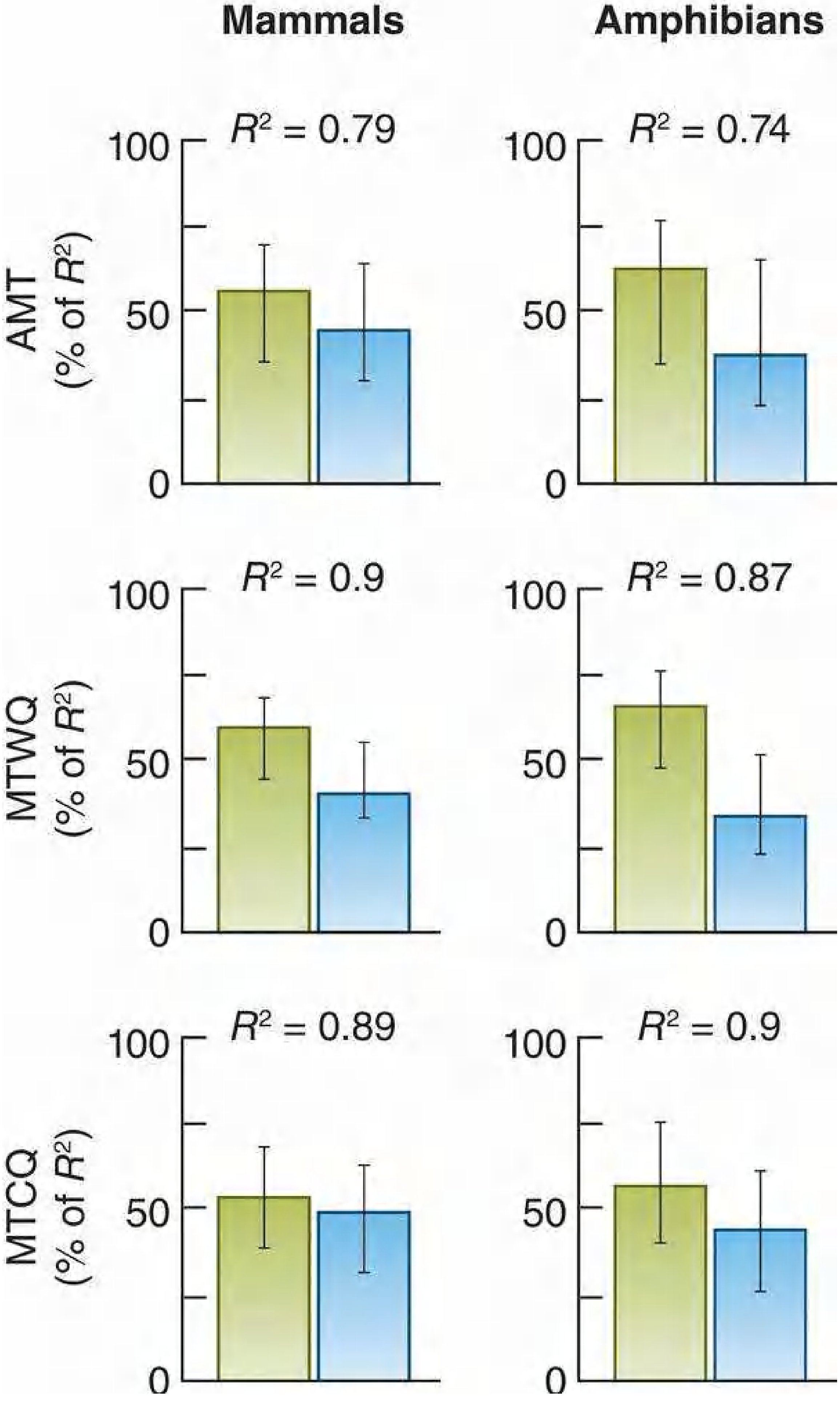
Figure 19.30 Species richness is higher nearer to the equator (and lower nearer to the poles). Latitudinal patterns in species richness in: (a) marine bivalves, (b) swallowtail butterflies, (c) quadruped mammals in North America and (d) woody angiosperms in North and South America.

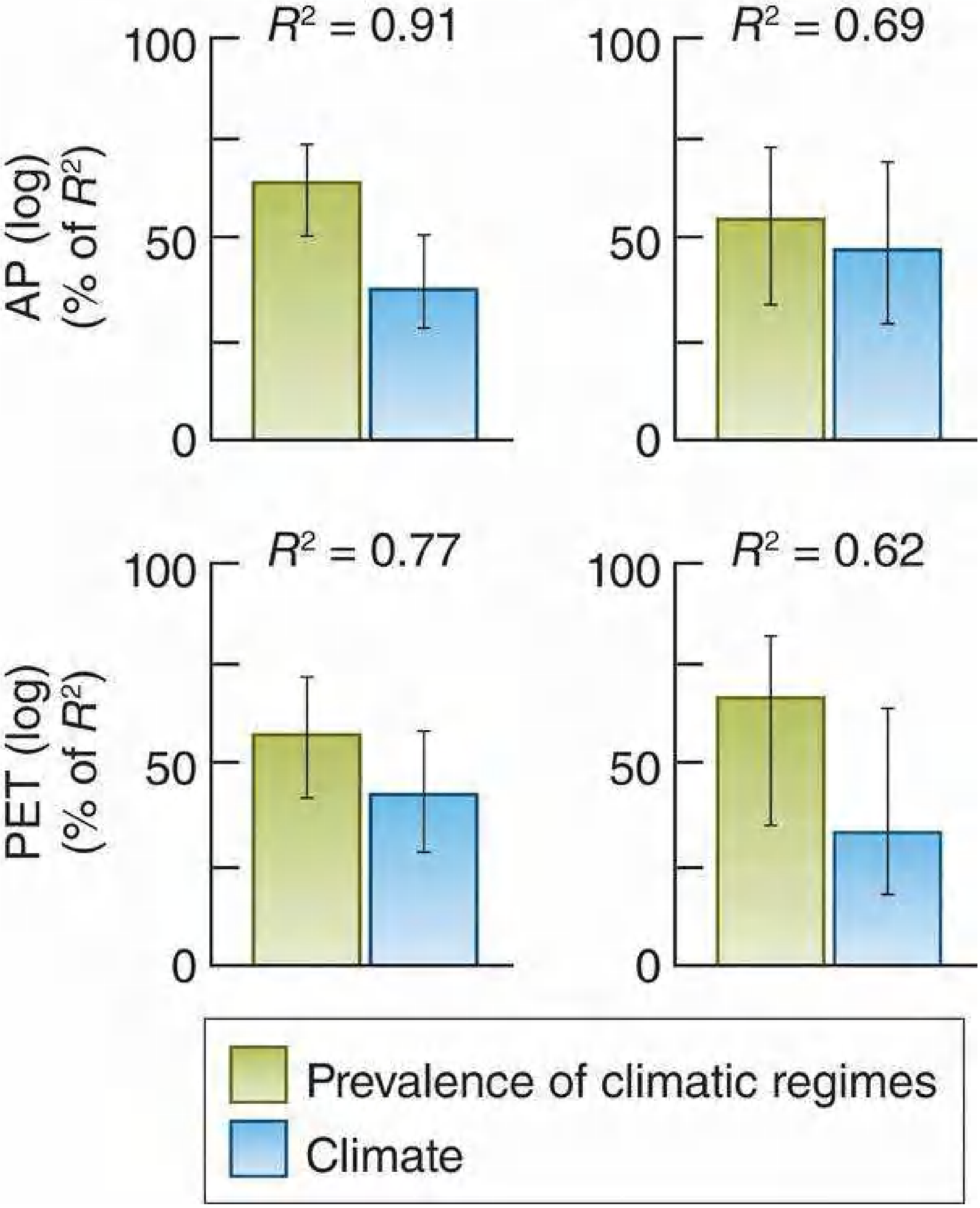
Source: (a) After Flessa & Jablonski (1995). (b) After Sutton & Collins (1991). (c) After Rosenzweig & Sandlin (1997). (d) After Kerkhoff *et al*. (2014).

energy as a core explanation

No doubt the most commonly proposed explanations for the latitudinal patterns have been variations in productivity and energy, which, as we discussed in Section 19.3, may often be difficult to disentangle. It was apparent there that variations in energy input, in their own right, are particularly important at 'global' scales, and the latitudinal trends certainly come into that category (see Figure 19.10). Also, while richness may often peak at intermediate productivities, this is least likely to be observed at larger spatial scales, which again certainly applies to global latitudinal trends. It seems likely, therefore, that the underlying variation in energy input, acting directly on metabolism and the consequences of metabolism at all trophic levels, but acting too through gradients in primary productivity, is a core explanation for the latitudinal gradient in species richness (latitudinal trends in productivity are dealt with in Section 20.2.1). On the other hand, to repeat the general point made earlier, this should not blind us to the exceptions to this trend (new, small or remote islands, etc.), where richness is no doubt more dependent on factors other than energy input.

What is more, there is support, too, for other explanations for the latitudinal trend. For example, Meyer and Pie (2018) have pointed out that while temperatures are higher and levels of precipitation typically greater at lower latitudes, those climatic regimes with higher temperatures and greater precipitation tend also to be more prevalent on a global scale. Thus, when climate itself and the prevalence of climatic regimes are compared in their ability to account for the species richnesses of mammals and amphibians, environmental prevalence is consistently the more powerful (Figure 19.31). More prevalent regimes may then allow greater specialisation (see Figure 19.3b), and some, at least, of the greater richness of the tropics may be attributable to the global prevalence of the climatic regimes represented there.


Mammals
Amphibians
$R^2 = 0.79$
$R^2 = 0.74$
AMT (% of R^2)
100
50
0
$R^2 = 0.9$
$R^2 = 0.87$
MTWQ (% of R^2)
$R^2 = 0.89$
$R^2 = 0.9$
MTCQ (% of R^2)



$R^2 = 0.91$
$R^2 = 0.69$
$R^2 = 0.77$
$R^2 = 0.62$
AP (log) (% of R^2)
PET (log) (% of R^2)
100
50
0
Prevalence of climatic regimes
Climate

Figure 19.31 Mammal and amphibian species richness is better predicted by the prevalence of climatic regimes than by climate itself. The relative contributions of the prevalence of climatic regimes and of climate itself to non-linear regressions accounting for the species richness (number of species per unit area) of mammals (left) and amphibians (right) are shown at a global scale, estimated for cells 110 km^2 in area, for five climatic variables, as indicated. AMT is annual mean temperature with regimes defined by 2°C intervals. MTWQ is the mean temperature of the warmest quarter (2°C intervals). MTCQ is mean temperature of the coldest quarter (2°C intervals). AP is annual precipitation (intervals of 0.2 log(mm) per year). PET is annual potential evapotranspiration (intervals of 0.07 log(mm) per year). Column heights are normalised to sum to 100% of the overall R^2 indicated. Error bars are 95% CIs.

Source: After Meyer & Pie (2018).

The greater evolutionary age of the tropics has also been proposed as a reason for their greater species richness, for example in the tropical conservatism hypothesis (Wiens & Donogue, 2004), which emphasises that many groups with high tropical richness originated in the tropics and have spread to temperate regions only more recently, if at all – they have been evolutionarily conservative in their tolerance of temperate conditions. Moreover, this greater tropical richness may be related to the much greater geographic extent of tropical compared with temperate regions until relatively recently, in geological terms at least – around 30–40 million years ago. We see evidence in support of the hypothesis in Figure 19.32, based on Kerkhoff *et al.*'s (2014) study of the distribution and phylogenetic relationships of woody flowering plants in North and South America (already referred to in Figure 19.30d). As predicted by the hypothesis, the floras of temperate regions were less phylogenetically diverse, even given their species richnesses, than those of the tropics. Furthermore, the vast majority of these temperate lineages have arisen within more ancient tropical lineages since global cooling began around 36 million years ago. Nonetheless, there have been few evolutionary transitions from tropical to temperate environments, nor vice versa: 94% of evolutionary divergences in the tropics retained tropical affinities, while for temperate regions the figure was 90%. As a result, the long-established species richness of the tropics has spread polewards little, in an evolutionary sense, as the earth has cooled. Hence, the latitudinal trend may reflect the combined and complementary influences of climate and age. More temperate regions may offer fewer opportunities for the coexistence of species that would generate increases in species richness. But there may also have been insufficient time, given evolutionary conservatism, for species to evolve to exploit the opportunities that do exist there.

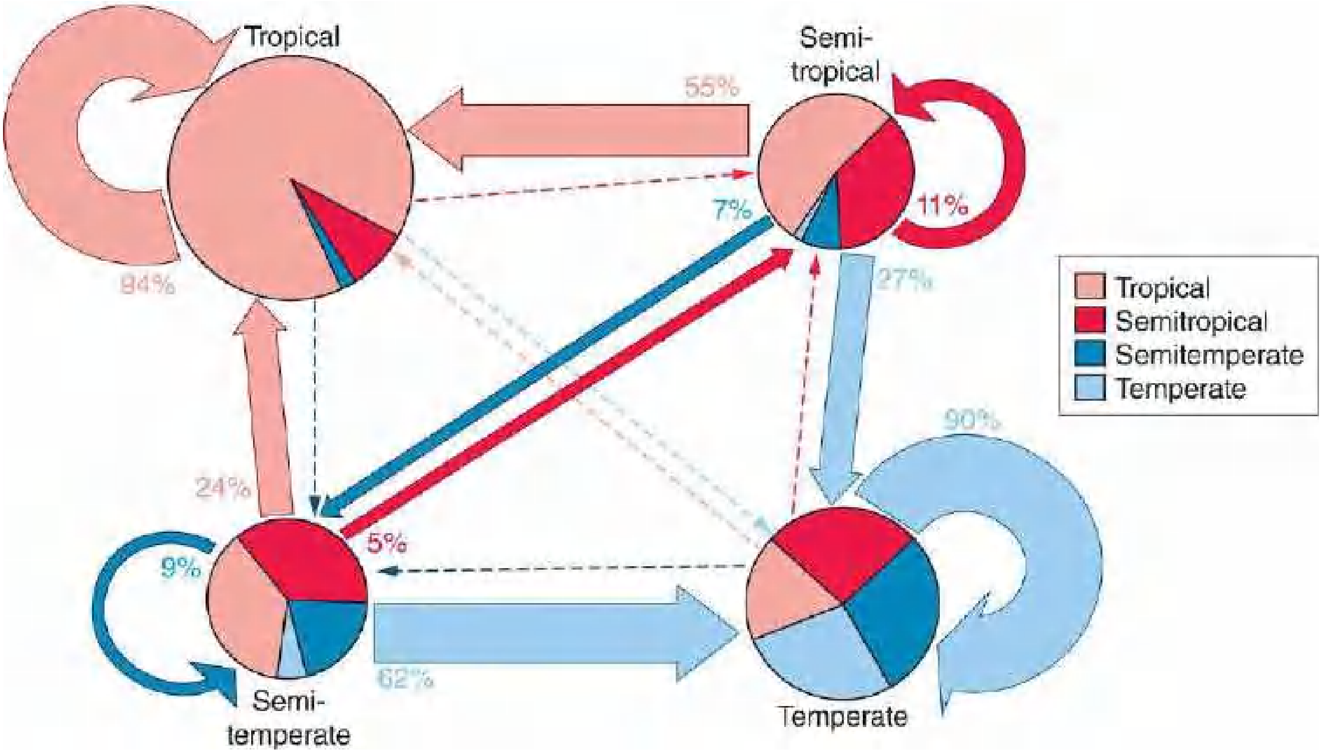


Figure 19.32 New species of woody flowering plants predominantly remain in the same latitudinal zone (temperate or tropical) as their immediate ancestors, hindering the transfer of long-established tropical richness to temperate regions. Ancestral–descendant transitions are shown within and between different latitudinal zones, as indicated. Circle areas are proportional to the number of lineages in each zone, and the coloured sectors of each circle represent the fractional ancestries of each zone: pink is tropical (more than three-quarters of the latitudinal range between the tropics), dark pink is semitropical (between half and three-quarters between the tropics), dark blue is semitemperate (between half and three-quarters outside the tropics) and light blue is temperate (more than three-quarters outside the tropics). The breadths of arrows (similarly colour coded) represent the fractions of transitions from each zone to each zone, as indicated (dashed arrows <5%).

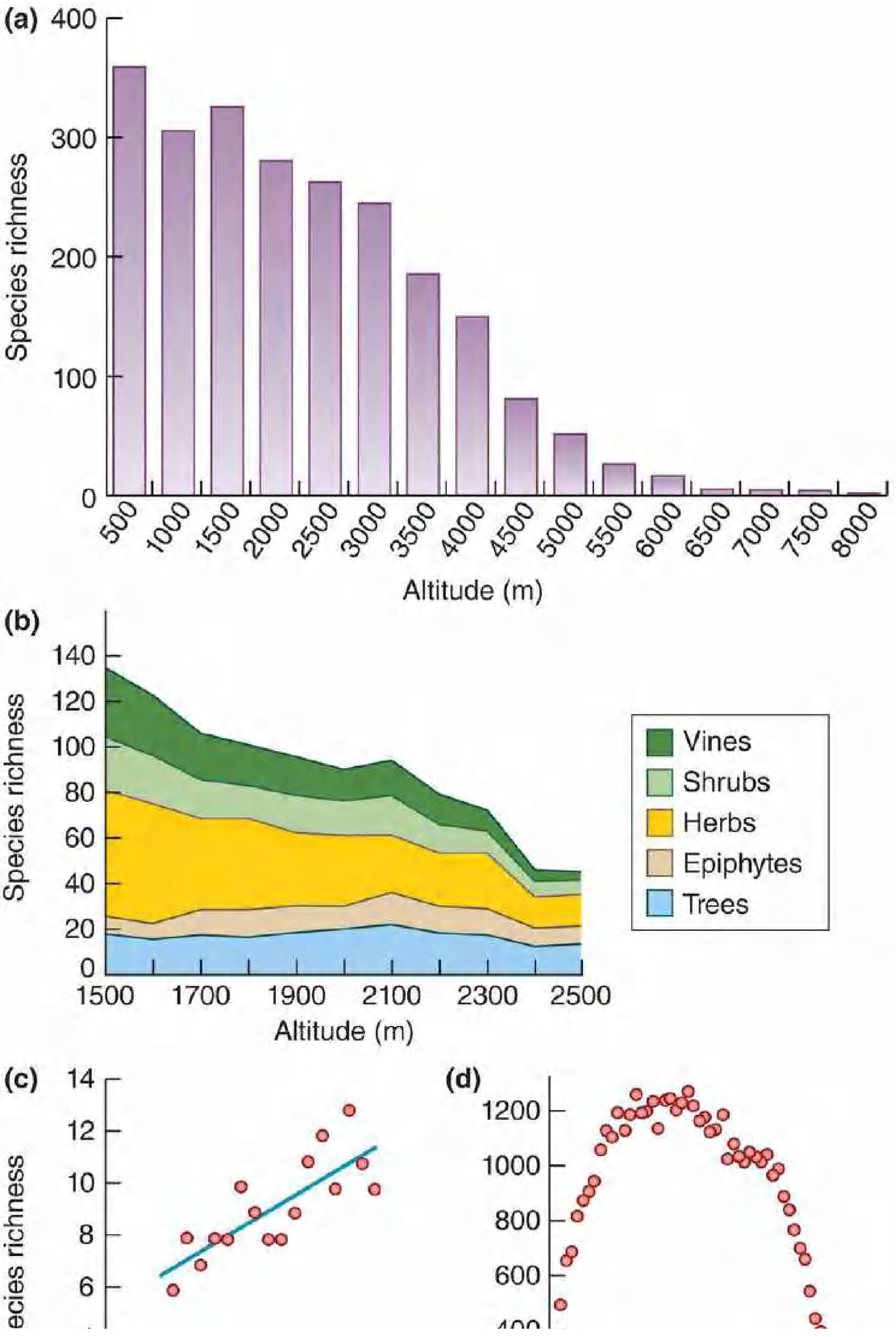
Source: After Kerkhoff *et al.* (2014).

More intense predation has also been proposed as a reason for greater species richness in the tropics, reducing the importance of competition and permitting greater niche overlap (see Figure 19.3c). However, predation cannot be put forward as the root cause of tropical richness, since this simply invites the question of what gives rise to the richness and abundance of the predators themselves. But focusing on predation raises a point that applies to all trends in species richness: the various factors acting on a trend may interact with and even reinforce one another. For example, greater energy may lead to higher species richness, to more predation, and hence to even higher species richness.

19.6.2 Gradients with elevation and depth

decreasing, increasing or hump-shaped richness relationships with elevation

A decrease in species richness with elevation has frequently been reported in terrestrial environments (e.g. Figure 19.33a, b). (‘Elevation’ is the height of a location above sea level, in contrast to ‘altitude’, which is its height relative to a local ground level.) At least some of the factors instrumental in the latitudinal trend in richness are also likely to be important as explanations for such elevational trends. For example, declines in species richness have often been explained in terms of decreasing productivity associated with lower temperatures and shorter growing seasons at higher elevations, or physiological stress associated with climatic extremes near mountaintops. On the other hand, high-elevation communities may occupy smaller areas than lowlands at equivalent latitudes, and they will usually be more isolated from similar communities than lowland sites. Therefore the effects of area and isolation are likely to contribute to observed decreases in species richness with elevation.


(a)
Species richness
400
300
200
100
0
Altitude (m)
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
6500
7000
7500
8000
(b)
Species richness
140
120
100
80
60
40
20
0
Altitude (m)
1500
1700
1900
2100
2300
2500
Vines
Shrubs
Herbs
Epiphytes
Trees
(c)
ecies richness
14
12
10
8
6
(d)
1200
1000
800
600

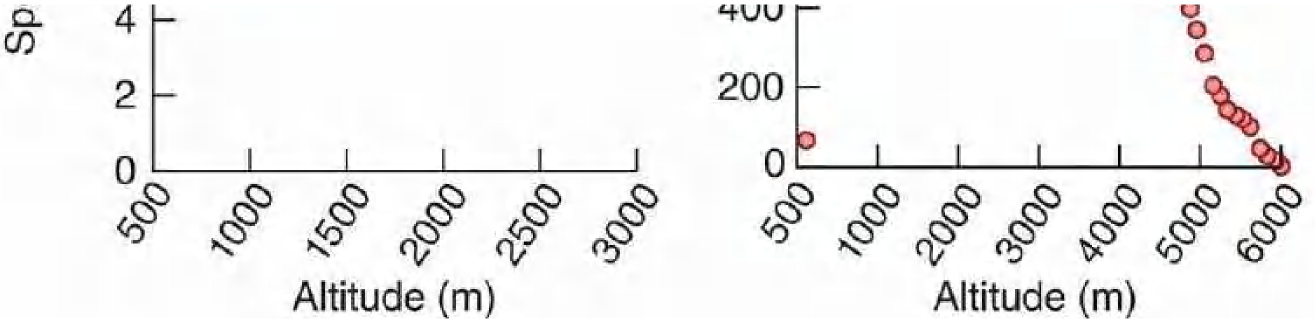


Figure 19.33 Species richness may decrease, increase or show a hump-shaped relationship with elevation. Relationships between species richness and elevation for: (a) breeding birds in the Nepalese Himalayas, (b) plants in the Sierra Manantlán, Mexico, (c) ants in Lee Canyon in the Spring Mountains of Nevada, USA, and (d) flowering plants in the Nepalese Himalayas.

Source: (a) After Hunter & Yonzon (1992). (b) After Vázquez & Givnish (1998). (c) After Sanders *et al.* (2003). (d) After Grytnes & Vetaas (2002).

Moreover, some have reported a monotonic *in*crease in richness with elevation, which itself may reflect an increase in productivity. Figure 19.33c, for example, illustrates a positive relationship between ant species richness and elevation in an arid region, where precipitation increased with elevation, resulting in higher productivity.

biotic and geometric explanations for mid-altitude peaks

In fact, though, about 50% of studies have described hump-shaped patterns (e.g. Figure 19.33d) – a mid-elevation peak in species richness (Rahbek, 1995). Explanations for this fall broadly into two types. In the first, the underlying drivers of richness themselves generate the hump, for example because productivity declines with elevation, but species richness peaks at intermediate productivities (see Section 19.3.1). Alternatively, factors may interact to generate optimal conditions at intermediate elevations. For example, in arid regions, productivity may be low at the lowest elevations because high temperatures are combined with a scarcity of water, but low too at the highest elevations because temperatures are low. Species richness is then highest where temperatures are moderate and water in ready supply.

There can, though, also be ‘geometric’ explanations for these mid-elevation peaks. The *mid-domain effect*, in particular (Colwell & Hurtt, 1994) assumes, as a null model, that if species distributions are arranged at random right across the gradient, the greatest overlap, and hence the greatest richness, will occur at intermediate elevations. Furthermore, if we acknowledge that mountains are mostly not isolated cones (where the higher the elevation is, the rarer the habitat) but complex landscapes with multiple, local valleys and peaks, then mid-elevation habitats are the commonest (Bertuzzo *et al.*, 2016), and a species–area relationship can be invoked to explain richness peaks at these intermediate elevations. Thus, for example, Grytnes and Vetaas (2002) analysed the elevational pattern in Himalayan flowering plants illustrated in Figure 19.33d and found that this fitted best to a model that combined the mid-domain effect with an underlying monotonic decline in richness with elevation – for which declining temperatures would be a plausible but unproven explanation.

Of course, this combination of factors need not be replicated in studies of other groups of species. Figure 19.34, for instance, shows some results from a worldwide study of the elevational species richness relationships for geometrid moths (19 elevational gradients in all). Most of the 19 showed a mid-elevation peak in richness (Figure 19.34a), and this was irrespective of climate or geographic location (and so was not confined, for example, to arid zones). Although single measures like temperature or humidity had explanatory power in a few individual cases (and so may have been proposed had one of these been the only study site), they had little or no support in statistical models that sought the optimal combination of factors in accounting for the distributions overall. The two factors that received the strongest support (Figure 19.34b) were the area occupied by the elevation band in question and productive space (see Section 17.2.8), that is, productivity multiplied by area. The mid-domain effect also received good support, whereas productivity and temperature effects had weak support and precipitation-related factors none at all.

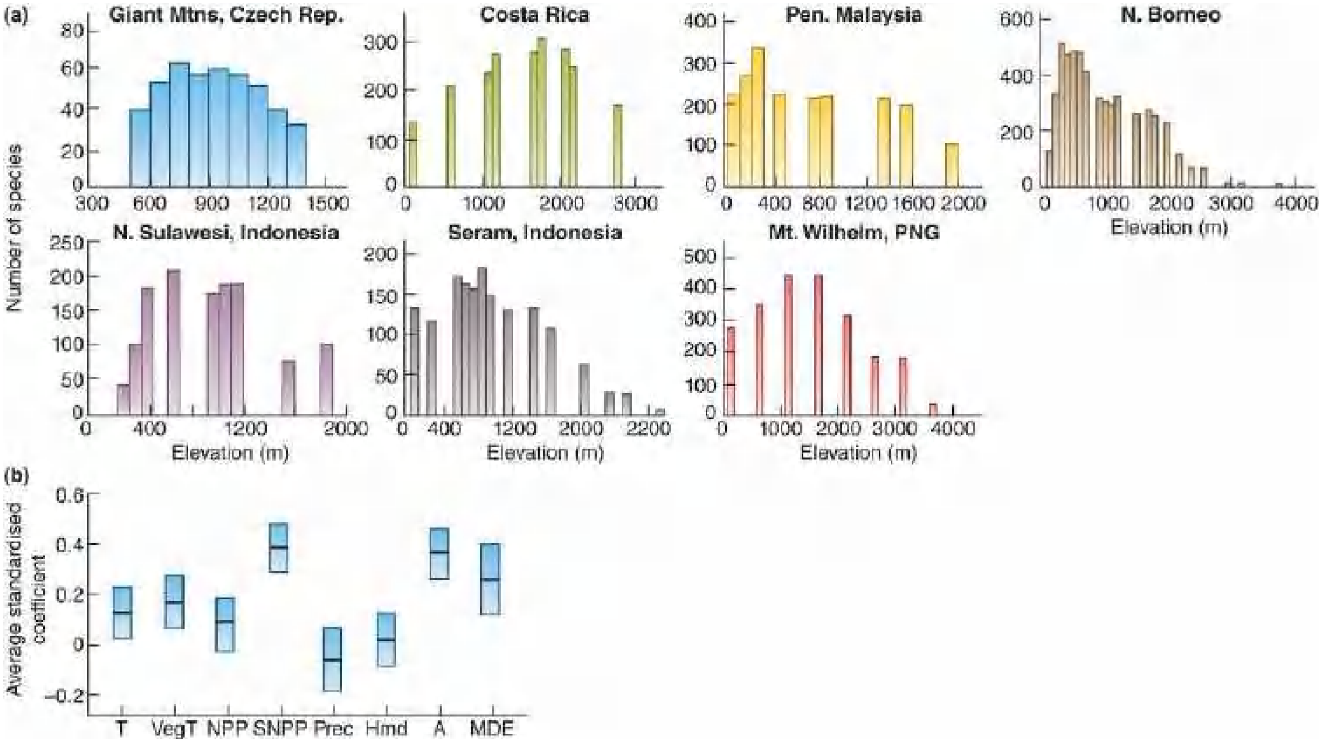


Figure 19.34 Mid-elevation peaks in species richness of geometrid moths are best accounted for by effects of area and productivity and mid-domain effects. (a) Mid-elevation peaks in moth richness for a selection of sites, as indicated. (b) Analysis for 19 sites (including the seven in (a)) of the strengths of eight factors, as indicated, in accounting for elevational distributions overall. The values are standardised coefficients averaged across 44 candidate models incorporating the eight factors. T, mean annual temperature; VegT, mean non-freezing temperature; NPP, net primary productivity; SNPP, NPP multiplied by areal extent of an elevation class; Prec, precipitation; Hmd, humidity; A, areal extent of an elevation class; MDE, the estimated mid-domain effect. Means and 95% CIs are shown.

Source: After Beck *et al.* (2017).

This illustrates another point that is pertinent to all trends in species richness: a multiplicity of factors may account for any particular pattern, and which does so, or which combination of factors does so, is likely to vary from case to case. We should not expect single, universal explanations.

patterns with depth in aquatic environments

In aquatic environments, the change in species richness with depth shows some similarities to the terrestrial gradient with elevation. In larger lakes, the cold, dark, oxygen-poor abyssal depths contain fewer species than the shallow surface waters. Likewise, in marine habitats, plants are confined to the photic zone (where they can photosynthesise), which rarely extends below 30 m. In the open ocean, therefore, there is a rapid decrease in richness with depth, reversed only by the variety of often bizarre animals living on the ocean floor.

Variations in species richness with depth among bottom-dwelling communities are inevitably more complex, since we are mostly comparing communities living at different depths at different locations, rather than communities at the same location distributed along a continuous slope. Hence, as with altitudinal patterns, they are affected by depth itself but also by what is above and around them. This is illustrated by the patterns in species richness with depth, and latitude, exhibited by the marine ophiuroid echinoderms, brittle stars (Figure 19.35). The effects of depth and latitude interact. Close to the equator, richness is clearly higher in shallower seas. At much higher latitudes, however, around 30–50°, no such clear trend is apparent. Thus, whereas in shallow seas, richness is highest in more tropical regions, in much deeper seas, richness peaks at intermediate latitudes.

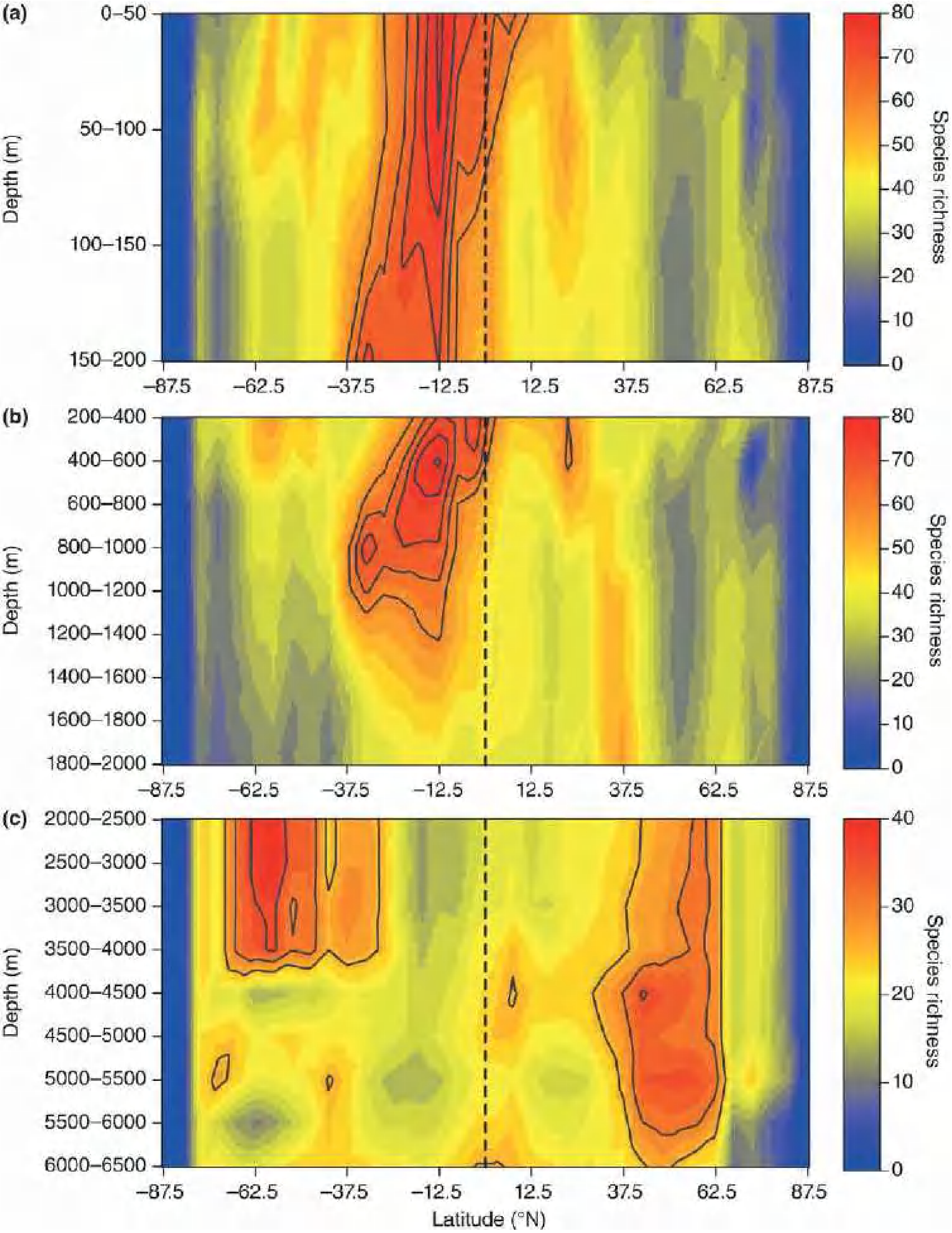

(a)
Depth (m)
0–50
50–100
100–150
150–200
−87.5
−62.5
−37.5
−12.5
12.5
37.5
62.5
87.5
Species richness
0
10
20
30
40
50
60
70
80
(b)
Depth (m)
200–400
400–600
600–800
800–1000
1000–1200
1200–1400
1400–1600
1600–1800
1800–2000
−87.5
−62.5
−37.5
−12.5
12.5
37.5
62.5
87.5
Species richness
0
10
20
30
40
50
60
70
80
(c)
Depth (m)
2000–2500
2500–3000
3000–3500
3500–4000
4000–4500
4500–5000
5000–5500
5500–6000
6000–6500
−87.5
−62.5
−37.5
−12.5
12.5
37.5
62.5
87.5
Latitude (°N)
Species richness
0
10
20
30
40

Figure 19.35 Brittle stars in shallow seas show peak richness near the equator, but those at much deeper sites peak at intermediate latitudes. The estimated distribution of the species richness of brittle stars with depth and latitude is shown based on combining results from 1614 research expeditions carried out over 130 years. Separate estimations were carried out for three different depth zones (a, b and c), as indicated. Note that the species richness scale is different for the lowest depths. The vertical dashed line indicates the position of the equator. The black contour lines encompass the top 20% of richnesses at each depth.

Source: After Woolley *et al*. (2016).

Statistical models fitted to account for these variations in brittle star richness found important roles for mean annual temperature in shelf and upper slope environments (less than 2000 m in depth) and, in the shallowest seas, an important role too for high seasonal peaks in net primary productivity (NPP). At greater depths, however, temperature was not important, though the seasonal peaks in NPP were, as was the flux of organic carbon from above. There was a role, too, at these deep sites, for the distance from the continental margin. Taking the joint latitude–depth distribution together, therefore, energy, productivity and isolation are all important – another example with multiple, interacting factors.

APPLICATION 19.4 Marine protected areas

Most marine protected areas (MPAs) are in coastal areas, and since coastal ecosystems make up a relatively small portion of the global oceans, MPAs occupy a larger percentage of these coastal ecosystems: almost 3% of the waters lying within 320 km of coastlines (the ‘200-mile’ limit of territorial waters claimed by most of the world’s nations as part of their territories) and over 6% within 20 km of coastlines (the ‘12-mile’ limit that many nations more intensively manage). In contrast, the pelagic zone of the open ocean comprises 99% of the volume of the world’s biosphere but currently only a small portion has any protection for pelagic fish, invertebrates, cetaceans and seabirds (e.g. the 360 000 km^2 Papahānaumokuākea Marine National Monument of Hawaii and the 87 000 km^2 Pelagos sanctuary in the north-west Mediterranean) (Game *et* al., 2009 ; Young *et al.*, 2015). In fact, considerable ingenuity will be required to plan appropriate pelagic MPAs (e.g. Robertson *et al.*, 2017), not least because pelagic systems respond to ocean currents and are not static like coral reefs or benthic communities.

As the name suggests, the space included in an MPA is all or mostly marine, although portions of adjacent land areas are often included, particularly when the land is intimately tied to the marine ecosystem in function or culturally. For instance, a tropical MPA may include an atoll as well as the associated fringing coral reefs and adjacent marine waters. MPAs were developed first in the 1960s as a way to better protect marine biodiversity and fishery resources, using as an analogy the way national parks on land had helped conserve some of the ecological functioning of terrestrial ecosystems. By 2019, MPAs around the world had grown to include almost 17 000 sites encompassing almost 8% of the world’s oceans (Protected Planet, 2019).

MPAs vary tremendously in their effectiveness in conserving biodiversity and managing fishery resources. Coral reefs are a particular focus for many MPAs, yet only a relatively few MPAs have been judged to be fully effective in protecting coral diversity and health (Figure 19.36).

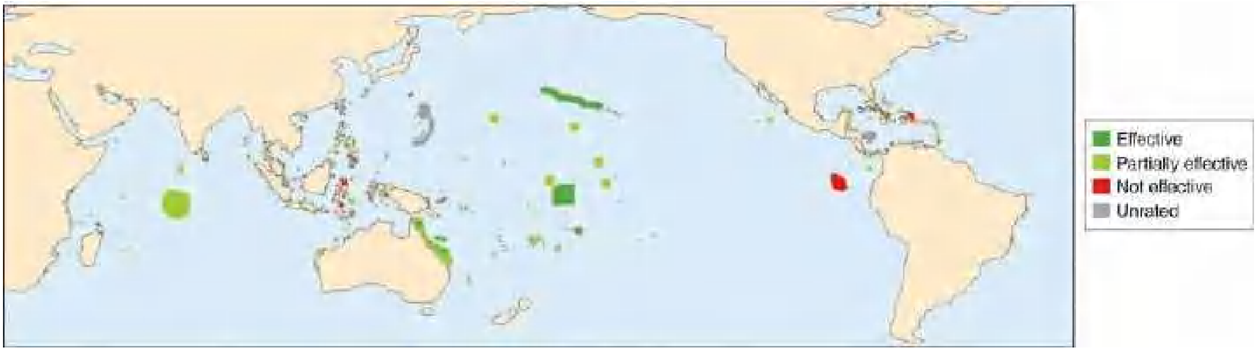


Figure 19.36 **Marine-protected areas in coral reef regions**. These have been established in many regions, but only a relatively few of these are judged to be fully effective in protecting the corals. Effectiveness was rated by regional experts using a three-point score. 1, ‘Effective’: managed sufficiently well that local threats were not undermining natural ecosystem function. 2, ‘Partially effective’: managed such that local threats were significantly lower than adjacent non-managed sites, but there may still be some detrimental effects on ecosystem function. 3, ‘Not effective’: unmanaged, or management was insufficient to reduce local threats in any meaningful way.

Source: After Burke *et al.* (2011).

no-take zones

Some MPAs are *no-take zones*: areas where no fishing is allowed. These areas are also well protected from land-based pollution sources and other human disturbances. In no-take zone MPAs, biodiversity is often maintained, and fishery resources grow in size and even help replenish fish populations in adjacent waters where fishing is still allowed. Hence, these MPAs are an overall benefit to commercial fishers despite being off limits to them. The Goat Island Marine Reserve in New Zealand, established in 1977, provides an example of such success – one that many subsequent MPAs have tried to emulate. Populations of snapper fish and crayfish that had been decimated by overfishing were re-established within 10 years, and now these and the near-pristine coastal ecosystem support both scientific research and extensive ecotourism.

However, no-take zones do not always accomplish this goal. Sometimes the populations of predatory fish increase greatly, leading to a trophic cascade of fewer herbivores, an increase in macroalgae and a decrease in biodiversity. This may be more likely to happen in areas under stress for other reasons, such as nutrient pollution, where the lack of predatory fish and the resulting high abundance of herbivores can mask some of the symptoms of nutrient pollution. Clearly, the ecological details matter. Most MPAs, however, are not no-take zones, and fishing and many other practices are commonly allowed. In the USA, less than 1% of the area of all MPAs consists of no-take zones.

MPAs must be sufficiently large to be effective

The size of an MPA matters. Fish will of course swim freely in and out of the defined area of an MPA. To protect a fish population, therefore, the size of the MPA must be large relative to the size of the home range of the fish of most concern. One calculation suggests that the MPA should be around 12.5 times the area of the species' home range, in order to keep fishing pressure on the population within the MPA at 2% or less of the pressure outside (Kramer & Chapman 1999). Since home range increases with body size, this means that for a typical species 20 cm in length, the MPA should be 1.8 km^2 in area. However, 30% of the world's MPAs are smaller than this, providing inadequate protection to typical fish species that grow to 20 cm in length or more.

limits to the powers of MPAs

While MPAs are an important step forward in the effort to protect marine ecosystems, they cannot do the job alone. They are focused on local protection of resources and provide little protection against global and regional threats. Global change poses high risk to many coastal ecosystems. Coral reefs, for example, are endangered by the increasing acidification of the world's oceans as levels of atmospheric carbon dioxide continue to rise. Corals are damaged too by rising temperatures of surface-ocean waters, which can cause coral 'bleaching': the loss of symbiotic algae in the corals, which causes the coral to lose colour and become white (see Section 13.8).

More generally, regional-scale nitrogen pollution can lead to eutrophication, dead zones and loss of biodiversity and fishery resources at scales much larger than those of most MPAs. Hence, despite the growing area covered by MPAs, global biodiversity continues to fall, and the health of corals continues to decline (see Figures 19.1 and 19.36). The trend globally for MPAs is similar to that for protected areas on land. For both, while protected areas have increased in size and number, they remain only a small percentage of the total, and biodiversity is clearly not adequately conserved.

19.6.3 Gradients during community succession

We saw in Section 18.4 how, in community successions, if they run their full course, the number of species tends first to increase (because of colonisation) but eventually to decrease (because of competition). This is most firmly established for plants, but the relatively few studies that have been carried out on animals in successions indicate, at least, a parallel increase in species richness in the early stages of succession. For example, in Figure 19.37 we see increases in bird species richness following reclamation of open-cast mines in the Czech Republic (Figure 19.37a), and in amphibians following the abandonment of old corn fields in Mexico (Figure 19.37b). However, we see a decrease in habitat generalists following initial colonisation among small mammals following regrowth of Atlantic forest in Brazil (Figure 19.37c), and no significant increase in habitat specialists (Figure 19.37d).

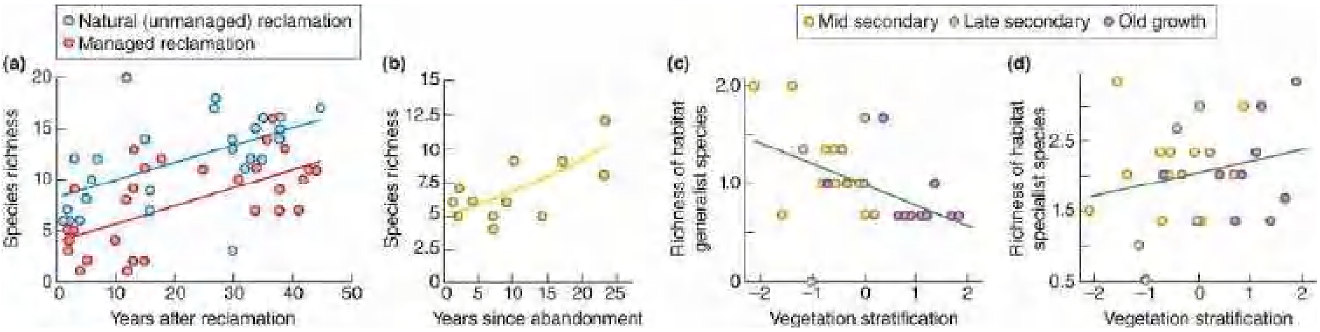


Figure 19.37 Animal species richness tends to increase during successions, though there may be a loss of initial colonisers. (a) Bird species richness increased during either natural or managed reclamation of abandoned open-cast mines in the Czech Republic. Managed reclamation involved uniform planting of grass and trees. Site age effect: $F^2 = 33.4$, $P < 0.001$. Management effect: $F^2 = 24.8$, $P < 0.001$. (b) The species richness (per 80 person-hours of observation) of amphibians increased following the abandonment of corn fields in Mexico. $r^2 = 0.52$, $P < 0.05$. (c) The richness of generalist small mammals decreased during regrowth of Atlantic forest in Brazil. The successional stages are mid secondary, late secondary and old growth. Vegetation stratification is a proxy for successional status on a continuous scale. Coefficient = –0.21, SE = 0.08. (d) The equivalent to (c) but for forest specialists, where richness increased but not significantly. Coefficient = 0.16, SE = 0.14.

Source: (a) After Salek (2012). (b) After Hernandez-Ordonez *et al.* (2015). (c, d) After Pinotti *et al.* (2015).

To a certain extent, the successional gradient is a necessary consequence of the gradual colonisation of an area by species from surrounding communities that are at later successional stages; that is, later stages are more fully saturated with species (see Figure 19.3d). However, this is a small part of the story, since succession involves a process of replacement of species (see Figure 19.37c, d) and not just the mere addition of new ones. Indeed, as with the other gradients in species richness, there is something of a cascade effect with succession: one process that increases richness kick-starts a second, which feeds into a third, and so on.

The earliest species will be those that are the best colonisers. They immediately provide resources (and introduce heterogeneity) that were not previously present. For example, the earliest plants generate resource-depletion zones in the soil that inevitably increase the spatial heterogeneity of plant nutrients. The plants themselves provide a new variety of microhabitats, and for the animals that might feed on them they provide a much greater range of food resources (see Figure 19.3a).

The increase in herbivory and predation may then feed back to promote further increases in species richness (predator-mediated coexistence: see Figure 19.3c), which provides further resources and more heterogeneity, and so on. In addition, temperature, humidity and wind speed are much less variable (over time) within a forest than in an exposed early successional stage, and the enhanced constancy of the environment may provide a stability of conditions and resources that permits specialist species to build up populations and persist (see Figure 19.3b). As with the other gradients, the interaction of many factors makes it difficult to disentangle their actions, but with the successional gradient of richness, the tangled web of cause and effect appears to be of the essence. The relevance of succession theory to restoration and conservation was dealt with in Chapter 18 (Application 18.2).

19.7 Selecting areas for conservation

is conservation focused on biodiversity hot spots?

We are bound to want to devise priorities so that the restricted number of new protected areas, in terrestrial and marine settings, can be evaluated systematically and chosen with care. We know that the biotas of different locations vary in species richness (with particular centres of diversity), in the extent to which the biota is unique (with centres of endemism), and in the extent to which the biota is endangered (with hot spots of extinction, for example because of imminent habitat destruction). Locally, no doubt, these are important considerations when protected areas are chosen. But on a global scale, it is not so clear that protected areas are most concentrated where they are most needed.

For example, if we compare the global distribution of ‘biodiversity hot spots’ (as indicated by the number of globally threatened birds and amphibians in an area) with the global distribution of protected areas (Figure 19.38), we can see how the needs of conservation are being much better satisfied in the richer than in the poorer parts of the world. The reality is that protected areas will be placed not only where they are most needed but also where, simply, it is possible to place them.

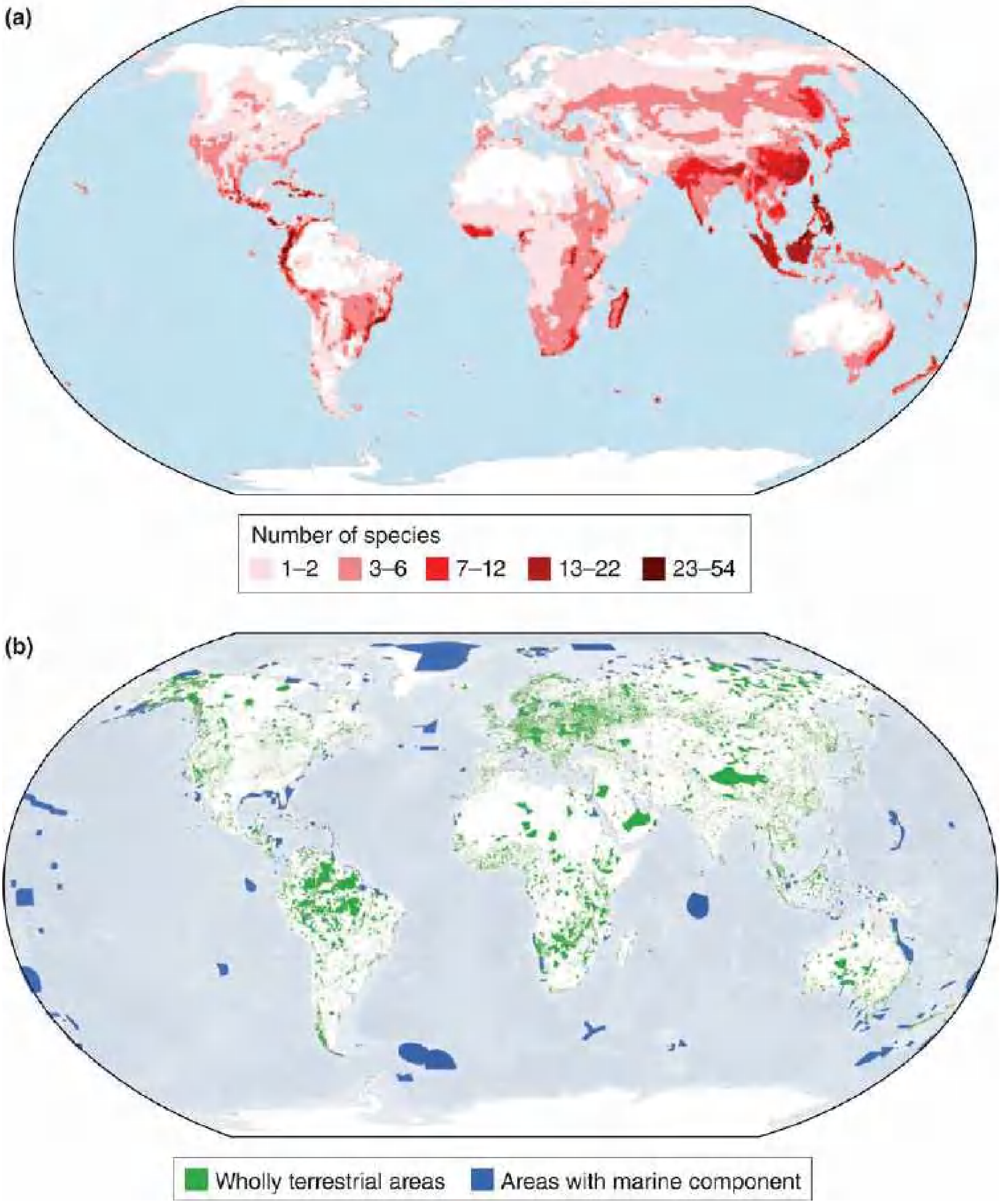
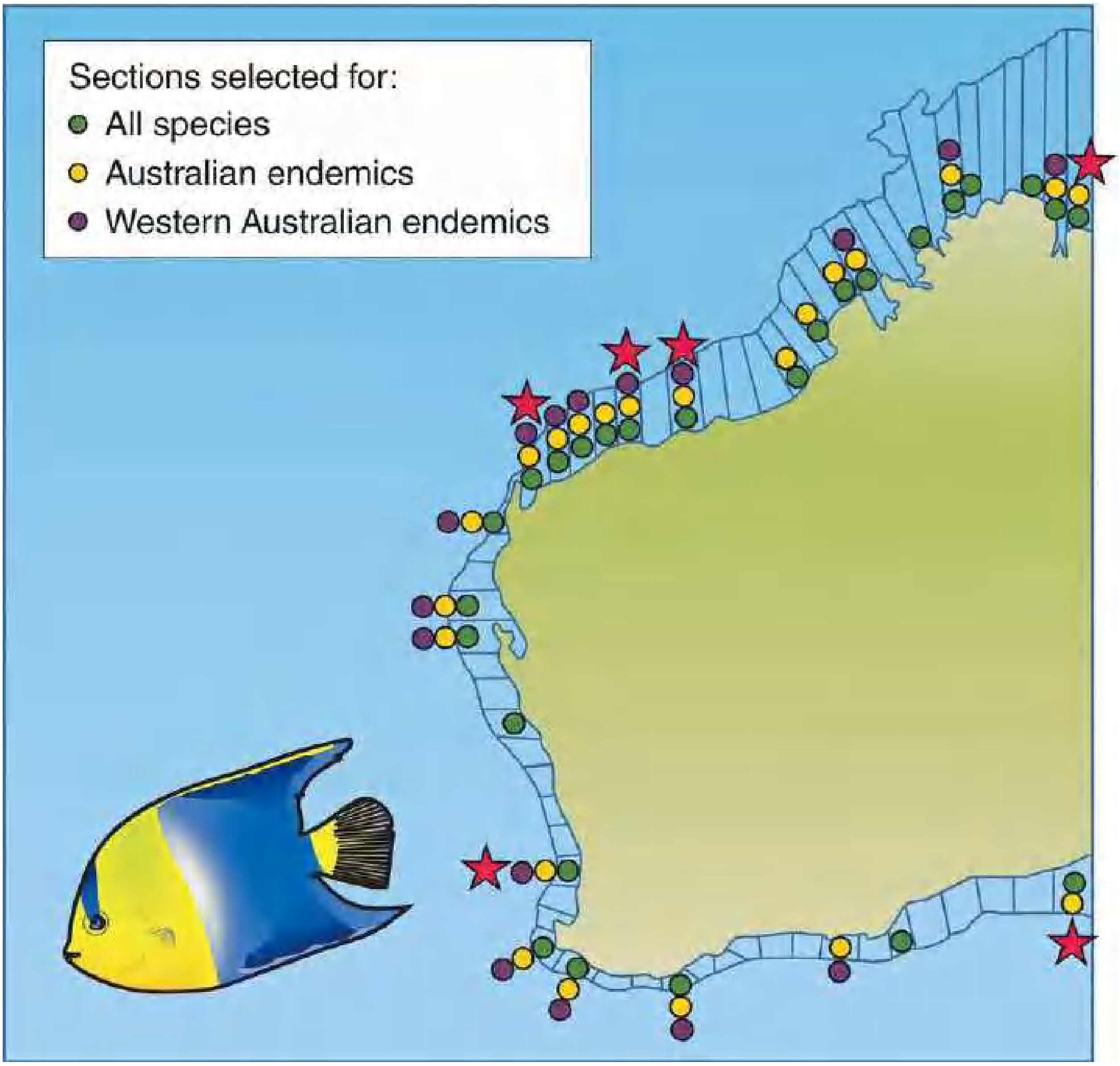


Figure 19.38 Protected areas are concentrated more in the 'rich world', especially Western Europe, despite this not being the focus of biodiversity hot spots. (a) Distribution of biodiversity hot spots as measured by the numbers of species of globally threatened birds plus amphibians mapped on an equal area basis (each grid cell is 3113 km^2). (b) Global distribution of the 177 547 nationally designated protected areas. Wholly terrestrial areas are green; those with at least a marine component are blue.

Source: (a) After Rodrigues *et al*. (2006). (b) After Bertzky *et al*. (2012).

APPLICATION 19.5 Site selection based on complementarity and irreplaceability

Choosing the best areas for conservation is not limited to identifying the best individual sites. Conservation programmes typically seek to identify whole sets of sites that are collectively best suited to achieve the conservation aims in view. One approach is *complementarity selection*. Here, we proceed in a stepwise fashion, starting with the single best site, but then selecting at each step the site that is most complementary to those already selected, that provides most of what the existing sites lack. In the case of the coastal marine fishes around Western Australia, for example, the results of a complementarity analysis showed that more than 95% of the total of 1855 species could be represented in just six appropriately located sections, each 100 km long (Figure 19.39).



[Figure 19.39](#) **Just six areas in Western Australia would conserve 95% of the fish species there.** The coastline of Western Australia divided into 100 km lengths and showing the results of complementarity analysis to identify the minimum number of sites needed to include all the fish biodiversity for the region. Analyses were performed using all fish species, and separately for species endemic to Australia (found nowhere else) or those endemic to Western Australia. In the case of total fish biodiversity, 26 areas were needed if all 1855 fish species were to be incorporated (green circles) but only six areas (red stars) would be needed to incorporate more than 95% of the total.

Source: After Fox & Beckley (2005).

We might also consider the *irreplaceability* of each potential area, defined as the likelihood of it being necessary if we are to achieve conservation targets for it to be included or, conversely, the likelihood that one or more targets will not be achieved if the area is not included. Cowling *et al.* (2003) used irreplaceability analysis as part of their conservation plan for South Africa's Cape Floristic Province – a global hot spot with more than 9000 plant species. A variety of conservation targets were identified, including, among others, the minimum acceptable number of species of *Protea* plants to be safeguarded (for which the region is famous), the minimum permissible number of ecosystem types and even the minimum permissible number of individuals of large mammal species.

The researchers used an irreplaceability approach to guide the choice of areas to add to existing reserves that would best achieve the conservation targets (Figure 19.40), and they concluded that, in addition to areas that already have statutory protection, 42% of the Cape Floristic Province, comprising some 40 000 km^2, will need some level of protection. This includes all cases of high irreplaceability (>0.8). It therefore includes some areas that are unimportant in terms of *Protea* and ecosystem types but that are critical to provide for the needs of large mammals in lowland areas.

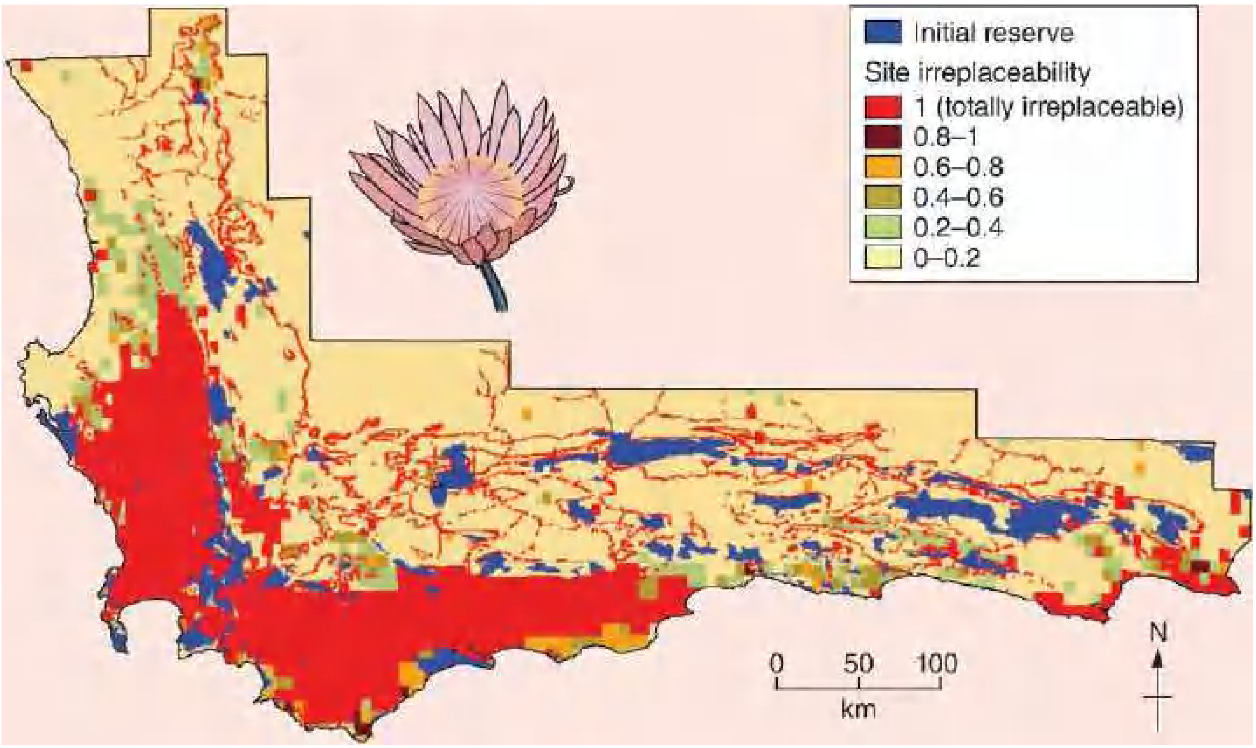


Figure 19.40 An example of irreplaceability analysis. Map of South Africa's Cape Floristic Region showing site irreplaceability values for achieving a range of conservation targets in the 20-year conservation plan for the region. Irreplaceability is a measure, varying from 0 to 1, which indicates the relative importance of an area for the achievement of regional conservation targets. Existing reserves are shown in blue.

Source: After Cowling *et al*. (2003).

19.8 Managing for multiple objectives – beyond biodiversity conservation

the triple bottom-line of sustainability

The main emphasis up to this point has been on the use of ecological theory to help solve conservation problems and establish strategies that are likely to be sustainable in the long term. However, we have already come across examples where ecological aspects of sustainability cannot be divorced from economic (see Application 19.2, and earlier examples in Applications 9.1 and 11.2) or social aspects (e.g. Applications 2.7, 11.3, 11.4 and 12.6) – the so-called triple bottom-line of sustainability.

the nature and people concept

Because of our dependence on ecosystem services (see Section 15.4.1), the degradation of ecosystems worldwide is not just a problem for biodiversity conservation but also for human well-being more generally. Sometimes managing for multiple objectives can uncover *win–win* situations, where biodiversity and certain ecosystem services can be sustained together (regulating ecosystem services such as erosion control, carbon storage, pollination potential; provisioning services such as wild foods and medicinal herbs; providing cultural services such as aesthetic fulfilment and educational, recreational and hunting opportunities). But there are also largely unavoidable *trade-off* situations where a crucial provisioning ecosystem service, such as crop and stock production, which requires the use of what would otherwise be intact wilderness, can be sustained only at the expense of biodiversity. A changing vision of sustainability, therefore, combines conservation with all the other human needs in what Mace (2014) has termed the ‘nature and people’ concept, which requires the development of relevant science within an inclusive focus so that tools and decisions can emerge that ensure a better future for both people and nature.

multiple-use reserves in the marine environment

In this context, many of the new generation of marine protected areas are designed as multiple-use reserves, accommodating many different users (environmentalists, indigenous harvesters, recreational fishers, commercial fishers, tourism operators, etc.).

APPLICATION 19.6 Marine zoning plans

New Zealand's Fiordland region – gifts, gains and china shops

A pioneering and innovative example of a marine zoning plan was the development of a management plan for the extensive Fiordland region in the south-west of New Zealand (Teirney, 2003). This was an entirely 'bottom-up' effort by the local community (with no top-down direction by government agencies), which took eight years from first meetings to the publication of a comprehensive plan. The diverse groups worked face-to-face from the beginning. The Guardians of Fiordland's Fisheries and Marine Environment comprise Maori, recreational and commercial fishers, tourism operators, marine scientists and environmentalists. While challenging to manage (a skilled facilitator was involved), this approach provides a model for minimising conflict, stimulating reciprocal learning and formulating objectives for sustainable ecosystem use that have often proved difficult to achieve by top-down means. Thus, political imperatives have often made it difficult to achieve a workable consensus among groups with very different objectives.

A significant feature of the proposal was the concept of *gifts* and *gains* by the various groups. Thus the plan called for new fishing behaviour: a reduction in bag limits for recreational fishers, the withdrawal of commercial fishers from the inner fiords and a voluntary suspension of certain customary fishing rights by Maori. In addition, a number of marine reserves and protected areas were identified to protect representative ecosystems, on the one hand, and 'china shops', on the other – areas with outstanding but vulnerable natural values. These gains in sustainability and conservation were balanced by the gift from environmentalists to refrain from pursuing their original goal of a much more extensive marine reserve programme. As a result, the plan does not represent either extreme of the resource-use spectrum – preserve everything *or* exploit as a free-for-all. Instead, a sustainable middle ground was identified, with the Maori concept of *kaitiakitanga* or guardianship at its root. The New Zealand government agreed to implement the plan in its entirety and passed the new legislation necessary.

an Italian marine zoning plan

Another example of multipurpose design in a marine setting is provided by Villa *et al.* (2002), who used a systematic approach to design one of the first marine reserve zoning plans in Italy. They involved all the different interest groups (fishing, recreation, conservation) in defining priorities, and used a geographic information system to map marine areas for different uses and degrees of protection. Italian law recognises reserves with three levels of protection: 'integral' reserves (only available for research), 'general' reserves and the less restrictive 'partial' reserves. Villa *et al.*'s starting point was to accept 'partial' and 'general' reserves but to split 'integral' reserves into two categories: no-entry, no-take zones (where only non-destructive research is permitted) and public-entry, no-take zones, which allow visitors a full experience of the reserve, apart from exploitation. Permitted activities for the four categories are shown in Table 19.2.

Table 19.2 Activities permitted or prohibited for different levels of protection in a marine reserve. There were four planned levels of protection (from left to right in order of decreasing protection) for the Asinara Island National Marine Reserve of Italy. Key to activities: A, allowed without authorisation; Aa, allowed upon authorisation; L, subject to specific limitations; P, prohibited. *Source*: After Villa *et al.* (2002).

Category	Activity	No-entry, no-take	Entry, no-take	General reserve	Partial reserve
Research	Non-destructive research	Aa	Aa	A	A
Sea access	Sailing	P	L	A	A
	Motor boating	P	P	L	L
	Swimming	P	P	A	A
Staying	Anchorage	P	P	L	L
	Mooring	P	L	Aa	A
Recreation	Diving	P	L	Aa	A
	Guided tours	P	L	Aa	A
	Recreational fishing	P	P	L	A
Exploitation	Artisanal	P	P	L	L
	Sport	P	P	P	L
	Scuba	P	P	P	P
	Commercial fishing	P	P	P	P

The next step was to produce maps of 27 variables important to one or more interest groups. These included fish diversity, fish nursery areas, sites used by life history stages of key species (e.g. limpets, sea mammals, marine birds), archaeological interest, suitability for various forms of fishing (e.g. traditional artisanal, commercial), suitability for various recreational activities (e.g. snorkelling, whale watching), tourist infrastructure and pollution status. Planning sessions with each interest group yielded weightings or relative importance values for the variables. Taking these into account, five higher level maps were produced (using an approach developed for economic analysis and urban planning known as multiple-criteria analysis): natural value of the marine environment (NVM – aggregating values related to biodiversity, rarity, crucial habitats such as nursery areas); natural value of the coastal environment (NVC – aggregating endemic coastal species including seabirds and habitat suitable for the reintroduction of turtles and seals); recreational activity value (RAV – aggregating values for all recreational activities); commercial resource value (CRV – aggregating traditional fishing sites plus other suitable areas); ease of access value (EAV – aggregating marine access routes and harbours). Aggregated maps for NVM, NVC and RAV are shown in Figure 19.41a–c.

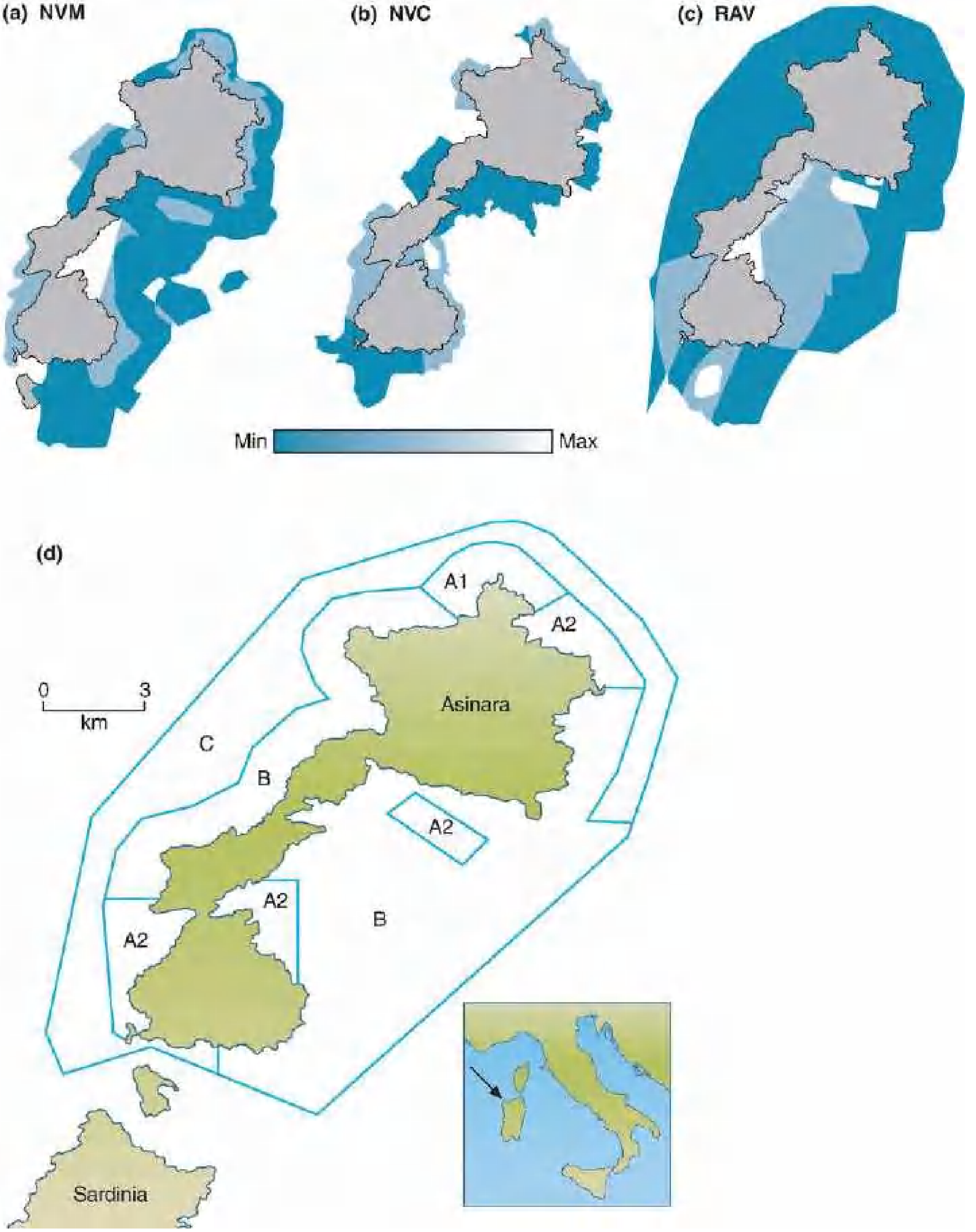


Figure 19.41 Developing a multipurpose zoning plan for the Asinara Island National Marine Reserve, Italy. Maps of the natural value of (a) the marine environment (NVM), (b) the coastal environment (NVC) and (c) recreational activities (RAV) for areas around Asinara Island (island land area is shown in the centre in grey). Lighter shades of colour represent higher values. (d) Final zoning plan: A1, no-entry, no-take; A2, entry, no-take; B, general reserve; C, partial reserve. The inset map shows the location of the reserve in relation to the mainland of Italy.

Source: After Villa *et al*. (2002).

The final stage was the production of a zoning plan. The researchers sought to avoid complex zoning that would make management and enforcement difficult and paid particular attention to the views of the various interest groups to reduce remaining conflicts to a minimum. The final plan (Figure 19.41d) had one no-entry, no-take zone (reflecting biological importance and relative remoteness), four entry, no-take zones to protect specific values such as endangered species (reflecting biological value but with easy access), two general reserve zones (to protect sensitive benthic assemblages, such as seagrass meadows that suffer little from permitted activities; see Table 19.2) and one partial reserve zone as a buffer for adjacent reserve zones (in an area where traditional fishing practices are compatible with conservation). The zoning proposal also identified three channels providing maximum boat access where environmental disturbance would be minimal.

terrestrial landscape planning

Conservation and sustainable use of land (forestry, agriculture) can also proceed hand in hand as long as planning has a scientific basis and the negotiated objectives are clear. A conservation plan that aims to enhance sustainability should try to maximise the co-benefits from ecosystem services and biodiversity conservation goals (the win–win situations of ‘land sharing’) as well as minimising trade-offs (‘land sparing’) (see Application 19.2) by allocating land uniquely to provisioning ecosystem services that are incompatible with biodiversity conservation (biomass production and food production), while designating other areas for biodiversity protection.

Such integrated planning is receiving increasing support from international policies. The European Union, for example, is working on designing a network of so-called ‘green infrastructure’ to conserve and reconnect fragmented natural areas (highlighting the importance of connectivity; see Application 19.3), while at the same time fostering the maintenance of a wide range of ecosystem services and their associated socioeconomic benefits (European Commission, 2018). By including the green infrastructure, the planning and decision-making process should help maintain both the supply of ecosystem services and biodiversity goals, especially those threatened by land-use changes, and to improve and restore ecosystem functioning more generally. An example of how this might be achieved is described in Application 19.7.

APPLICATION 19.7 Holistic landscape planning for Catalonia, Spain

Catalonia, in north-east Spain, has a population of 7.5 million people, covers an area of 32 000 km^2, and has a central depression that is bordered by a coastal mountain range along its Mediterranean coast and by the Pyrenees in the north. Agriculture currently comprises 35% of the total area of Catalonia, mostly in the central depression. Twenty nine per cent of Catalonia has diverse conservation status (with varying degrees of protection), while 61% of the total area is forest (conservation and forestry are not mutually exclusive).

Lanzas *et al.* (2019) used a systematic planning approach (called 'Marxan with zones'; see Watts *et al.*, 2009) aimed at identifying three kinds of zone: (i) core areas for biodiversity conservation (conservation zones), (ii) areas of connection between core conservation areas that also have important value for delivery of ecosystem services (green infrastructure zones), and (iii) areas for provisioning ecosystem services that are incompatible with conservation (exploitation zones). They compiled biodiversity data for birds (as well-studied representatives of faunal biodiversity) and diversity of habitats, at a 1 km cell scale (Figure 19.42). They grouped ecosystem service data for Catalonia into three classes: (i) ecosystem services considered compatible with biodiversity conservation (avoidance of soil erosion, carbon storage, water retention and recreation opportunity); (ii) ecosystem services considered to have a low impact on biodiversity conservation (pollination potential, mushroom production (an important cultural activity in Catalonia with about 1.2 million mushroom pickers) and potential areas for extensive (rather than intensive) grazing and hunting potential); (iii) ecosystem services considered non-compatible with conservation (biomass production and food crop production). Urban areas were excluded from the exercise.

The first step in the analysis is arbitrary in quantitative detail but based on expert opinion. Thus, targets were set that locked in all grid cells currently designated as reserves, and aimed to achieve 75% coverage of the current distribution of the rarest birds and habitats (and 25% or 10% for birds or habitats that are currently widely or very widely distributed), and for each ecosystem service they aimed to achieve in their zoning plan 20% of what is currently the case in the area as a whole. Those parts of the territory surplus to that necessary to achieve the targets were designated no-management zones.

The details of the process of optimising the spatial allocation of management zones need not concern us here (see Lanzas *et al.* (2019), for further explanation), but the outcome, whenever possible, was to allocate conservation and green infrastructure zones in areas with low human impact, to allocate exploitation zones in areas of high human impact, to ensure that the green infrastructure zones buffer the conservation zones, and to make the exploitation zones as spatially disconnected from the conservation zones as possible. They also investigated the consequences of different levels of strictness in how zones were allocated: at one extreme, where conservation zones were exclusively devoted to conservation (called 'strict'), at the other, where a more flexible strategy was employed with an increased co-occurrence of conservation and ecosystem service objectives allowed within a zone ('flexible').

The results are shown in Figure 19.42. The areas of highest conservation priority for threatened bird species were along the coast and in river networks, reflecting the dependence of some of the species on aquatic habitats, while the highest concentrations of priority habitats were in the north and north-east. Exploitation zones were concentrated in the central depression, reflecting high values there of the ecosystem service 'food crop production’. These areas also had high values for the ecosystem services ‘water retention’ and ‘hunting potential’. In the Pyrenean mountains were high values for ecosystem services related to forest ecosystems – ‘biomass production’, ‘carbon retention’, ‘water retention’ and ‘avoidance of soil erosion’ – and, in natural mountain pastures, high values for the ecosystem service ‘extensive grazing potential’.

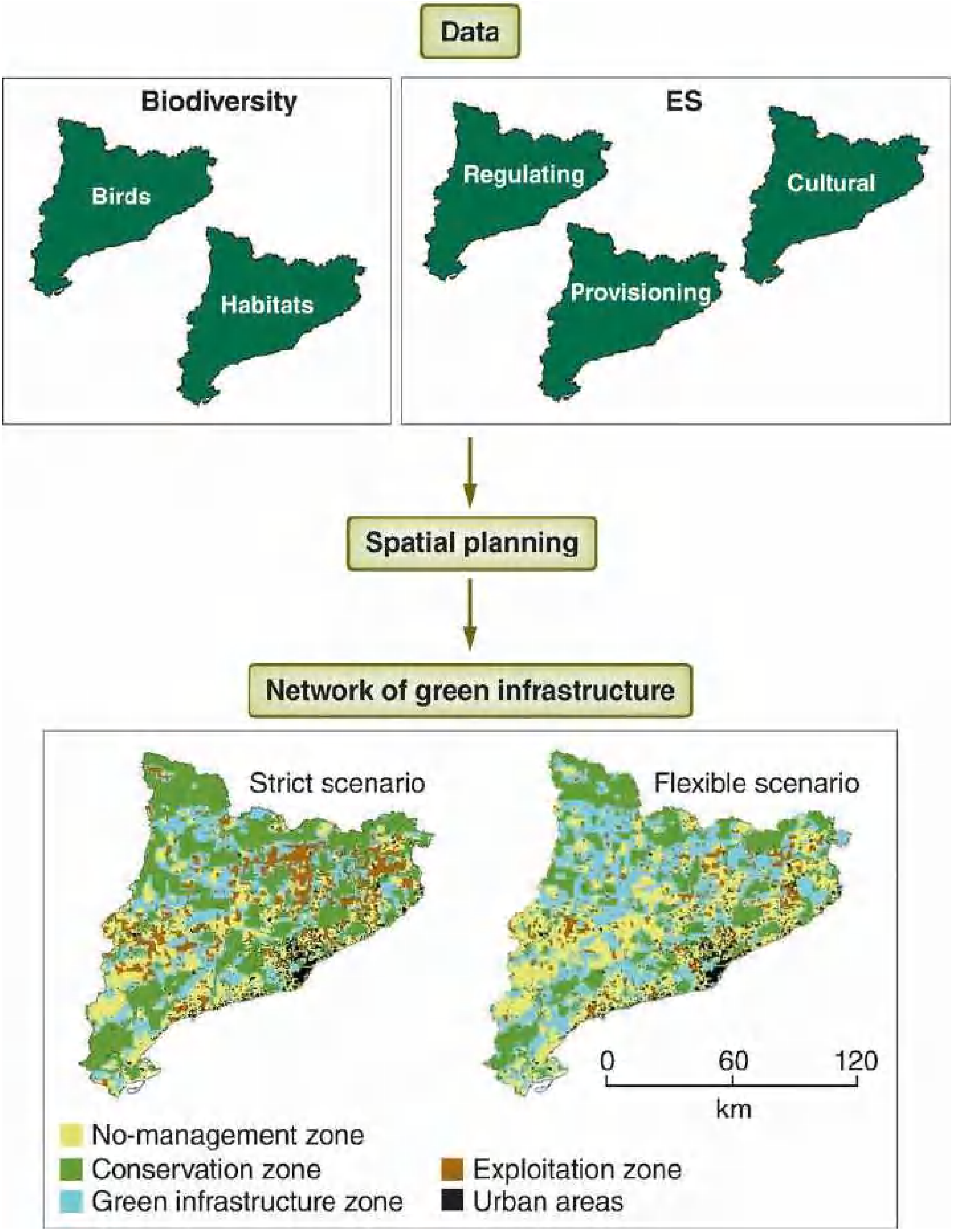

Data
Biodiversity
Birds
Habitats
ES
Regulating
Provisioning
Cultural
Spatial planning
Network of green infrastructure
Strict scenario
Flexible scenario
0
60
120
km
No-management zone
Conservation zone
Green infrastructure zone
Exploitation zone
Urban areas

Figure 19.42 Spatial planning for a network of green infrastructure in the Catalonia region of Spain. Biodiversity data consisted of bird and habitat diversity at a 1 km^2 scale, while ecosystem service (ES) data included regulating, provisioning and cultural services. The results of the analysis show the optimal distribution of conservation zones, green infrastructure zones and exploitation zones for two scenarios (strict and flexible; see text). No-management zones comprise those parts of the territory surplus to what is necessary to achieve the set targets. Note that the Mediterranean coast runs from north-east to south-west, while Catalonia is bordered by land along its north and west boundaries. The Pyrenean mountains are in the north.

Source: From Lanzas *et al.* (2019).

The set representation targets were met by both strict and flexible allocation strategies, but the flexible approach (where conservation and some ecosystem services were allowed to co-occur) led to a lower overall extent of exploitation zones and an increase in conservation and green infrastructure zones (Figure 19.42). The green infrastructure zones not only played a role in buffering and connecting conservation zones, they also contributed to achieving many ecosystem service targets (mushroom production, pollination potential, extensive grazing and hunting potential). The flexible allocation strategy saw an increase to 15% in area devoted to green infrastructure zones from 10% in the strict strategy. Lanzas *et al.* (2019) argue that their approach can help unlock new conservation priorities (enhancing the role of existing reserves by adding the value of ecosystem services they provide) and help to minimise land-use conflicts in landscapes with widespread human usage, while meeting the European Union mandate to identify green infrastructure networks.

Chapter 20

The Flux of Energy through Ecosystems

20.1 Introduction

All biological entities require matter for their construction and energy for their activities. This is true for individual organisms, but also for the populations and communities that they form in nature. We use the term *ecosystem* to denote the biological community *plus* the physicochemical environment that provides the living conditions for it and acts both as a source and a sink for energy and matter. In Chapters, we saw how community ecology concerns itself with patterns in the taxonomic composition and diversity of communities. We might expect the ecosystem concept to incorporate the community concept, but, in practice, ecosystem ecologists have focused on the flows of energy and matter into, within and out of a spatially defined entity (the ecosystem), more or less irrespective of the species that are present. Ecosystem ecology and community ecology have thus developed as rather distinct disciplines, which have only recently begun to meld together (Loreau, 2010b). In what follows, therefore, we will highlight, when we can, examples that successfully bring the two viewpoints together (e.g. Sections 20.2.4, 20.3.6 and 20.5.1). Overall, however, our emphasis will be on the intrinsic importance of fluxes themselves – of energy (this chapter) and of matter (Chapter 21).

Lindemann laid the foundations of ecological energetics

A classic paper by Lindemann (1942) laid the foundations of a science of ecological energetics. He attempted to quantify the concept of food chains by considering the efficiency of energy transfer between trophic levels – from incident radiation received by a community, through its capture by green plants in photosynthesis, to its subsequent use by herbivores, carnivores and decomposers. Lindemann's paper was a major catalyst for the International Biological Programme (IBP; 1964–74), which, with a view to human welfare, aimed to understand the biological basis of the productivity of areas of land, freshwaters and the seas (Worthington, 1975). The IBP provided the first occasion on which ecologists throughout the world were challenged to work together towards the common end of enhancing scientific knowledge.

In 1988, a further pressing issue again galvanised the scientific community into action, this time to synthesise scientific information as a basis for policy formulation. Realisation was dawning that the burning of fossil fuels, deforestation and other pervasive human influences were causing dramatic changes to global climate and atmospheric composition, which in turn would influence patterns of productivity on a global scale. The Intergovernmental Panel on Climate Change (IPCC) was set up to provide policymakers with regular assessment reports of the scientific basis of climate change, its impacts and future risks (IPCC, 2014). Recognising the extent to which humans have changed ecosystems – not just by altering the climate but also through habitat destruction and modification, overexploitation and pollution – another major collaborative effort began in 2001, the Millennium Ecosystem Assessment (MEA, 2005a). The MEA was concerned with the consequences of ecosystem change for human well-being and the actions needed for conservation and sustainable use of ecosystems. Its particular focus has been on the services that ecosystems provide, such as food, forest products, clean water and flood control. In the context of the present chapter, much of the recent work on productivity (the rate of production of biomass) has a prime objective of providing the basis for predicting the effects of changes in climate, atmospheric composition and land use on terrestrial and aquatic ecosystems and biodiversity conservation (a theme taken up again in Chapter 22).

APPLICATION 20.1 Ecosystem services

linking ecosystem and community ecology

We have already introduced the idea of a utilitarian view of ecosystem processes that focuses on the services that ecosystems provide for people to use and enjoy (supporting, provisioning, regulating and cultural ecosystem services – see Section 15.4.1). The concept of ecosystem services provides an important link between ecosystem ecology and community ecology because ecosystem productivity supports the biodiversity that is so important to human welfare and happiness. The ecosystem service concept also melds together ecological, economic and sociological concerns: a key feature of ecosystem services is that values can be assigned to them so that the dollar gains associated with particular enterprises at a location can be assessed against the dollar losses associated with lost or damaged ecosystem services.

Some ecologists have argued that the concept of ecosystem services provides a very anthropocentric view of nature; that nature does not exist to ‘service’ us and has its own intrinsic value that may sometimes be undermined by dealing solely in monetary terms (Bekessy *et al.*, 2018). In other words, policies based on the valuation of ecosystem services may not always protect biodiversity and achieve conservation goals. Others maintain that the concept of ecosystem services is not one in which *Homo sapiens* is the only species that matters but rather one where the whole system is the focus – of humans embedded in society and within the rest of nature (Costanza *et al.*, 2017). At any rate, ecosystem services have become the focus of a burgeoning research effort and the concept is playing an increasing role in formulating policy directed at the sustainable use of resources.

progressive improvements in technology to assess productivity

The decades since Lindemann’s classic work have seen progressive improvements in technology to assess productivity. Early calculations in terrestrial ecosystems involved sequential measurements of plant biomass (usually just the above-ground parts) and estimates of energy transfer efficiency between trophic levels. In aquatic ecosystems, production estimates relied on photosynthesis-related changes in the concentrations of oxygen or carbon dioxide measured in experimental enclosures. More recently, meteorological flux towers (more than 600 around the world) have been used in a wide range of terrestrial habitats to measure the uptake or emission of carbon dioxide and provide on-the-ground production estimates. Now, with the development of satellite remote-sensing techniques, local estimates of productivity can be extrapolated to regional and global scales. Thus, satellite sensors can measure vegetation cover on land and chlorophyll concentrations in the sea, from which rates of light absorption are calculated and, based on mathematical models that incorporate our understanding of photosynthesis, are converted to estimates of productivity (e.g. Cullen *et al*., 2012 ; Masek *et al*., 2015).

20.1.1 The fundamentals of energy flux

standing crop and biomass, ...

Before proceeding further it is necessary to define some new terms. The bodies of the living organisms within a unit area constitute a *standing crop* of biomass. By *biomass* we mean the mass of organisms per unit area of ground (or per unit surface area or unit volume of water) and this is usually expressed in units of energy (e.g. J m^{-2}) or dry organic matter (e.g. t ha^{-1}) or carbon (e.g. g C m^{-2}). The great bulk of the biomass in terrestrial communities is almost always formed by plants, and in many aquatic communities by algae, which are the primary producers of biomass because of their almost unique ability to fix carbon in photosynthesis. (We have to say 'almost unique' because photosynthesis and chemosynthesis by bacteria and archaea also contribute to forming new biomass.) Biomass includes the whole bodies of the organisms even though parts of them may be dead. This needs to be borne in mind, particularly when considering woodland and forest communities in which the bulk of the biomass is dead heartwood and bark. The living fraction of biomass represents active capital capable of generating interest in the form of new growth, whereas the dead fraction is incapable of new growth. In practice, we include in biomass all those parts, living or dead, which are attached to the living organism. They cease to be living biomass when they fall off and become litter, humus or peat (*dead organic matter*).

... primary and secondary productivity, autotrophic respiration, ...

The *primary productivity* of a community is the rate at which biomass is produced per unit area by the primary producers, the *autotrophs* (plants, algae and photosynthetic or chemosynthetic bacteria and archaea). It can be expressed either in units of energy (e.g. J m^{-2} day^{-1}) or dry organic matter (e.g. kg ha^{-1} $year^{-1}$) or carbon (e.g. g C m^{-2} $year^{-1}$). The total fixation of energy by photosynthesis is referred to as *gross primary productivity* (GPP). A proportion of this is respired away by the autotrophs and is lost from the community as respiratory heat (*autotrophic respiration* or RA). The difference between GPP and RA is known as *net primary productivity* (NPP) and represents the actual rate of production of new biomass that is available for consumption by heterotrophic organisms.

... net ecosystem productivity, and heterotrophic and ecosystem respiration

The rate of production of biomass by heterotrophs is called *secondary productivity*. Another way to view energy flux in ecosystems involves the concept of *net ecosystem productivity* (NEP, using the same units as GPP or NPP). This acknowledges that the carbon fixed in GPP can leave the system as inorganic carbon (usually carbon dioxide) via either RA or, after consumption by heterotrophs, via *heterotrophic respiration* (RH) – the latter consisting of respiration by non-autotrophic bacteria and archaea, fungi, non-photosynthetic eukaryotes (formerly referred to as protozoans) and animals. Total *ecosystem respiration* (RE) is the sum of RA and RH. NEP then is equal to GPP – RE. When GPP exceeds RE, the ecosystem is fixing carbon faster than it is being released and thus acts as a carbon sink. When RE exceeds GPP, carbon is being released faster than it is fixed and the ecosystem is a net carbon source. That the rate of RE can exceed GPP may seem paradoxical. However, it is important to note that an ecosystem can receive organic matter from sources other than its own photosynthesis – via the import of dead organic matter that has been produced elsewhere. Organic matter produced by photosynthesis within an ecosystem's boundaries is known as *autochthonous*, whereas that imported from elsewhere is called *allochthonous*.

In what follows, we deal first with large-scale patterns in biomass and primary productivity (Section 20.2) before considering the factors that limit productivity in terrestrial (Section 20.3) and aquatic (Section 20.4) settings. We then turn to the fate of primary productivity and consider the flux of energy through food webs (Section 20.5), placing particular emphasis on the relative importance of grazer and decomposer systems (a more detailed consideration of the population interactions in food webs was discussed in Chapter 17). We finally turn to seasonal and longer term variations in energy flux through ecosystems.

20.2 Patterns in primary productivity

Net primary production of the planet is estimated to be about 109 petagrams of carbon per year (1 Pg = 10^{15} g = 1 billion tonnes). Global estimates of terrestrial NPP average about 58.5 Pg C year^{-1} and aquatic NPP is of a similar order at 50.6 Pg C year^{-1} (Table 20.1). GPP is approximately double NPP in both terrestrial and aquatic ecosystems.

Table 20.1 Net primary production (NPP) per year for major biomes and for the planet in total (in units of petagrams of carbon). *Source*: From (1) Geider *et al.* (2001), (2) Gough (2011) and (3) Rousseaux & Gregg (2014).

Marine	**NPP (Pg C yr^{-1})**	**Terrestrial^{2}**	**NPP (Pg C yr^{-1})**
Tropical and subtropical oceans^{3}	13.1	Tropical forest	16.0–23.1
Temperate oceans^{3}	18.0	Temperate forest	4.6–9.1
Polar oceans^{3}	6.9	Boreal forest	2.6–4.6
Coastal^{1}	10.7	Tropical savanna and grasslands	14.9–19.2
Salt marsh/estuaries/seaweed^{1}	1.2	Temperate grasslands and shrublands	3.4–7.0
Coral reefs^{1}	0.7	Deserts	0.5–3.5
		Tundra	0.5–1.0
		Croplands	4.1–8.0
Total	50.6	Total	48.0–69.0

> primary productivity depends on, but is not solely determined by, solar radiation

NPP is far from evenly distributed across the globe (Figure 20.1). Although oceans cover about two-thirds of the world's surface, they account for less than half of its production. On the land, tropical rainforests and savannas account between them for about 60% of terrestrial NPP, reflecting the large areas covered by these biomes and their high levels of productivity. Oceans account for three-quarters of aquatic NPP, again reflecting the huge areas they cover, but the open ocean is generally only a quarter as productive per unit area as coastal and estuarine systems (Chavez *et al.*, 2011). All biological activity is ultimately dependent on received solar radiation but solar radiation alone does not determine primary productivity. In very broad terms, the fit between solar radiation and productivity is far from perfect because incident radiation can be captured efficiently only when water and nutrients are available and when temperatures are in the range suitable for plant growth. Many areas of land receive abundant radiation but lack adequate water, and most areas of the oceans are deficient in mineral nutrients.

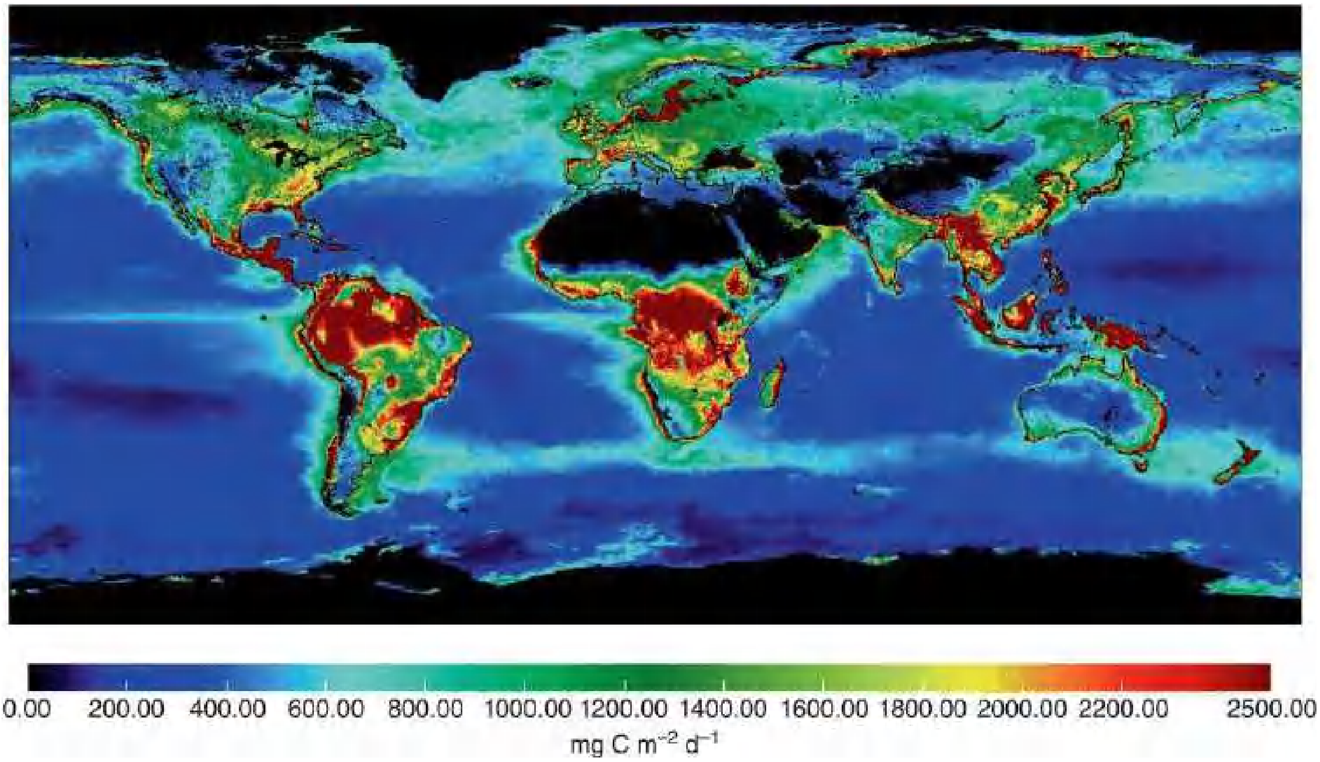


Figure 20.1 **Annual average rates of net primary productivity** across the planet for both oceans and terrestrial ecosystems.

Source: Data for 2006 from the Ocean Productivity Site at http://www.science.oregonstate.edu/ocean.productivity (last accessed October 2019).

APPLICATION 20.2 Human appropriation of net primary production

Humans appropriate nearly a quarter of the world's total terrestrial NPP for their own use, and high-density industrial nations appropriate nearly 50% of terrestrial NPP within their national boundaries. Human appropriation of net primary production (HANPP), as defined by Haberl *et al.* (2014), includes not only harvested HANPP (HANPPharv – whether *used* to feed people or domestic animals, or *unused*, such as residues left in the field or human-induced fires) but also HANPP associated with land-use change (HANPPluc), measured as the difference between the potential NPP of natural vegetation that would exist in the absence of human land use (NPPpot) and the actual NPP in the human-modified ecosystem (NPPeco). NPPpot is usually greater than NPPeco (Figure 20.2a).

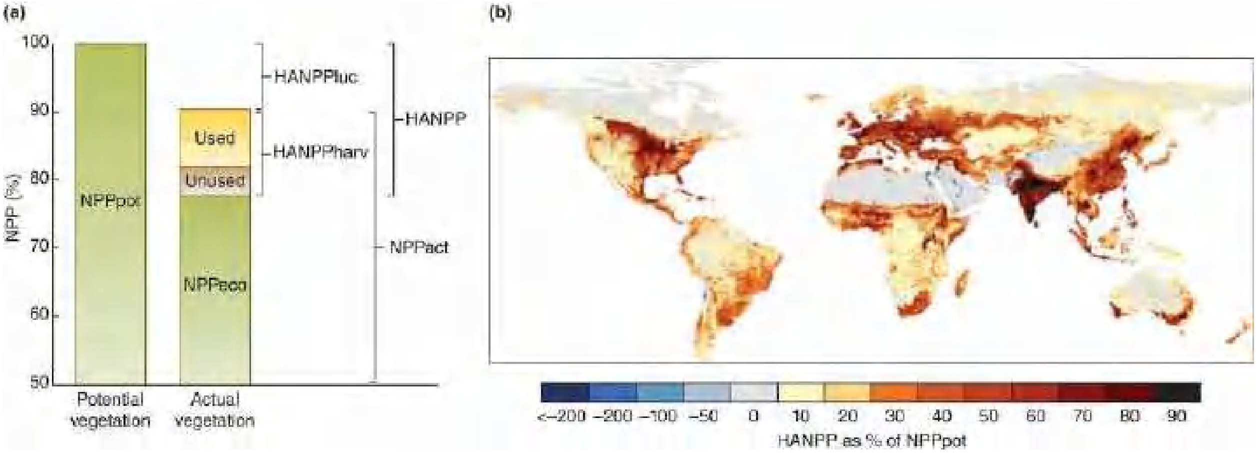


Figure 20.2 Patterns in the human appropriation of net primary production (HANPP). (a) The calculation of HANPP illustrated using global estimates for the year 2000. See text for details. (b) HANPP as a percentage of NPP of natural vegetation that would exist in the absence of human land use (NPPpot). Note that negative values occur in regions where actual NPP in the human-modified ecosystems (NPPeco) exceeds NPPpot because of high inputs of fertilisers and particularly irrigation in arid regions.

Source: From Haberl *et al.* (2014).

Globally, annual HANPP increased from 6.9 to 14.8 Pg carbon from 1910 to 2005, equivalent to an increase from 13% to 25% of NPPpot (Haberl *et al.*, 2014). The proportion of NPPpot used is particularly high in Europe, India, mid-west and eastern USA, South-east Asia, eastern China and the seasonal tropical forests of western Africa.

Human use of NPP can be viewed as more efficient if HANPPluc is minimal and HANPP more or less equals HANPPharv. Wise decisions about which crops and forestry species to harvest, together with the use of high-intensity technologies, may enable increases in harvest output without necessarily increasing HANPP overall. However, on the downside, the spread of industrial agriculture and forestry can be expected to diminish other ecosystem services important to human welfare (see Section 20.1). Haberl *et al.* (2014) also argue that contemplation of a possible switch from fossil fuels to biomass fuels needs to be approached with caution because of an accompanying massive increase in pressure on ecosystems. So we can see that careful land-management choices can help minimise human impacts on NPPeco and to maximise land-use efficiency, while also accounting for the value of ecosystem services potentially lost in the process.

20.2.1 Latitudinal trends in productivity

productivity of terrestrial ecosystems diminishes towards the poles ...

A general pattern of declining terrestrial productivity from the equator to the poles is evident from Figure 20.1. Of course, there is also considerable variation around this trend, much of it due to differences in water availability, local topography and associated variations in microclimate (referred to previously in Section 19.6.1). Particularly extensive data are available for forests (Figure 20.3a), which show an obvious productivity decline from temperate to boreal forests, but a less marked decline between tropical and temperate forests.

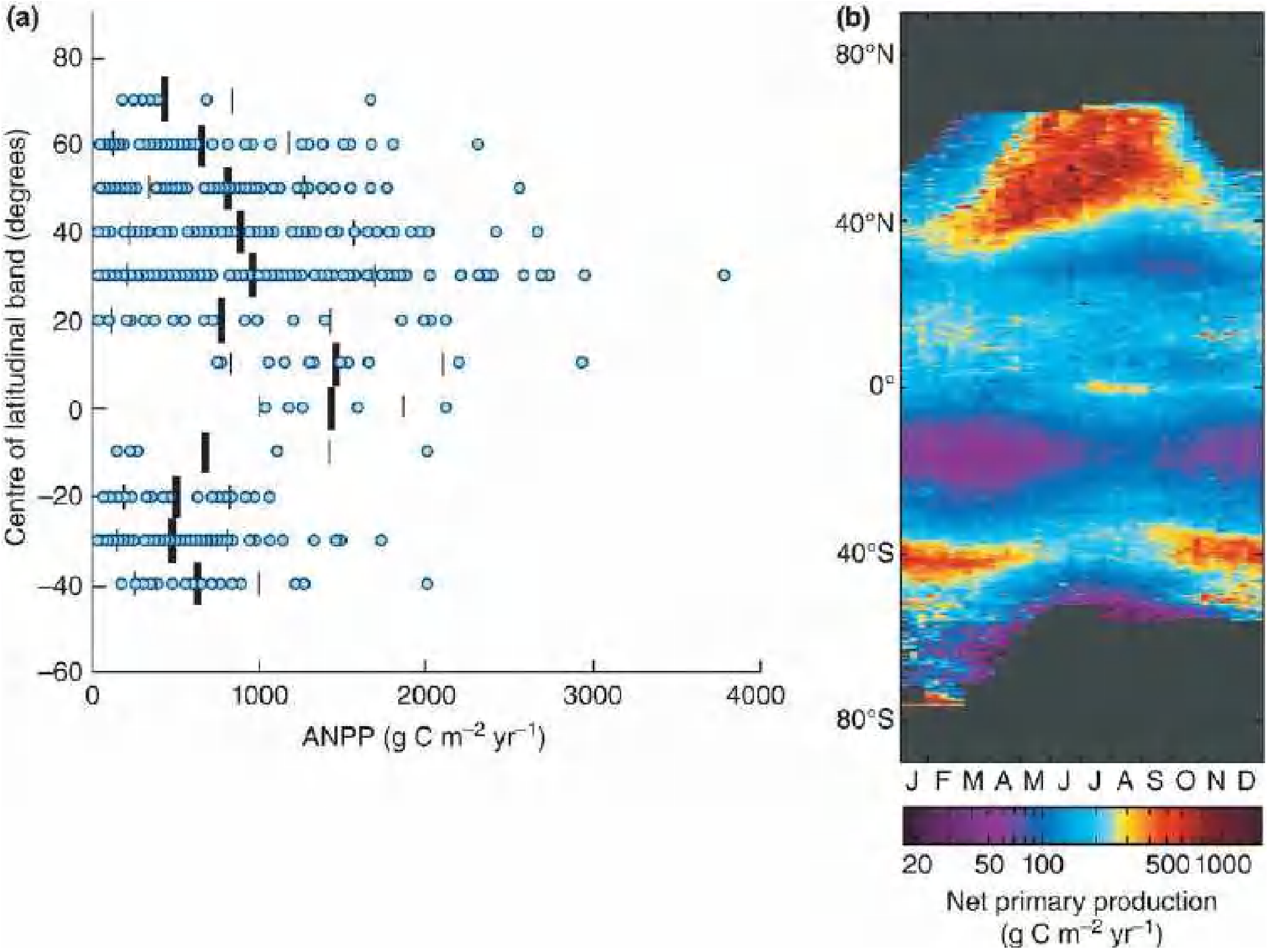


Figure 20.3 Global patterns in forest and ocean net primary production (NPP). (a) Global distribution of estimated annual above-ground NPP (ANPP) for 755 forest sites in relation to latitude. All data points are shown with mean (thick bar) ± SD (thin vertical bars). (b) Latitudinal pattern of satellite-based estimates of annual NPP in the Atlantic Ocean. Dark grey areas are landmasses or permanent ice cover.

Source: (a) After Huston & Wolverton (2009). (b) From Lutz *et al.* (2007).

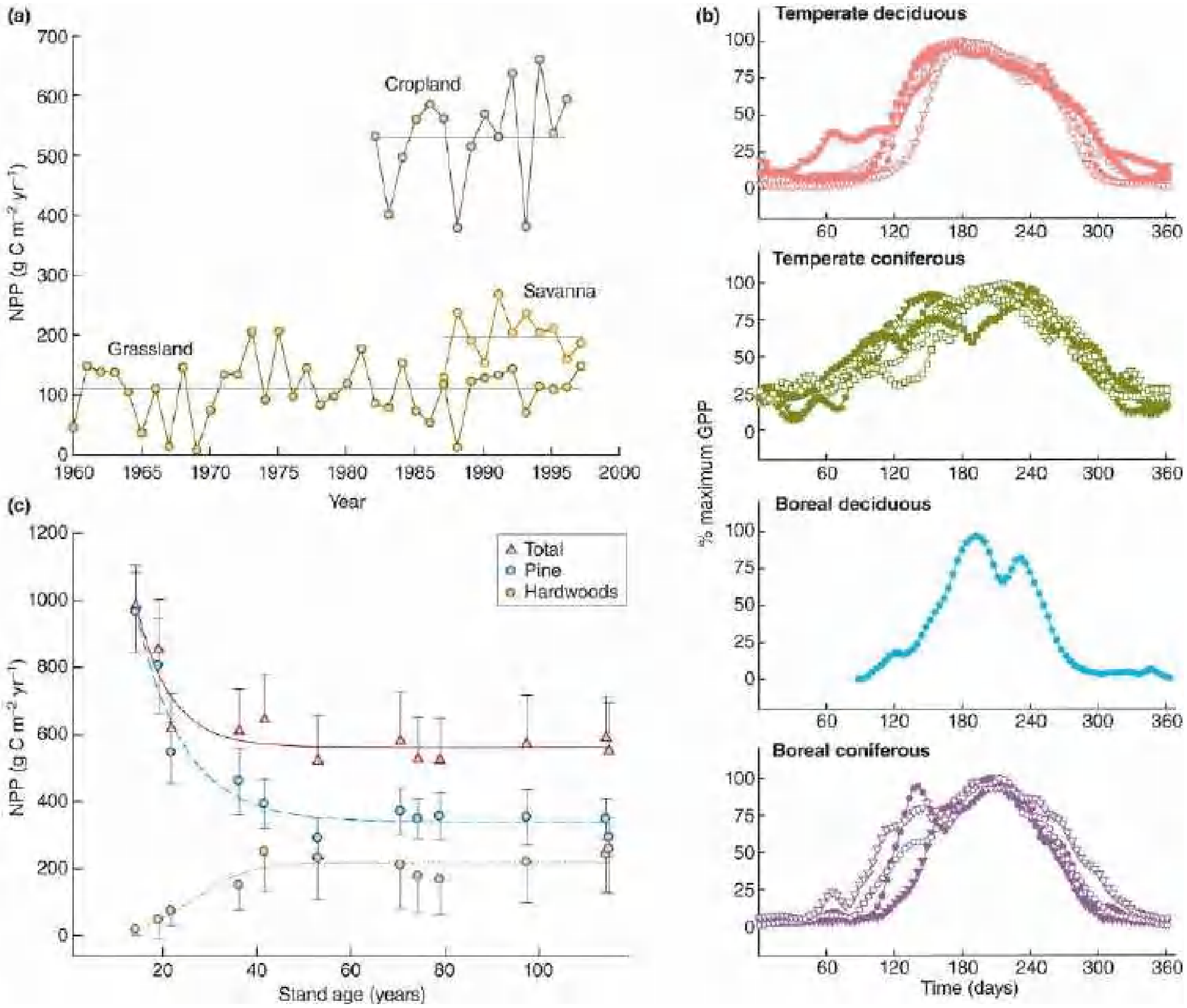
... but ocean productivity shows the opposite pattern

Phytoplankton production in the oceans provides an interesting contrast; NPP is higher in northern and southern seas than nearer the equator (Figure 20.3b). Ocean productivity is often limited by a shortage of nutrients. Very high productivity occurs in marine communities where nutrients are supplied in upwellings of nutrient-rich waters (especially in the Southern Ocean) or from adjacent continental landmasses (especially in the northern oceans), despite less incident radiation and lower temperatures in these regions. Thus, while radiation (a resource) and temperature (a condition) may often limit the productivity of communities, other factors frequently constrain productivity within even narrower limits.

20.2.2 Temporal trends in primary productivity

> productivity shows considerable temporal variation: year to year, ...

The large ranges in productivity in Table 20.1 and the wide confidence intervals in Figure 20.3a emphasise the considerable variation that exists within a given class of ecosystems. It is important to note also that productivity varies from year to year in a single location. This is illustrated for a temperate cropland, a tropical grassland and a tropical savanna in Figure 20.4a. Such annual fluctuations no doubt reflect year-to-year variation in cloudless days, temperature and rainfall.



[Figure 20.4](#) Temporal patterns in primary production. (a) Interannual variation in net primary productivity (NPP) in a grassland in Queensland, Australia (above-ground NPP), a cropland in Iowa, USA (total above- and below-ground NPP) and a tropical savanna in Senegal (above-ground NPP). Black horizontal lines show the mean NPP for the whole study period. (b) Seasonal development of maximum daily gross primary productivity (GPP) for deciduous and coniferous forests in temperate (Europe and North America) and boreal (Canada, Scandinavia and Iceland) locations. The different symbols in each panel relate to different forests. Daily GPP is expressed as the percentage of the maximum achieved in each forest during 365 days of the year. (c) NPP (mean and SD) of loblolly pines, hardwood trees and in total for 12 forest stands that vary in age.

Source: (a) After Zheng *et al*. (2003). (b) After Falge *et al*. (2002). (c) From Drake *et al*. (2011).

... seasonally, ...

At a smaller temporal scale, productivity reflects seasonal variations in conditions, particularly in relation to the consequences of temperature or water availability for the length of the growing season. For example, the period when daily GPP is high persists for longer in temperate than in boreal (more polar) situations (Figure 20.4b). Moreover, the growing season is more extended but the amplitude of seasonal change is smaller in evergreen coniferous forests than in their deciduous counterparts (where the growing season is curtailed by the autumn shedding of leaves).

… during ecological succession …

Over a longer time period, during ecological succession, while the general pattern certainly is that forest NPP (and GPP) increases with stand age, different trees may show contrasting patterns. Thus, in a study of different-aged forests dominated by loblolly pine (*Pinus taeda*) in North Carolina, USA, the pines showed a decline in NPP with forest age while later successional hardwood trees (including sweetgum *Liquidambar styraciflua* and oaks *Quercus* spp.) growing alongside the pines showed an increase in NPP (Figure 20.4c). The decline in NPP with pine age was linked to increasing hydraulic limitation (greater stomatal closure because the taller trees have further to transport water in the face of increasing gravitational potential opposing its ascent), while the increase in hardwoods was associated with a higher leaf area index (leaf area per unit surface area of ground) in older trees.

… and in relation to climate change

At a larger temporal scale, the increasing concentration of atmospheric carbon dioxide (CO_2) and associated rise in surface temperatures, due in particular to the burning of fossil fuels, might be expected to have increased photosynthesis and global primary production in recent decades. In this context, models have been used to scale up GPP data from flux tower stations to predict global terrestrial GPP based on worldwide data for the variables that account for GPP at the tower station sites. The models vary in their details, but Anav *et al.* (2015) report that estimates ('predictions') of the trend in global GPP for the period 1990–2009 were all positive, with increases ranging from 0.05 to 0.621 Pg C year^{-1} (i.e. an increase in annual GPP of between 0.05 and 0.621 Pg C each year). We clearly need to improve both observation-based datasets and our carbon cycling models to better understand the future uptake of CO_2 by the earth's vegetation.

20.2.3 Autochthonous and allochthonous production

autochthonous and allochthonous production …

All biotic communities depend on a supply of energy for their activities. In most terrestrial systems this is autochthonous production, contributed *in situ* by the photosynthesis of green plants. Exceptions exist, however, particularly where colonial animals deposit faeces derived from food consumed at a distance from the colony (e.g. bat colonies in caves, seabirds on coastland) – guano is an example of allochthonous organic matter (dead organic material formed outside the ecosystem).

… vary in systematic ways in lakes, rivers and estuaries

In aquatic communities, the autochthonous input is provided by the photosynthesis of large plants and attached benthic algae in shallow waters (littoral zone) and by microscopic phytoplankton in the open water. However, a substantial proportion of the organic matter in aquatic communities comes from allochthonous material that arrives in rivers, via groundwater or is blown in by the wind. The relative importance of the two autochthonous sources (littoral and planktonic) and the allochthonous source of organic material in an aquatic system depends on the dimensions of the body of water and the types of terrestrial community that deposit organic material into it.

A small stream running through a wooded catchment derives most of its energy input from litter shed by surrounding vegetation. Shading from the trees prevents any significant growth of planktonic or attached algae or aquatic higher plants. As the stream widens further downstream, shading by trees is restricted to the margins and autochthonous primary production increases. Still further downstream, in deeper and more turbid waters, rooted higher plants contribute much less, and the role of the microscopic phytoplankton becomes more important. Where large river channels are characterised by a flood plain, with associated oxbow lakes, swamps and marshes, allochthonous dissolved and particulate organic matter may be carried to the river channel from its flood plain during episodes of flooding.

The sequence from small, shallow lakes to large, deep ones shares some of the characteristics of the river continuum just discussed (Figure 20.5). A small lake is likely to derive quite a large proportion of its energy from the land because its periphery is large in relation to its area. Small lakes are also usually shallow, so internal littoral production is more important than that by phytoplankton. In contrast, a large, deep lake will derive only limited organic matter from outside (a small periphery relative to lake surface area) and littoral production, limited to the shallow margins, may also be low. The organic inputs to the community may then be due almost entirely to photosynthesis by the phytoplankton.

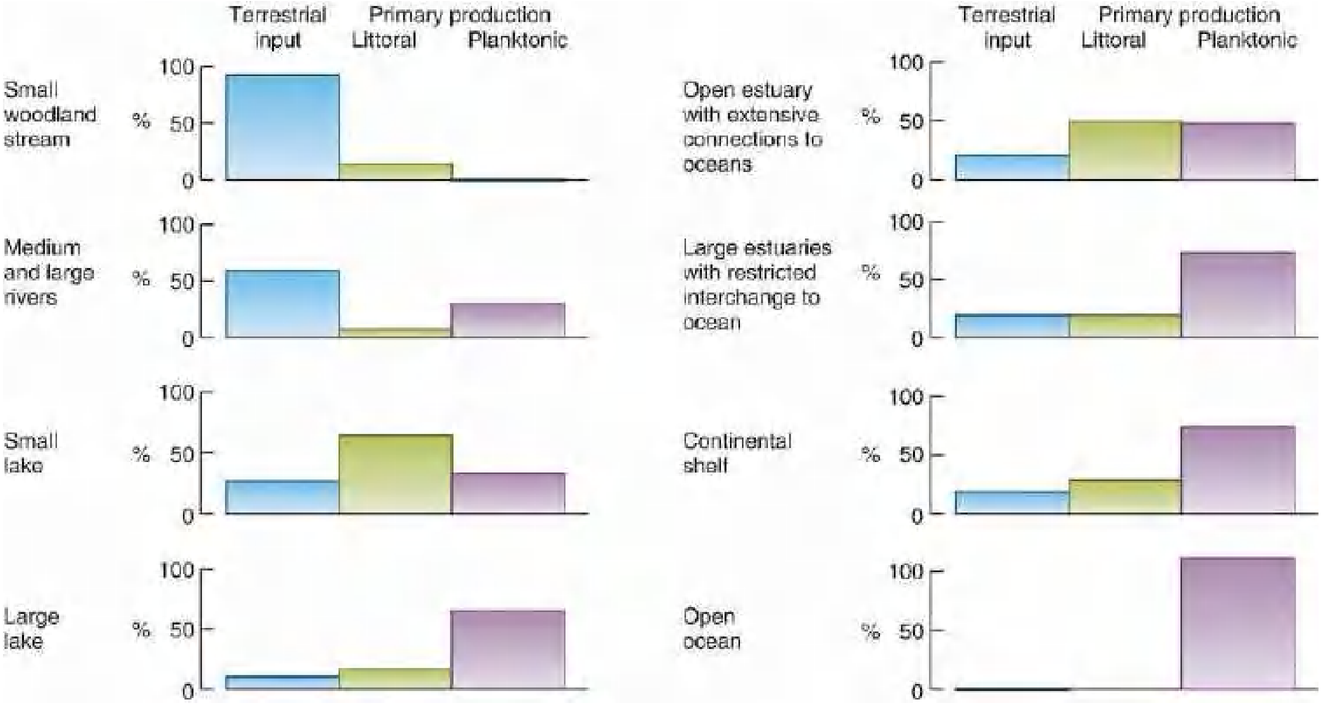


Figure 20.5 **Contrasting sources of organic matter in aquatic communities.** Variation in the importance of terrestrial input of organic matter and littoral and planktonic primary production in contrasting aquatic communities.

Estuaries are often highly productive systems, receiving allochthonous material and a rich supply of nutrients from the rivers that feed them. In some situations, there is distinct temporal variation in the relative importance of carbon sources. Thus, where there are marked wet and dry seasons, such as in the intermittently connected estuaries of tropical Australia, a flood-related influx of allochthonous material of distant terrestrial origin occurs in the post-wet season, whereas autochthonous (phytoplankton and benthic algae) and local sources (riparian vegetation) predominate in the dry season (Abrantes & Sheaves, 2010). In larger estuarine basins, with restricted interchange with the open ocean and with small marsh peripheries relative to basin area, phytoplankton tend to dominate. By contrast, seaweeds dominate in some open basins with extensive connections to the sea. In turn, continental shelf communities derive a proportion of their energy from terrestrial sources (particularly via estuaries) and their shallowness often provides for seaweed (macroalgae) or sea grass (angiosperm) communities that are among the most productive systems of all. The high levels of GPP in sea grass meadows, for example, is contributed almost equally by the sea grasses themselves (which are angiosperms) and the epiphytes and macroalgae growing on and around them (Duarte *et al*., 2010).

Finally, the open ocean can be described in one sense as the largest, deepest ‘lake’ of all. Except around the mouths of large rivers, the input of organic material from terrestrial communities is negligible, and the great depth precludes photosynthesis in the darkness of the sea bed. The phytoplankton are then all-important as primary producers.

20.2.4 Variations in the relationship of productivity to biomass

NPP : B ratios are very low in forests and very high in aquatic communities

We can relate the productivity of a community to the standing crop biomass that produces it (the interest rate on the capital). Alternatively, we can think of the standing crop as the biomass that is sustained by the productivity (the capital resource that is sustained by earnings). Overall, there is a dramatic difference in the total biomass that exists on land (800 Pg) compared with the oceans (2 Pg) and freshwater (<0.1 Pg) (Geider *et al*., 2001). On an areal basis, biomass on land ranges from 0.2 to 200 kg m^{-2}, in the oceans from less than 0.001 to 6 kg m^{-2} and biomass in freshwater is generally less than 0.1 kg m^{-2} (Geider *et al*., 2001).

The average values of NPP and standing crop biomass (B) for a range of community types are plotted against each other in Figure 20.6. It is evident that a given value of NPP is produced by a smaller biomass when non-forest terrestrial systems are compared with forests, and the biomass involved is smaller still when aquatic systems are considered. Thus NPP : B ratios (kilograms of dry matter (DM) produced per year per kilogram of standing crop) average 0.042 for forests, 0.29 for other terrestrial systems and 17 for aquatic communities. The major reason for this is almost certainly that a large proportion of forest biomass is dead and has been so for a long time (dead limbs, tree heartwood, etc.), and also that much of the living support tissue is not photosynthetic. In grassland and scrub, a greater proportion of the biomass is alive and involved in photosynthesis, though half or more of the biomass may be roots. In aquatic communities, particularly where productivity is due mainly to phytoplankton, there is no support tissue, there is no need for roots to absorb water and nutrients, dead cells do not accumulate (they are usually eaten before they die) and the photosynthetic output per kilogram of biomass is thus very high indeed. Another factor that helps to account for high NPP : B ratios in phytoplankton communities is the rapid turnover of biomass (turnover times of biomass in oceans and freshwaters average 0.02–0.06 years, compared with 1–20 years on land; Geider *et al*., 2001). The annual NPP shown in Figure 20.6 is actually produced by a number of overlapping phytoplankton generations, while the standing crop biomass is only the average present at an instant.

OO	Open ocean	SM	Swamp and marsh	WS	Woodland and scrubland
CS	Continental shelf	TRF	Tropical rainforest	S	Savanna
UW	Upwelling zone	TSF	Tropical seasonal forest	TG	Temperate grassland
ABR	Algal beds and reefs	TEF	Temperate evergreen forest	TA	Tundra and alpine
E	Estuaries	TDF	Temperate deciduous forest	DSD	Desert and semi-desert
FW	Freshwater lakes and streams	BF	Boreal forest	CL	Cultivated land

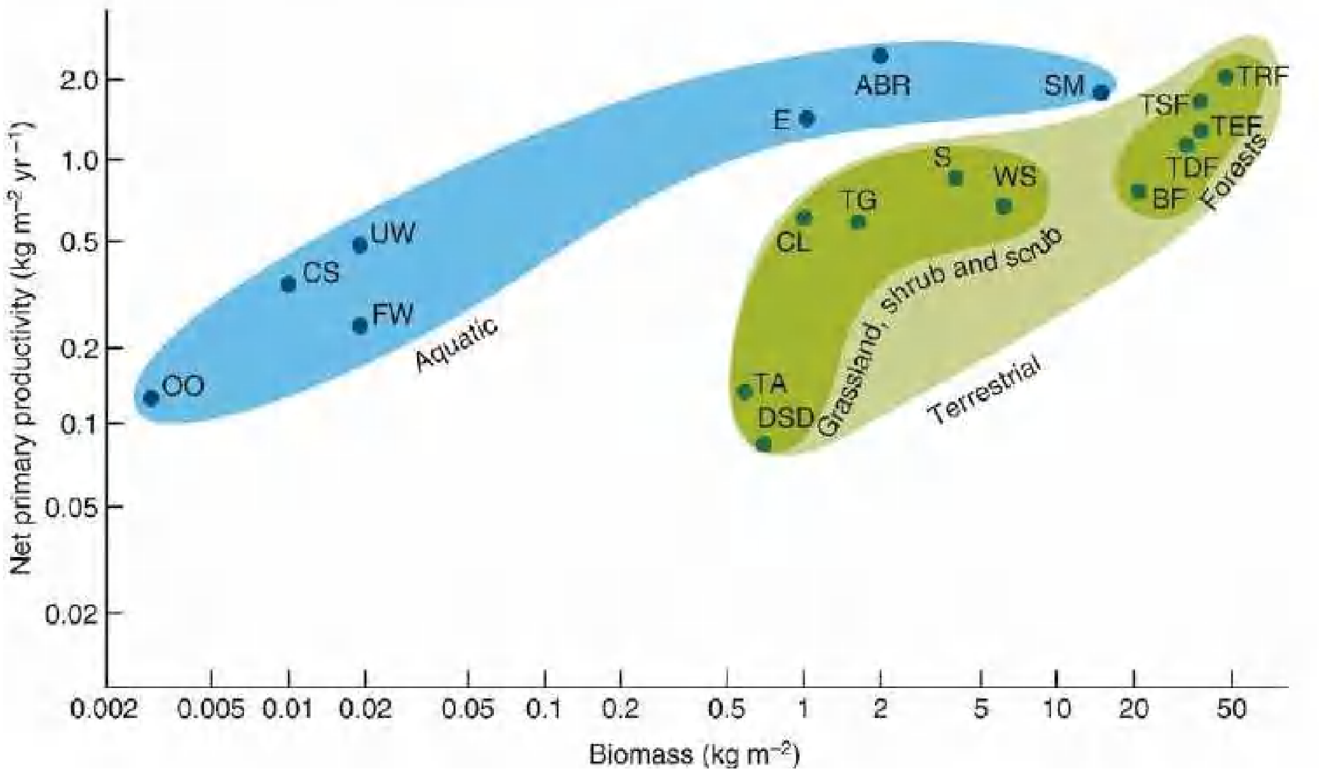


Figure 20.6 **NPP in relation to biomass**. The relationship between average net primary productivity and average standing crop biomass for a range of ecosystems.

Source: Based on data in Whittaker (1975).

NPP : B ratios provide a further link between community and ecosystem ecology

It is clear that community composition – whether dominated by phytoplankton or macrophytes, or in terrestrial settings by grasses or shrubs or trees – has a strong bearing on ecosystem energetics. This is also evident in ecological successions. Thus, the early successional pioneers are rapidly growing herbaceous species with relatively little support tissue and a high NPP : B ratio. However, the species that come to dominate later are generally slow growing, but eventually achieve a large size and come to monopolise the supply of space and light. Their structure involves considerable investment in non-photosynthesising and dead support tissues, and as a consequence their NPP : B ratio is low.

20.3 Factors limiting primary productivity in terrestrial communities

Sunlight, CO_2, water and soil nutrients are the resources required for primary production on land, while temperature, a condition, has a strong influence on the rate of photosynthesis. Atmospheric CO_2 has increased from 270 μmol mol^{-1} in preindustrial times to more than 400 μmol mol^{-1}

today. Experimental enrichment of CO_2 around enclosed plant communities usually leads to enhanced NPP, an effect that is particularly pronounced in C_3 species (14–31% rise in NPP; see Section 3.3.1) (Lenka & Lal, 2012). While an overall CO_2 fertilisation effect must has been quite substantial through recent decades with increasing atmospheric concentrations, CO_2 probably plays little role in determining differences between the productivities of different communities. On the other hand, the quality and quantity of light, the availability of water and nutrients, and temperature all vary dramatically from place to place. They are all candidates for the role of limiting factor. Which of them actually sets the limit to primary productivity?

20.3.1 Inefficient use of solar energy

terrestrial communities use radiation inefficiently

Depending on location, something between 0 and 5 joules (J) of solar energy strikes each square metre of the earth's surface every minute. If all this were converted by photosynthesis to plant biomass (that is, if photosynthetic efficiency were 100%) there would be a prodigious generation of plant material, one or two orders of magnitude greater than recorded values. However, much of this solar energy is unavailable for use by plants. In particular, only about 44% of incident shortwave radiation occurs at wavelengths suitable for photosynthesis. This photosynthetically active radiation (PAR) occurs in the 400–700 nm wavelength range. Even when this is taken into account, though, productivity still falls well below the maximum possible. Photosynthetic efficiency has two components – the fraction absorbed (f_{PAR}; a function of total foliage area per unit ground surface) and the plant's efficiency, ε, in converting this to biomass (with units such as gC MJ^{-1}); ε varies with species and environmental conditions (such as nutrient supply, water and temperature). Figure 20.7 shows the range in overall net photosynthetic efficiencies (% of PAR incorporated into above-ground NPP) in seven coniferous forests, seven deciduous forests and eight desert communities studied as part of the International Biological Programme (see Section 20.1). The conifer communities had the highest efficiencies, but these were only between 1% and 3%. For a similar level of incoming radiation, deciduous forests achieved 0.5–1%, and, despite their greater energy income, deserts were able to convert only 0.01–0.2% of PAR to biomass.

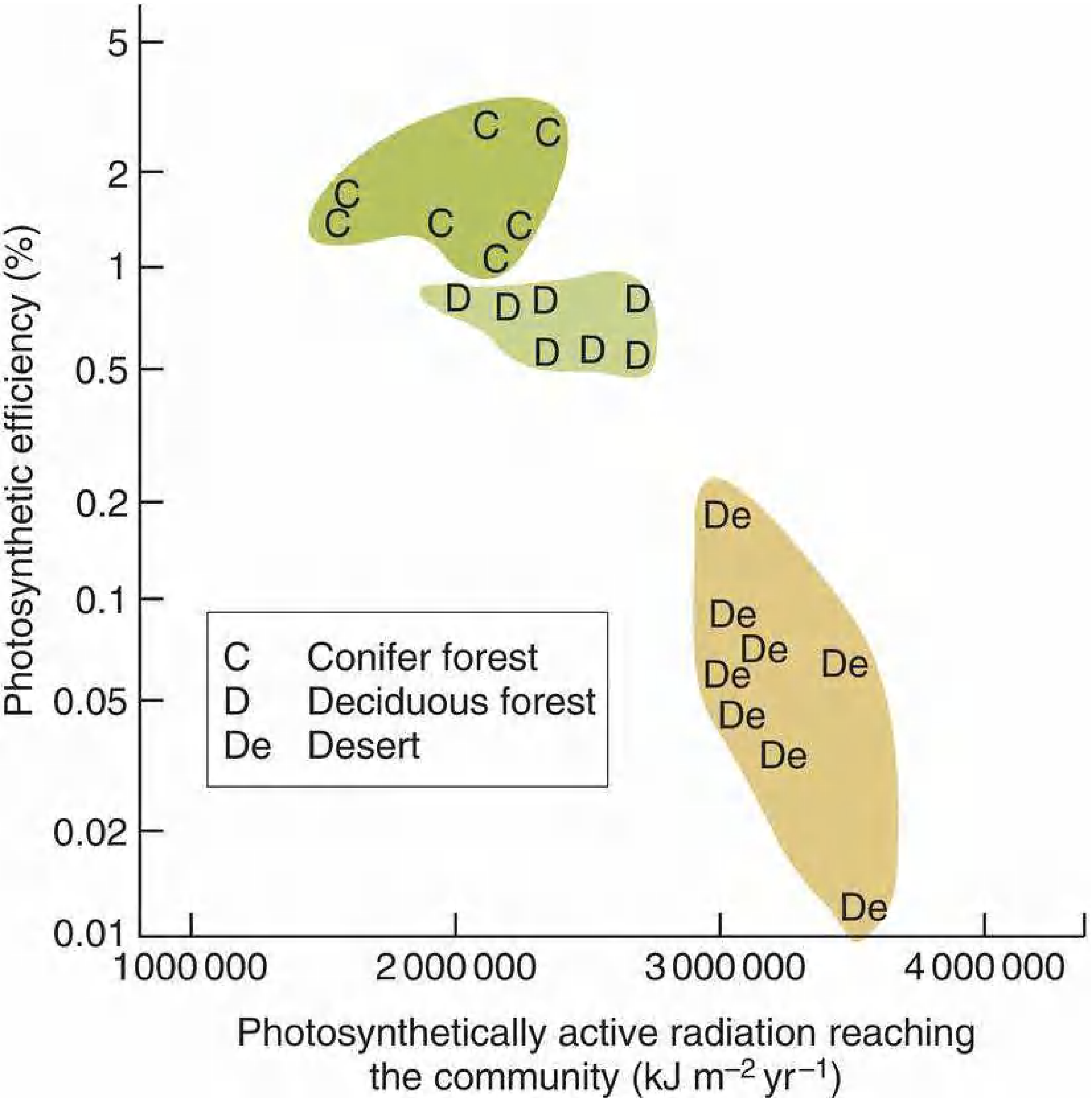


Figure 20.7 Patterns in photosynthetic efficiency in contrasting terrestrial communities. The percentage of incoming photosynthetically active radiation converted to above-ground net primary productivity is shown for three sets of terrestrial communities in the USA.

Source: After Webb *et al*. (1983).

scaling productivity up to the global level

Our understanding of the canopy's efficiency at converting PAR into biomass is central to accurate modelling of global patterns of productivity, where GPP (or NPP, if plant respiration is accounted for) is expressed as the product of PAR, the fraction of PAR absorbed by the canopy (f_{PAR}), and the plants' efficiency in converting this fraction to biomass (ε). The starting point for global modelling is satellite-derived data on temporal and spatial variability in PAR at the earth's surface, and advances have been made in inferring reasonable values for f_{PAR} and ε according to spatial patterns in vegetation cover and composition (Hilker *et al.*, 2008). Such approaches have furnished the global patterns of productivity displayed elsewhere in this chapter (see Figures 20.1 and 20.2).

productivity may still be limited by a shortage of PAR

Note that the fact that radiation is not used efficiently does not in itself imply that it does not limit community productivity. We would need to know whether at increased intensities of radiation the productivity increased or remained unchanged. Some of the evidence given in Chapter 3 shows that the intensity of light during part of the day is below the optimum for canopy photosynthesis. Moreover, at peak light intensities, most canopies still have their lower leaves in relative gloom, and would almost certainly photosynthesise faster if the light intensity were higher. For C_4 plants, a saturating intensity of radiation never seems to be reached, and the implication is that productivity may in fact be limited by a shortage of PAR even under the brightest natural radiation.

There is no doubt, however, that what radiation is available would be used more efficiently if other resources were in abundant supply. The much higher values of community productivity recorded from agricultural systems, with irrigation and addition of synthetic fertilisers, bear witness to this.

20.3.2 Water and temperature as critical factors

shortage of water may be a critical factor

The relationship between the NPP of a wide range of ecosystems on the Tibetan Plateau and both precipitation and temperature is illustrated in Figure 20.8. Water is an essential resource both as a constituent of cells and for photosynthesis. Large quantities of water are lost in transpiration – particularly because the stomata need to be open for much of the time for CO_2 to enter. It is not surprising that the rainfall of a region is quite closely correlated with its productivity. In arid regions, there is an approximately linear increase in NPP with increase in precipitation, but in the more humid forest climates there is a limit beyond which productivity does not continue to rise. Note that a large amount of precipitation is not necessarily equivalent to a large amount of water available for plants; all water in excess of field capacity will drain away (field capacity is the water quantity that an initially saturated soil is still able to hold against gravity after 2–3 days). A positive relationship between productivity and mean annual temperature can also be seen in Figure 20.8. However, the pattern can be expected to be complex because, for example, higher temperatures are associated with rapid water loss through evapotranspiration; water shortage may then become limiting more quickly.

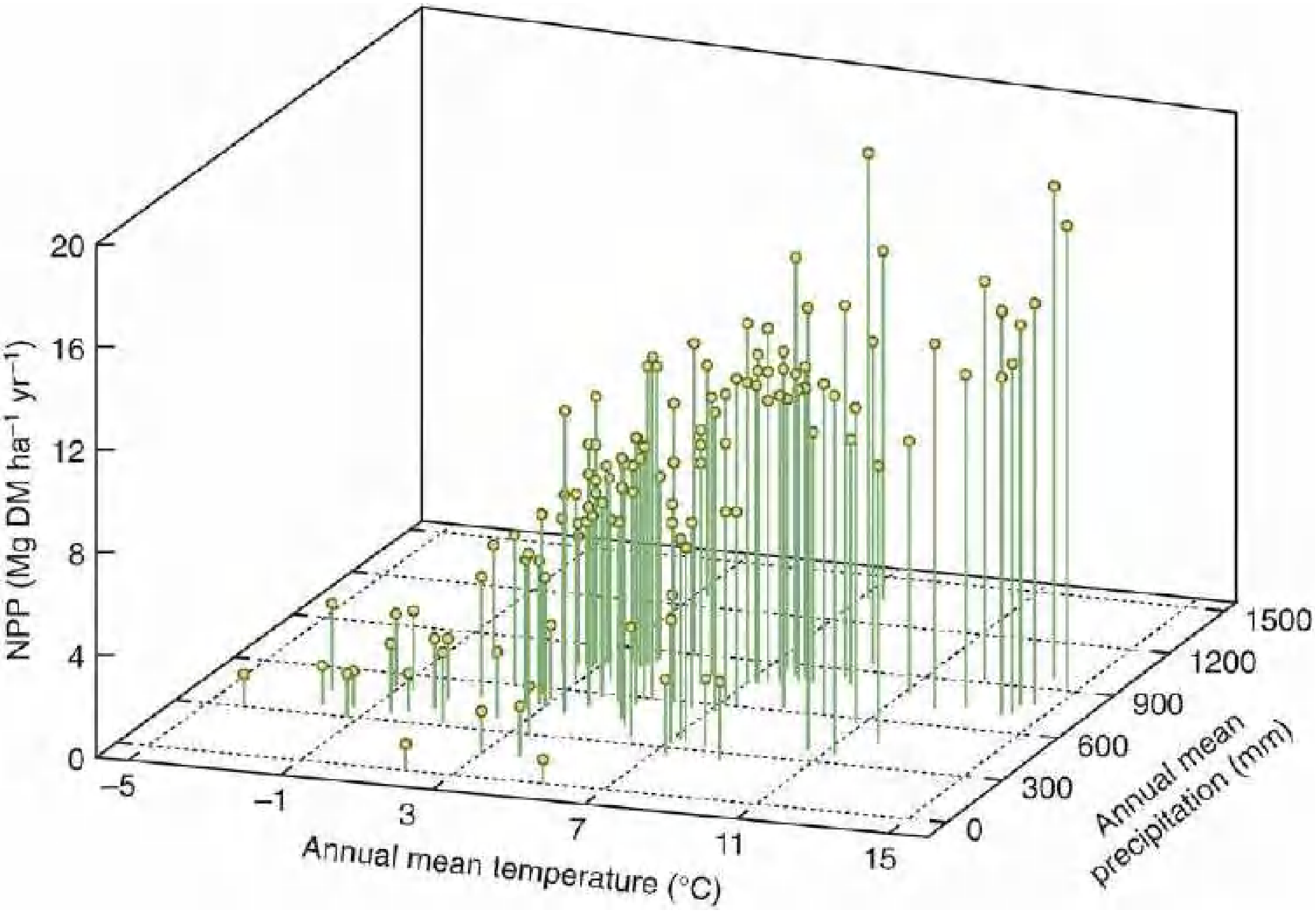


Figure 20.8 Net primary production (NPP) tends to increase with both precipitation and temperature on the Tibetan Plateau. Total NPP and annual precipitation and temperature are shown for ecosystems that include forests, woodlands, shrublands, grasslands and desert. DM, dry matter.

Source: After Luo *et al*. (2002).

productivity and the structure of the canopy

Water shortage has direct effects on the rate of plant growth but also leads to the development of less dense vegetation. Vegetation that is sparse intercepts less light (much of which falls on bare ground). This wastage of solar radiation is the main cause of the low productivity in many arid areas, rather than the reduced photosynthetic rate of drought-affected plants. This point is made by comparing the productivity per unit weight of leaf biomass instead of per unit area of ground for the studies shown in Figure 20.8. Coniferous forest produced 1.64 g DM g^{-1} $year^{-1}$, deciduous forest 2.22 g DM g^{-1} $year^{-1}$ and desert 2.33 g DM g^{-1} $year^{-1}$.

interaction of temperature and precipitation

To unravel the relationships between productivity, rainfall and temperature, it is more instructive to concentrate on a single ecosystem type. Above-ground NPP (ANPP) was estimated for a number of grassland sites along two precipitation gradients (running west to east) in the Argentinian pampas. One of these was in mountainous country and the other in the lowlands. Figure 20.9a shows the relationship between an index of ANPP and precipitation for the two sets of sites. There are strong positive relationships between ANPP and precipitation but the slope of the relationship is steeper for the lowland transect (Figure 20.9a).

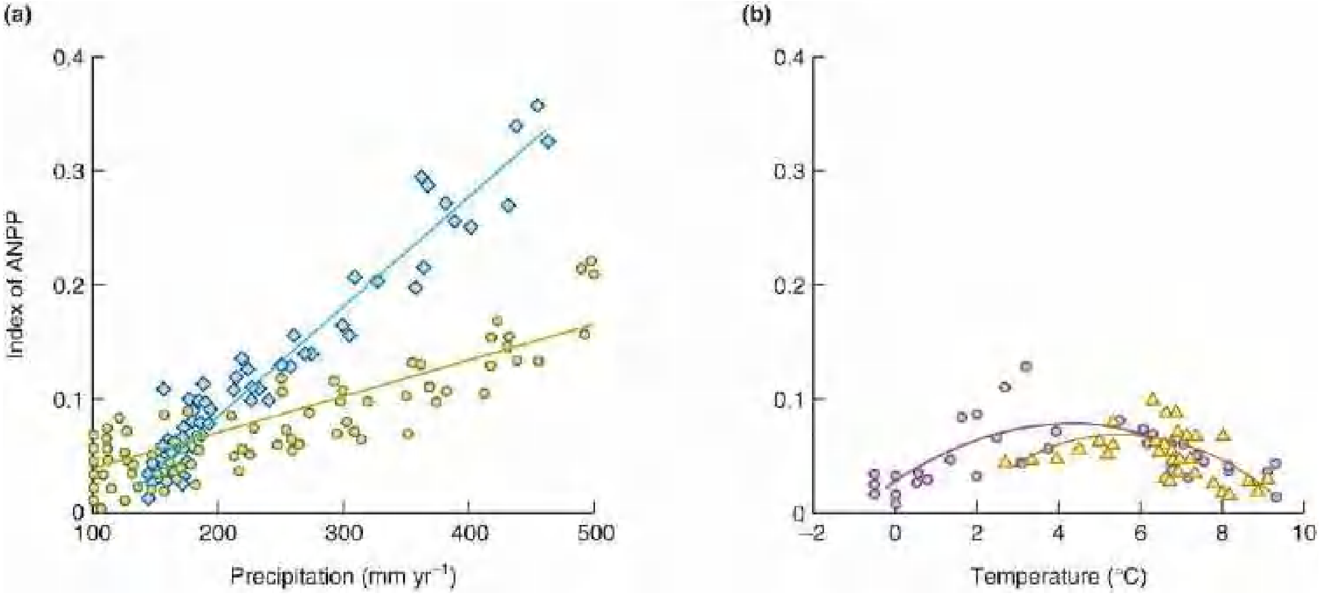


Figure 20.9 **The relationships between NPP, precipitation and temperature in Argentinian pampas grasslands**, showing the annual above-ground net primary productivity (ANPP) of grasslands along two environmental gradients. NPP is shown as an index based on satellite radiometric measurements with a known relationship to absorbed photosynthetically active radiation in plant canopies. (a) NPP in relation to annual precipitation at sites along a lowland (blue diamonds) and a mountainous (green circles) transect. (b) NPP in relation to annual mean temperature at sites along two elevation transects (purple circles and yellow triangles), with colder temperatures at higher elevations.

Source: After Jobbagy *et al.* (2002).

The relationships between ANPP and temperature along a pair of elevation gradients (running north to south) both show a hump-shaped pattern (Figure 20.9b). This pattern probably results from the overlap of two effects of increasing temperature: a positive effect on the length of the growing season and a negative effect through increased evapotranspiration at higher temperatures. Because temperature is the main constraint on productivity at the cool end of the gradients, NPP increases as we move from the cooler to warmer sites. However, there is a temperature value above which the growing season does not lengthen and the dominating effect of increasing temperature is now to increase evapotranspiration, thus reducing water availability and curtailing NPP.

20.3.3 Drainage and soil texture can modify water availability and thus productivity

There was a notable difference in the slopes of the graphs of NPP against precipitation for the mountainous and lowland sites in Figure 20.9. The slope was much lower in the mountainous case and it seems likely that the steeper terrain in this region resulted in a higher rate of water runoff from the land and, thus, a lower efficiency in the use of precipitation (Jobbagy *et al.*, 2002).

soil texture can influence productivity

A related phenomenon has been observed when forest production on sandy, well-drained soils is compared with production on soils consisting of finer particle sizes, which retain more water. Data are available for the accumulation through time of forest biomass at a number of sites where all the trees had been removed by a natural disturbance or human clearance. For forests around the world, Johnson *et al.* (2000) have reported the relationship between above-ground biomass accumulation (a rough index of ANPP) and accumulated growing season degree-days (stand age in years × growing season temperature × growing season as a proportion of the year). In effect, 'growing season degree-days' combine the time for which the stand has been accumulating biomass with the average temperature at the site in question. Figure 20.10 shows that productivity of broadleaf forests is generally much lower, for a given value for growing season degree-days, when the forest is on sandy soil. Such soils have less favourable soil-moisture-holding capacities and this accounts in some measure for their poorer productivity. In addition, however, nutrient retention may be lower in coarse soils, further reducing productivity compared to soils with finer texture. This was confirmed by Reich *et al.* (1997) who, in their compilation of data for 50 North American forests, found that soil nitrogen availability (estimated as annual net nitrogen mineralisation rate) was indeed lower in sandier soils and, moreover, that ANPP was lower per unit of available nitrogen in sandy situations.

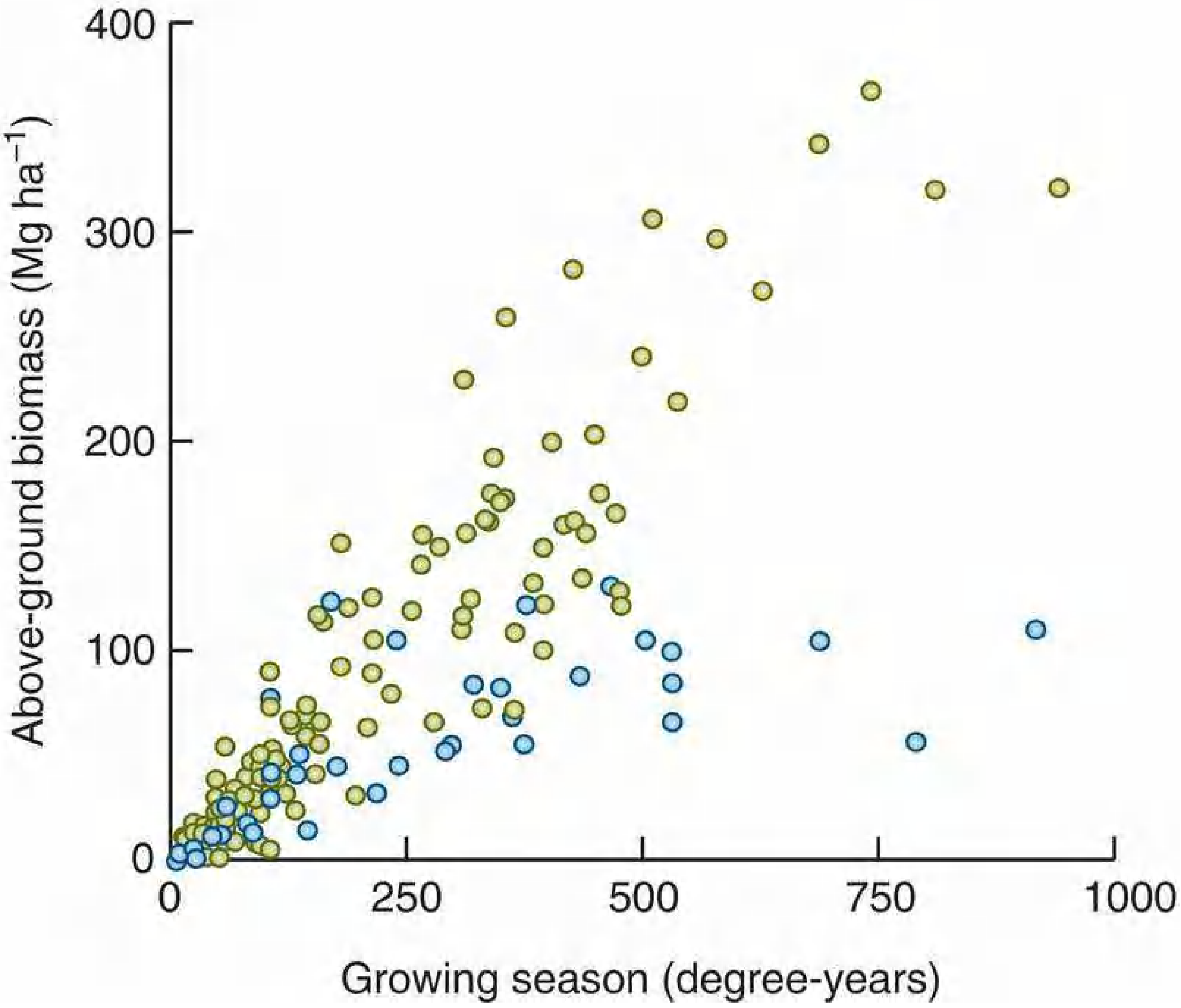


Figure 20.10 Net primary production (NPP) in broadleaf forests tends to increase with the length of growing season. Above-ground biomass accumulation (a rough index of NPP) is expressed as megagrams (= 10^6 g) per hectare in relation to accumulated growing season degree-days in broadleaf forest stands growing on sandy (blue circles) or non-sandy (green circles) soils.

Source: After Johnson *et al.* (2000).

20.3.4 Length of the growing season

The productivity of a community can be sustained only for that period of the year when environmental conditions, particularly temperature or water availability, are favourable for photosynthesis. Deciduous trees are tuned to these environmental conditions, and have evolved to shed leaves when conditions are unfavourable. In general, the leaves of deciduous species photosynthesise fast and die young, whereas evergreen species have leaves that photosynthesise slowly but for longer (see Figure 20.4b). Evergreen trees hold a canopy throughout the year, but during some seasons they may barely photosynthesise at all or may even respire faster than they photosynthesise. Evergreen conifers tend to dominate in nutrient-poor and cold conditions, perhaps because in other situations their seedlings are outcompeted by their faster growing deciduous counterparts (Becker, 2000). The latitudinal patterns in forest productivity seen earlier (see Figure 20.3a) are largely the result of differences in the number of days when there is active photosynthesis.

length of the growing season: a pervasive influence on productivity

Moreover, in our earlier discussion of the study of Argentinian pampas communities (see Figure 20.9) we noted that higher NPP was not only directly affected by precipitation and temperature but was partly determined by length of the growing season. Figure 20.11 shows that the start of the growing season in those pampas communities was unaffected by precipitation (Figure 20.11a, left panel), but positively related to mean annual temperature (earlier growing season start with warmer mean annual temperatures). On the other hand, the end of the growing season was determined partly by temperature (Figure 20.11b, right panel) but also by precipitation (it ended earlier where temperatures were high and precipitation was low such that plants had to close their stomata more often to conserve water). There is a complex interaction between water availability and temperature.

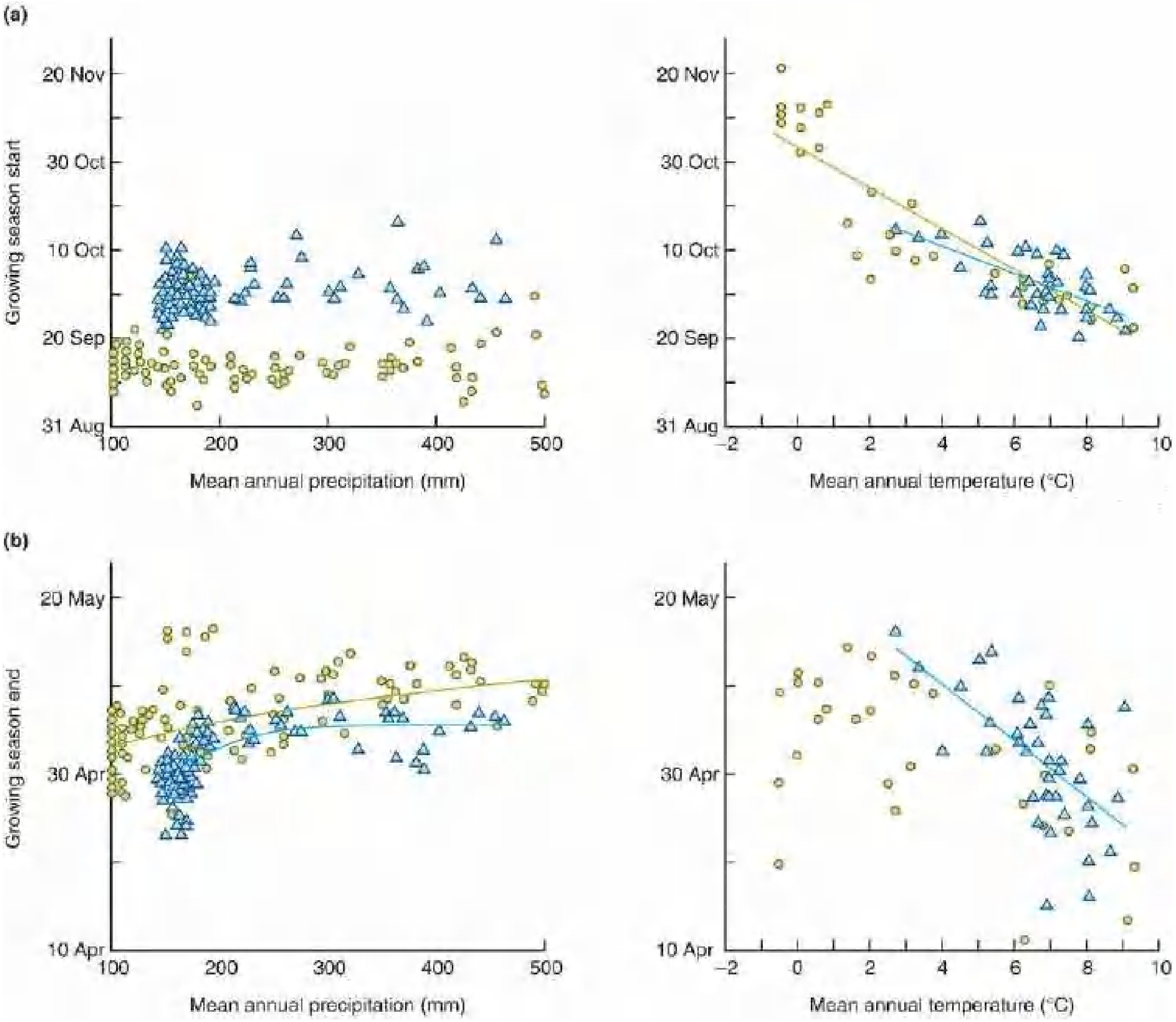


Figure 20.11 Length of growing season in relation to precipitation and temperature for Argentinian pampas communities. (a) Start and (b) end dates of the growing season for pampas communities described in the text. Remember that the austral spring begins in September. Circles represent sites along the precipitation gradient in the mountainous region and triangles represent sites along the lowland gradient. Lines are included where relationships are statistically significant.

Source: After Jobbagy *et al*. (2002).

20.3.5 Productivity may be low because mineral resources are deficient

the crucial importance of nutrient availability

No matter how brightly the sun shines and how often the rain falls, and no matter how equable the temperature is, productivity will be low if the soil is deficient in essential mineral nutrients. Plants need many nutrients but the supply of most is sufficient to meet their needs. Nitrogen or phosphorus, however, are often in relatively short supply and most commonly set limits to NPP. Indeed, in some ecosystems nitrogen and phosphorus are both limiting. Wang *et al*. (2010) built a global model to predict the spatial distribution of areas of nitrogen or phosphorus limitation (or co-limitation) by estimating for both nutrients the rate of supply in relation to the potential demand to support NPP as a function of vegetation type, climate and soil type (Figure 20.12).

Their model predicts that N-limitation or co-limitation is likely in most temperate grasslands and forests. Note that there is probably no agricultural system that does not respond to applied nitrogen by increased primary productivity, and nitrogen fertilisers added to temperate forest soils almost always stimulate forest growth. Nitrogen alone also limits productivity in deserts, tundra and boreal forests. In contrast, the model predicts that phosphorus alone should be more critical in many tropical forests and savannas, where soils are old and most of the original phosphorus contained in them has been eroded away or bound up in chemical forms that are not available to plants. The modellers estimated that phosphorus limitation may reduce NPP by about 20% and nitrogen limitation by up to 40% in comparison to what would be achieved if nutrients were not in short supply.

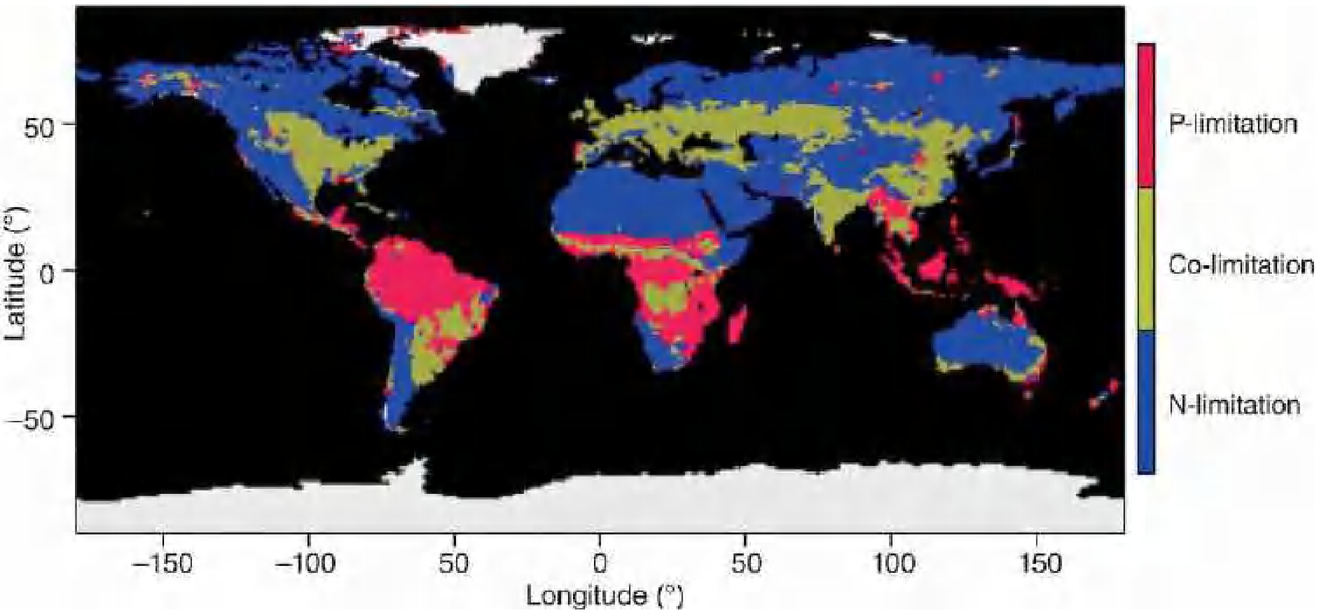


Figure 20.12 Predicted regions of the earth where terrestrial net primary production is N-limited, P-limited or co-limited.

Source: From Wang *et al*. (2010).

Young soils formed from recent weathering of bedrock, including many temperate and boreal areas subject to glaciation in recent millennia, often contain high levels of available phosphorus, and thus N-limitation is prevalent. But phosphorus is leached away or becomes bound in unavailable forms as soils age and P-limitation is commonly seen in the very old soils that prevail in tropical areas that have not experienced glaciation for millions of years.

limitation by a succession of factors

In the course of a year, the productivity of a community may (and usually will) be limited by a succession of factors including limitation by water, nutrients, temperature, soil depth and the efficiency with which leaves photosynthesise. In a grassland community, for instance, primary productivity may be far below the theoretical maximum because the winters are too cold and light intensity is low, the summers are too dry, the rate of nitrogen mobilisation is too slow, and for periods grazing animals may reduce the standing crop to a level at which much incident light falls on bare ground.

20.3.6 Do community composition and species richness affect ecosystem productivity?

higher productivity associated with higher species richness – a general finding

In Sections 17.2.3 and 17.2.4, we discussed the effects of species richness (and community complexity generally) on community stability and noted the associated effects on overall productivity. We now turn to these directly. Generally, experimental increases in species richness (a community property) result in increases in primary productivity (an ecosystem property) in both laboratory (e.g. Figure 20.13a) and field (e.g. Figure 20.13b) experiments. As explained in Section 17.2.3, two principal hypotheses have been advanced to account for these positive relationships (Tilman *et al.*, 2014).

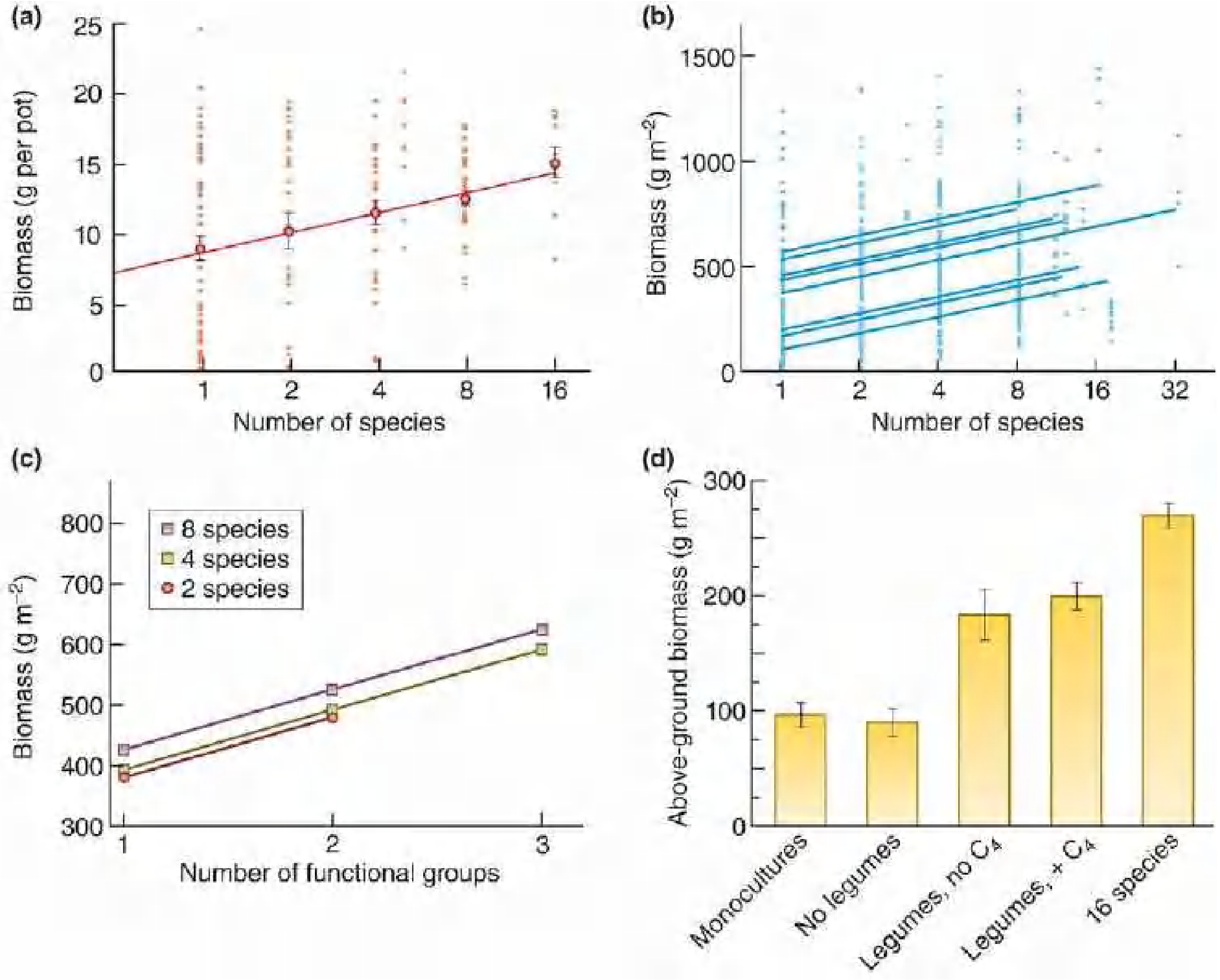


Figure 20.13 Primary production tends to increase with both species and functional group richness. (a) Primary productivity measured as biomass accumulation versus species richness in experimental plant assemblages grown under controlled conditions. Bold red circles are means (with SEs). Small points show values for each assemblage individually. (b) Primary productivity measured as biomass accumulation after 2 years in a large number of simulated grassland assemblages across Europe (regression lines are for different countries). (c) Primary productivity versus functional group richness (grasses, N-fixing legumes, other herbs) in the European grasslands combined. (d) Productivity measured as above-ground biomass after 13 years of the Cedar Creek grassland biodiversity experiment in relation to plant functional group composition.

Source: (a) After Naeem *et al.* (1995). (b, c) After Hector *et al.* (1999). (d) After Mueller *et al.* (2013).

complementarity, selection and transgressive overyielding

On one hand, the *complementarity hypothesis* proposes that if species show niche differentiation (see Chapter 8), then they may use resources in complementary ways, utilising, between them, a greater proportion of available resources and thereby achieving a higher level of productivity. Complementarity might also result from partitioning of natural enemies: for example, after soil fungicidal treatment, perennial grass biomass in monocultures increased to levels similar to plant mixtures, presumably as a result of reducing the negative influence of soil pathogen build-up on conspecific hosts (Maron *et al.*, 2011). As long as the complementarity hypothesis applies, a case can be made for the need for management to conserve biodiversity to maintain ecosystem functioning.

On the other hand, the *selection hypothesis* proposes that positive relationships between richness and productivity may arise simply because when more species are present in an assemblage, the assemblage is more likely, by chance, to contain a competitively dominant species that is highly productive.

As noted in Section 17.2.3, if the complementarity effects are strong enough (Loreau, 2004), and especially if there are also selection effects, then this will lead to ‘transgressive overyielding’, where the rate of production in more diverse communities is greater than can be achieved in any of the communities with fewer species. With selection effects alone, species-rich communities may on average be more productive because they are more likely to be dominated by especially productive species, but a multispecies community can be no more productive than a monoculture of its most productive species.

complementarity and selection both play a role

With improvements to experimental design and statistical analyses, the consensus now, from hundreds of studies involving terrestrial, freshwater and marine ecosystems, is that both species-specific selection effects *and* multispecies complementarity play significant roles in the net positive effect of species richness on productivity (Cardinale *et al.*, 2011). In the case of the greenhouse experiment shown in Figure 20.13a, for example, none of the higher richness species mixtures had greater productivity than the best single species growing alone, a result consistent with the selection hypothesis. On the other hand, complementarity effects are suggested by studies in which higher plant functional richness is associated with increased productivity. This is apparent for grasses, legumes and other herbs in Figure 20.13c and for legumes and C_4 grasses in Figure 20.13d.

More broadly, in a meta-analysis of 44 separate experiments on a wide range of plant communities, Cardinale *et al.* (2007) estimated increased productivity as the proportional difference between the yield in the most diverse polyculture in each case and the mean of the yields of all the constituent species grown in monoculture. They found an increased yield with greater diversity in 79% of cases (Figure 20.14a), though there was *transgressive* overyielding – polyculture yields greater than those of the most productive species – in only 12% of the experiments. Remember that the richness–productivity outcome depends on both complementarity and selection. Only if the most productive species is competitively dominant and there is strong complementarity is transgressive overyielding likely to be evident.

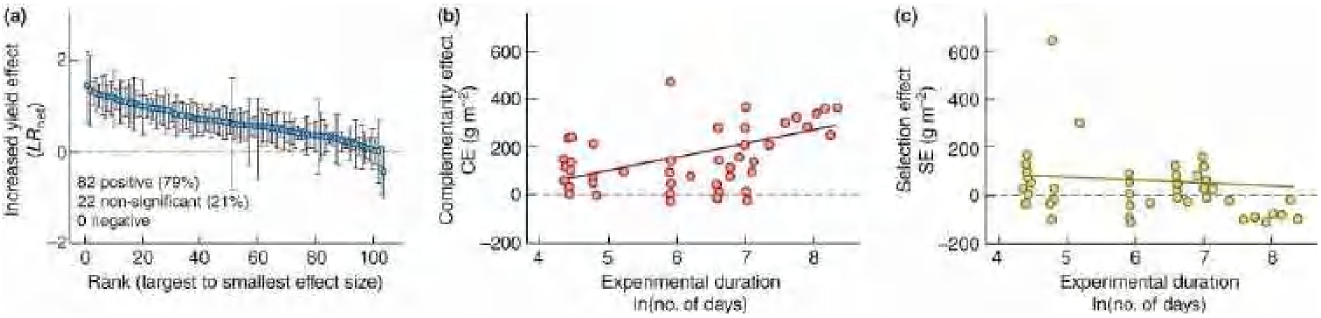


Figure 20.14 Increased yield with species richness in grassland experiments increases with experiment duration as a result of complementarity but not selection effects. (a) Increased yield effects in 104 experimental studies, measured as LR_{net}, the log ratio of biomass in the most diverse polyculture to the average of all species in monoculture. Bars are 95% CIs. (b) The increase in complementarity effect (CE) with experiment duration for 47 estimates available from the experiments in (a) ($F = 13.3$, $P < 0.01$). (c) No relationship was found between selection effect (SE) and experiment duration for 47 estimates available from the experiments in (a) ($F = 0.61$, $P = 0.44$). For methods for estimating CE and SE, see Loreau and Hector (2001).

Source: After Cardinale *et al*. (2007).

For some of the studies, Cardinale *et al*. (2007) then went on to partition the increased yield (not only the overyielding) into that attributable to complementarity and that attributable to selection effects, using a method developed by Loreau and Hector (2001). Complementarity was far more important than selection. Two-thirds of the species richness effects were due to complementarity and almost half of the selection effects were negative (dominance of low-yielding species) rather than positive. Similar results were found by Loreau and Hector (2001) for the eight grasslands experiments described in Figure 17.12. Moreover, these complementarity effects became increasingly important the longer the duration of an experiment, whereas there was no such trend with the selection effects (Figure 20.14b, c). Hence, given the inevitable constraints on performing lengthy experimental studies, the true importance of complementarity may have been underestimated.

APPLICATION 20.3 How important is biodiversity loss compared with other human-induced factors?

Biodiversity loss is a major and growing concern and, based on the results in this section, reductions in species richness can be expected to have substantial consequences for ecosystem functioning. But how do these consequences compare with changes to primary productivity attributable to other human impacts?

In their analysis of a wide variety of experiments in the Cedar Creek grassland study, Tilman *et al.* (2012) compared differences in biomass production between treatment and control plots for high richness versus low richness plots as well as a range of other potential human impacts: nitrogen pollution/fertilisation (three application rates), CO_2 increase (560 ppm versus ambient), water (irrigation versus ambient), herbivory (fenced versus unfenced), drought and annual burning (Figure 20.15). They concluded that reductions in species richness of the magnitude imposed by human actions had impacts as large or larger than other equally relevant human perturbations.

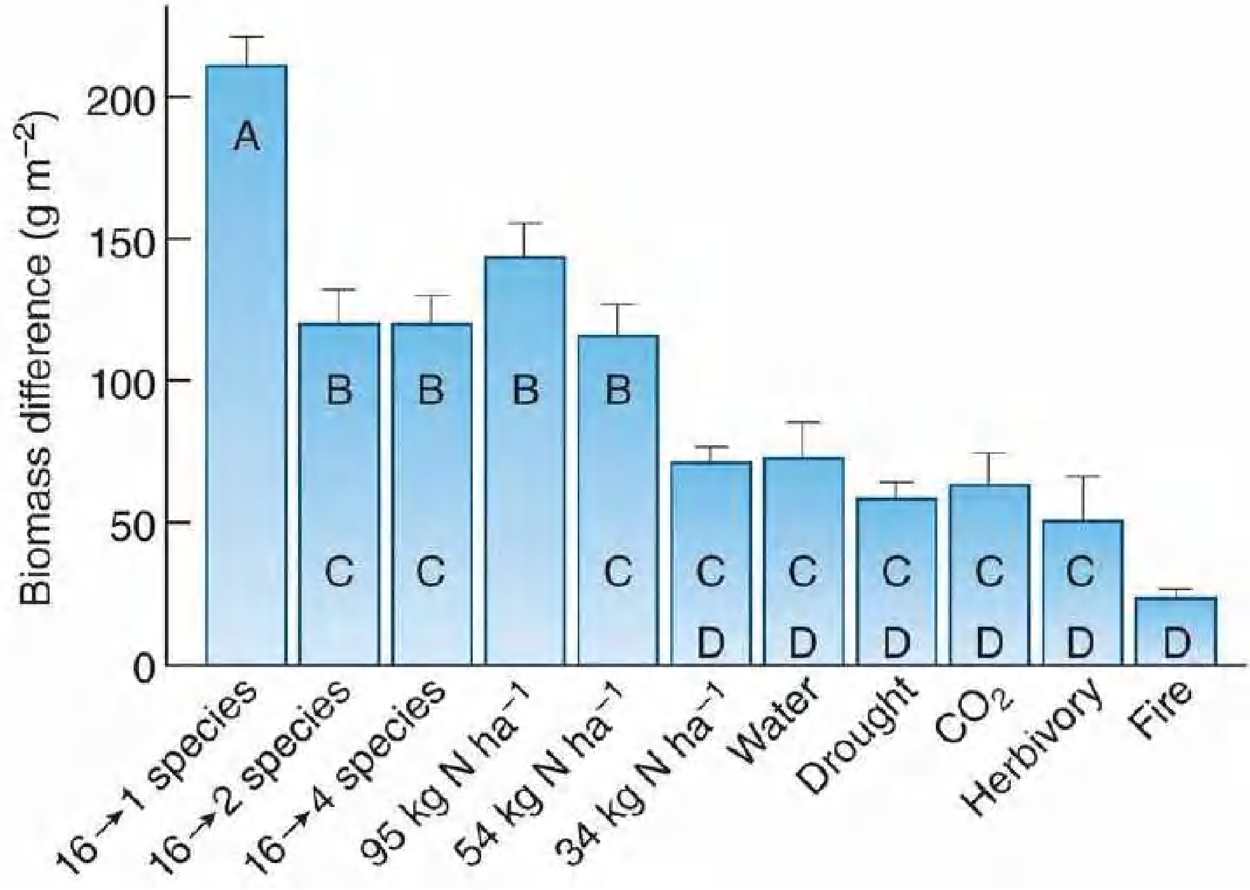


Figure 20.15 Comparing richness-based and human perturbation effects on biomass production. Differences in biomass production in experimental plots at Cedar Creek comparing 16 versus one, two and four species richness treatments, as well as various nitrogen addition rates, irrigation, drought, CO_2 enrichment, herbivory and burning. Bars with the same letter are not significantly different.

Source: After Tilman *et al.* (2012).

In a related meta-analysis, but including nearly 200 studies from terrestrial, freshwater and marine settings, Hooper *et al.* (2012) concluded that intermediate levels of species loss (21–

40%) produced drops in primary productivity comparable to those documented for climate warming or enhanced ultraviolet radiation, while higher levels of extinction (41–60%) rivalled the effects of elevated ozone, acidification, elevated CO_2 and nutrient pollution. These results add weight to the need to place a high priority on the restoration and conservation of biodiversity, not simply for the retention of the individual species but for the ecosystems services that, between them, they provide. We should note that in experimental studies, we can detect the indirect effects of biodiversity loss because the direct effects of other factors can be controlled. In observational studies, however, human perturbations may have already affected these other factors, perhaps with consequent effects on biodiversity loss, so that it is no easy matter to separate direct and indirect effects.

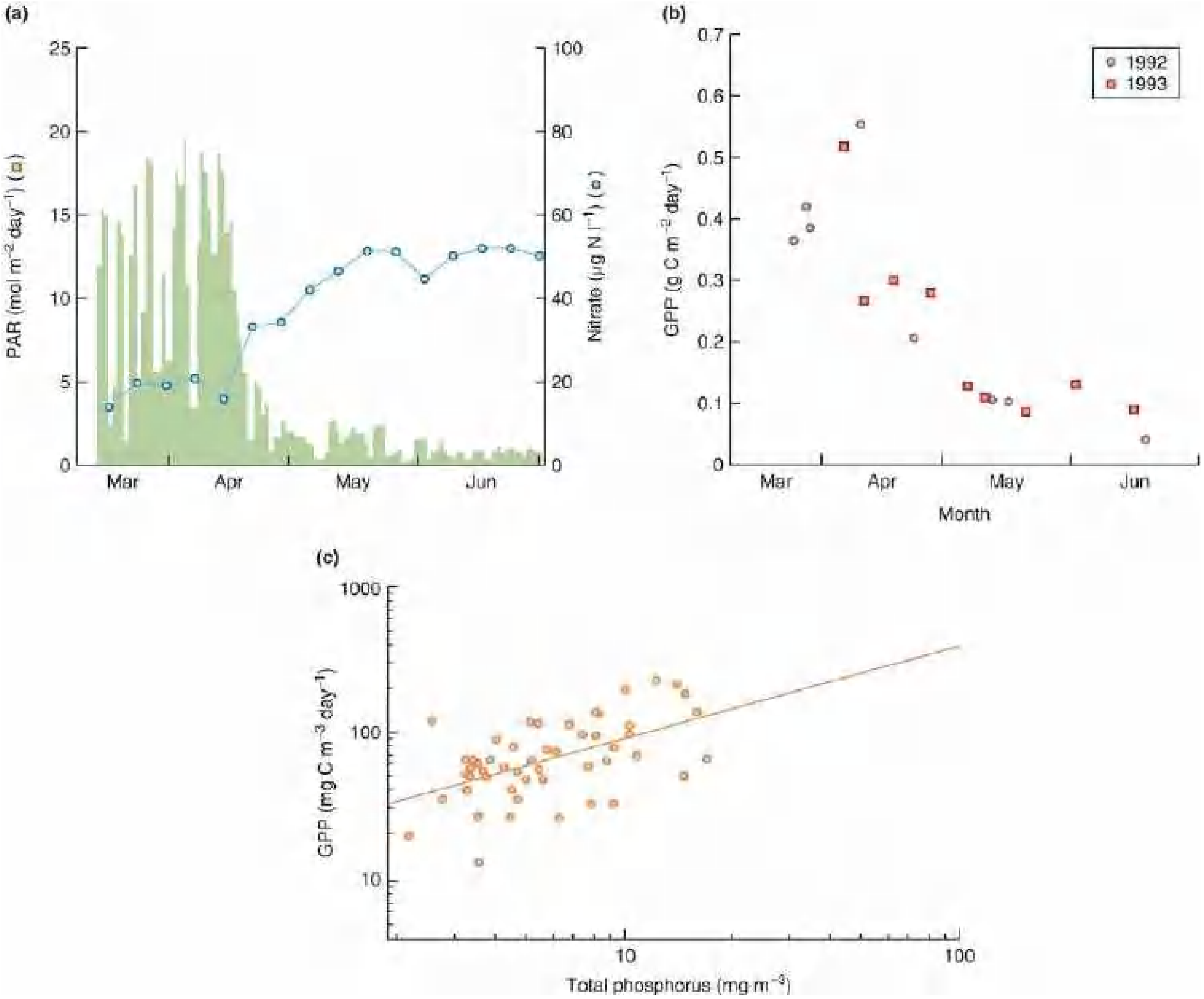
20.4 Factors limiting primary productivity in aquatic communities

The factors that most frequently limit the primary productivity of aquatic environments are the availability of light and nutrients. The most commonly limiting nutrients are nitrogen (usually as nitrate) and phosphorus (phosphate), but iron can be important in open ocean environments.

20.4.1 Limitation by light and nutrients in streams

in small forest streams, light and nutrients interact to determine productivity

Streams flowing through deciduous forests undergo marked transitions in primary production by algae on the stream bed during the growing season, as conditions shift from light-replete early in spring to severely light-limited when leaves develop on the overhanging trees. In a stream in Tennessee, USA, leaf emergence reduced PAR reaching the stream bed from more than 1000 to less than 30 μmol m^{-2} s^{-1} (Hill *et al.*, 2001). The reduction in PAR was paralleled by an equally dramatic fall in stream GPP (Figure 20.16). This is despite a large increase in photosynthetic efficiency from less than 0.3% to 2%; the higher efficiencies arose both because existing taxa acclimated physiologically to low irradiances and because more efficient taxa became dominant later in the season. Intriguingly, as PAR levels fell, the concentration in stream water of both nitrate (Figure 20.16a) and phosphate rose. It seems that nutrients limited primary production when PAR was abundant early in spring, with uptake by the algae reducing the concentration in the water at this time. When light became limiting, however, the reduction in algal productivity meant that less of the available nutrients were removed from the supply in the flowing water.



[Figure 20.16](#) **Relationship between primary production, photosynthetically active radiation (PAR) and phosphorus in freshwater ecosystems.** (a) PAR reaching the bed of a Tennessee stream (green bars) and stream water nitrate concentration (blue circles and line) during the spring of 1992 (the patterns were very similar in 1993). (b) Gross primary productivity (GPP) in the stream during the spring in 1992 and 1993 (calculated on the basis of whole-stream diurnal changes in oxygen concentration). (c) Relationship between GPP of phytoplankton in the open water of some Canadian lakes and phosphorus concentration. Metabolic measurements were made in bottles in the laboratory at lake temperatures on depth-integrated water samples taken from the field.

Source: (a, b) After Hill *et al*. (2001). (c) After Carignan *et al*. (2000).

20.4.2 Lakes and estuaries: the importance of nutrients and of autochthonous production

productivity in lakes and estuaries shows a pervasive role for nutrients ...

Like streams, lakes receive nutrients by the weathering of rocks and soils in their catchment areas, in the rainfall, and as a result of human activity (fertilisers and sewage input). They vary considerably in nutrient availability. A study of 12 Canadian lakes shows a clear relationship between GPP and phosphorus concentration and demonstrates the importance of nutrients in limiting lake productivity (Figure 20.16c). In a meta-analysis of several hundred aquatic ecosystem studies from around the world, Hoellein *et al.* (2013) confirmed that lake GPP was generally driven by phosphorus concentrations, and the same was the case for estuaries, which are among the most productive ecosystems of all.

It is notable that nutrients are far more limiting in aquatic than in terrestrial ecosystems. While a relative scarcity of nitrogen or phosphorus may depress terrestrial primary productivity by 20–40% (Section 20.3.5), lake GPP in Figure 20.16c increased by more than 300% as the concentration of phosphorus increased, and other aquatic studies have shown even steeper increases in relation to phosphorus or nitrogen availability.

... and gross primary production is well matched to ecosystem respiration

GPP in lakes and estuaries was highly correlated with ecosystem respiration (RE; the sum of autotrophic and heterotrophic respiration) (Figure 20.17); thus, these ecosystems are fixing carbon at roughly the same rate as it is being released by the respiration of plants and heterotrophs (see Section 20.1). The majority of streams (and some wetlands), on the other hand, had RE rates that easily exceeded their GPP. These patterns reflect the overriding importance of autochthonous production in larger lakes (and estuaries) and of allochthonous carbon inputs to streams (and wetlands).

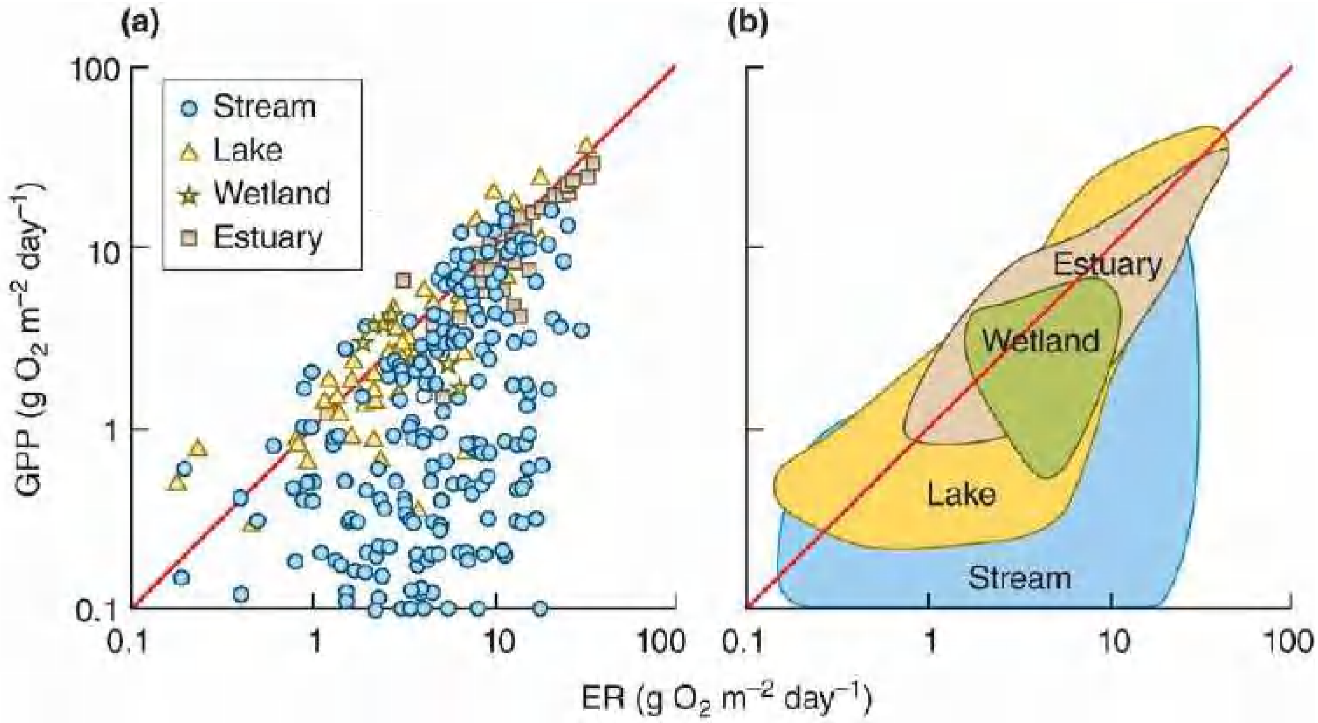


Figure 20.17 Relationship between gross primary productivity (GPP) and ecosystem respiration (RE) in aquatic ecosystems around the globe. (a) Expressed as ecosystem metabolism on the basis of daily oxygen production (GPP) and oxygen consumption (RE) in summer. (b) Schematic of the relationship.

Source: After Hoellein *et al.* (2013).

20.4.3 Nutrients and the importance of upwelling in oceans

both nitrogen and phosphorus limit NPP in nutrient-poor ocean gyres

Net primary productivity in the open ocean is very low and is limited by both nitrogen and phosphorus. In subtropical gyres (major ocean circulations around a high-pressure system), for example, the concentration of inorganic nitrogen in surface waters is low and inorganic phosphorus even lower (e.g. Smith, 1984). However, phytoplankton also contain less phosphorus than nitrogen, and the relative proportions of dissolved nutrient concentrations in ocean water and in phytoplankton biomass are virtually the same throughout much of the world's oceans. This N : P ratio by moles is 16 : 1, termed the Redfield ratio after Alfred Redfield for his pioneering work in the 1930s (Gruber & Deutsch, 2014). The relative constancy of the N : P ratio in seawater largely explains why both nitrogen and phosphorus limit production there.

The 'constancy' of the Redfield ratio is a relative term, because at particular times and places either nutrient may be in lower supply in the water (e.g. Karl *et al.*, 2001) and species of phytoplankton differ in their average stoichiometric ratios and the extent to which these vary according to local conditions. At larger spatial and temporal scales, however, the importance of the two nutrients in limiting production is similar. This relative constancy is unlikely to be a coincidence. Rather, homeostatic processes seem likely to be at play, whereby N-fixation by cyanobacteria and other microbes adds nitrate when the ratio is below 16 : 1, while denitrifying bacteria remove a nitrate excess when the ratio is above 16 : 1. The precise nature of the mechanisms governing such homeostasis is still not fully understood (Gruber & Deutsch, 2014). Experimental manipulation of nitrogen and phosphorus in ocean water bodies may not be possible but such manipulation in some Canadian lakes has provided support for the idea of homeostasis. Thus, when N-fertilisation was reduced while keeping P-fertilisation high, so that their proportions fell below the Redfield ratio, N-fixing bacteria appeared in a matter of weeks; these organisms were not observed in the lakes when nitrogen was plentiful (Flett *et al.*, 1980).

rich supplies of nutrients in marine environments … from estuaries …

Locally high levels of primary productivity in the oceans are associated with high nutrient inputs from two sources. First, nutrients may flow continuously into coastal shelf regions from estuaries. Productivity in the inner continental shelf region is particularly high both because of high nutrient concentrations and because the relatively clear water provides a reasonable depth within which net photosynthesis is positive (the *euphotic zone*). Closer to land, the water is richer in nutrients but is usually highly turbid and its productivity is less. The least productive zones are on the outer shelf (and, as noted above, in the open ocean) where primary productivity might be expected to be high because the water is clear and the euphotic zone is deep; here, however, productivity is low because of the extremely low concentrations of nutrients.

… and upwellings

Ocean upwellings are a second source of high nutrient concentrations. These occur on continental shelves where the wind is consistently parallel to, or at a slight angle to, the coast. As a result, water moves offshore and is replaced by cooler, nutrient-rich water originating from the bottom, where nutrients have been accumulating by sedimentation. Strong upwellings can also occur adjacent to submarine ridges, as well as in areas of very strong currents. Where it reaches the surface, the nutrient-rich water sets off a bloom of phytoplankton production. A chain of heterotrophic organisms takes advantage of the abundant food, and the great fisheries of the world are located in these regions of high productivity.

iron as a limiting factor in oceans

Iron is a limiting nutrient that potentially affects about one-third of the open ocean. Being very insoluble in seawater, iron is ultimately derived from wind-blown dust from continents and large areas of ocean receive insufficient amounts. When iron is added experimentally to ocean areas, massive blooms of phytoplankton can result (Boyd *et al*., 2007); such blooms are also likely to occur when large storms supply land-derived iron to the oceans (e.g. Qiu, 2015).

temperature and PAR also affect productivity

While nutrients are the most influential factors for local ocean productivity, temperature and PAR also play a role at a larger scale (Figure 20.18). This aids our ability to estimate ocean primary productivity because sea surface temperature and PAR (together with surface chlorophyll concentration, another factor correlated with NPP) can be measured using satellite telemetry.

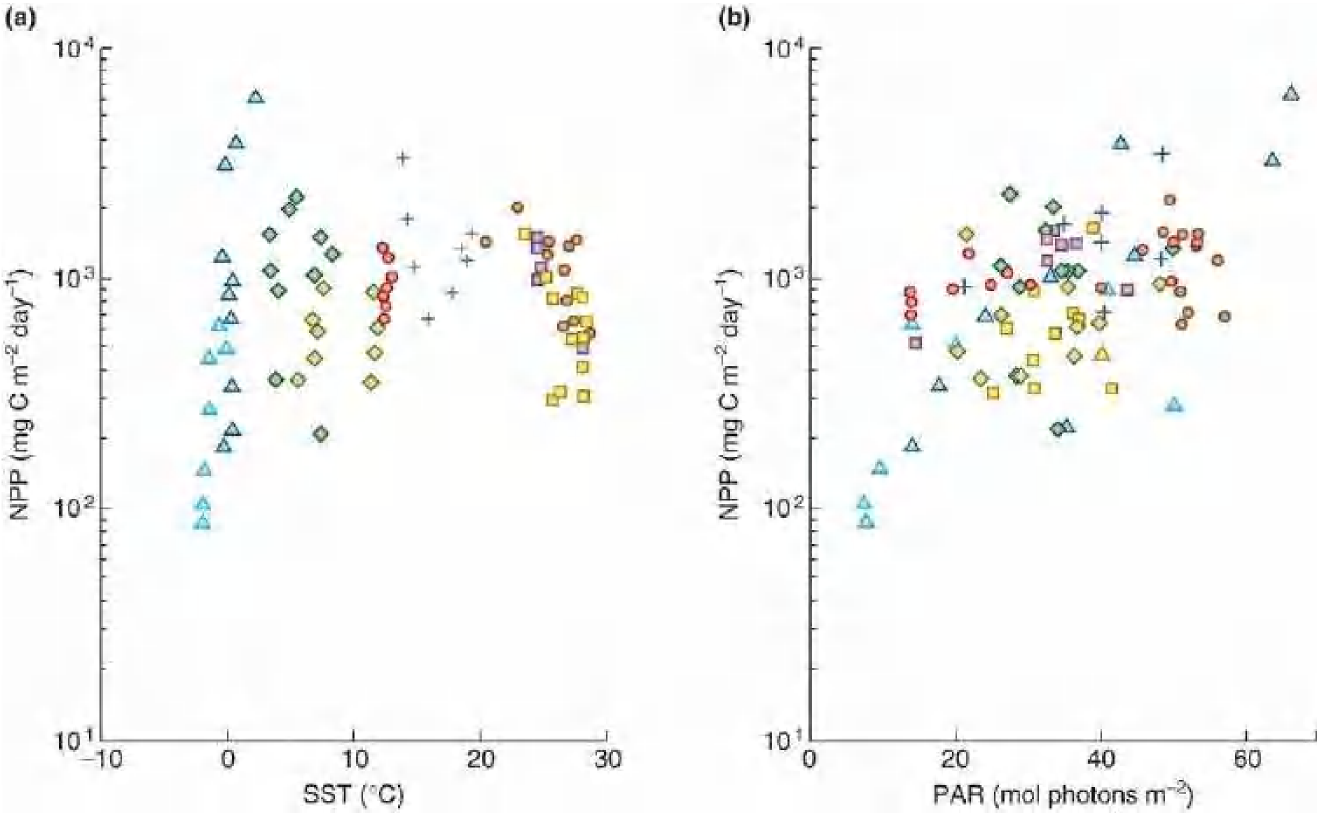


Figure 20.18 Net primary production (NPP) in relation to temperature and photosynthetically active radiation (PAR) in the ocean. Relationships between daily depth-integrated estimates of NPP and: (a) sea surface temperature (SST) and (b) above-water daily PAR. The different symbols relate to different datasets from various oceans.

Source: After Campbell *et al*. (2002).

20.4.4 Productivity varies with depth in aquatic communities

phytoplankton productivity varies with depth

Although the concentration of a limiting nutrient usually determines the productivity of aquatic communities on an areal basis, in any given water body there is also considerable variation with depth as a result of attenuation of light intensity. Light is absorbed by water molecules as well as by dissolved and particulate matter, and it declines exponentially with depth. Near the surface, light is superabundant, but at greater depths its supply is limited and light intensity ultimately determines the extent of the euphotic zone. Figure 20.19a shows how GPP declines with depth. The depth at which GPP is just balanced by phytoplankton respiration is known as the compensation point. Above this, NPP is positive. Very close to the surface, particularly on sunny days, there may even be photoinhibition of photosynthesis. This seems to be due largely to radiation being absorbed by photosynthetic pigments at such a high rate that it overflows into destructive photo-oxidation reactions.

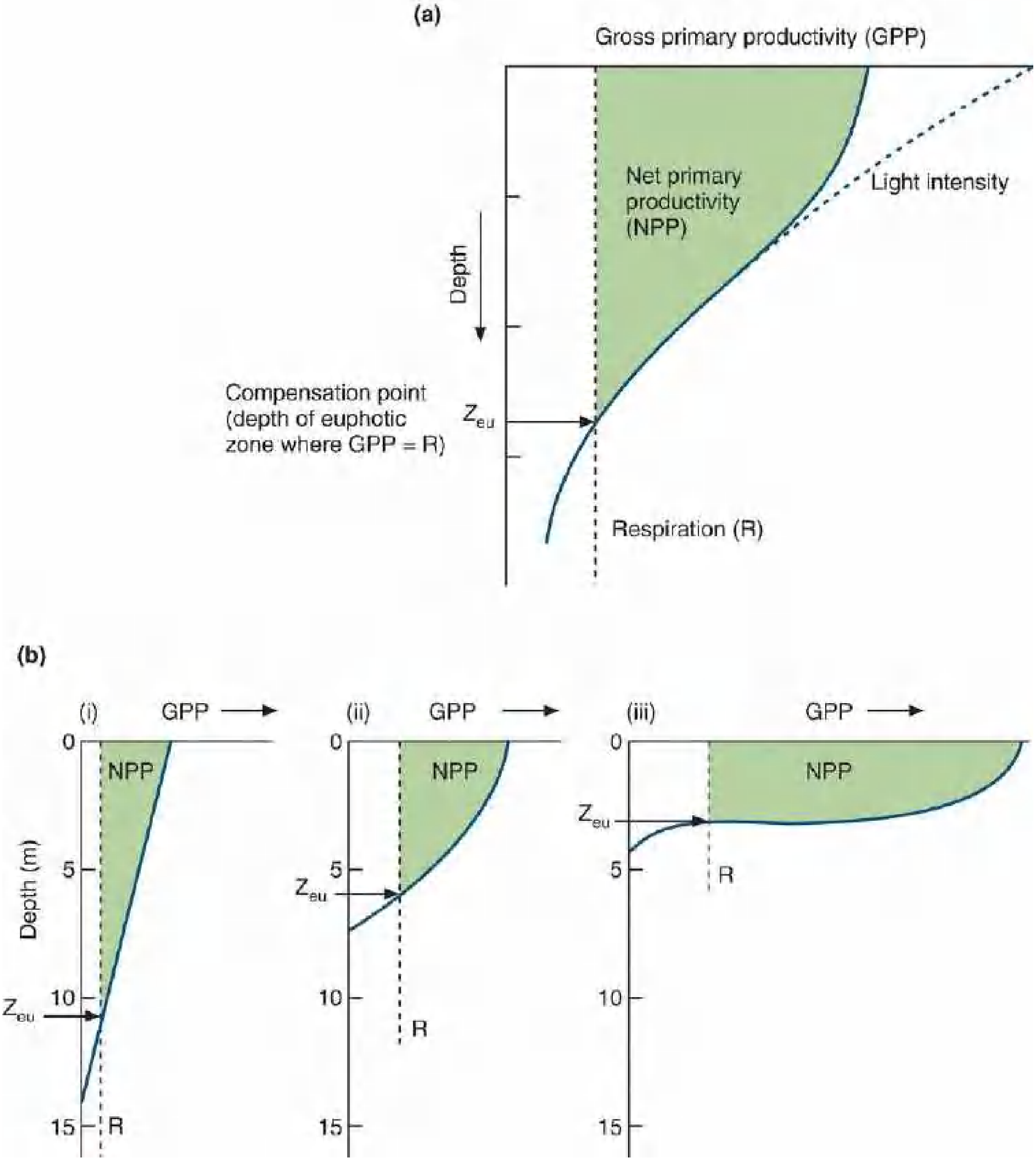


Figure 20.19 Gross primary production (GPP) declines with ocean depth. (a) The general relationship with depth, in a water body, of GPP, respiratory heat loss (R) and net primary productivity (NPP). The compensation point (or depth of the euphotic zone, eu) occurs at the depth (Z_{eu}) where GPP just balances R and NPP is zero. (The disparity between the dashed line for light intensity and GPP relates to near-surface photoinhibition of photosynthesis.) (b) Total NPP increases with nutrient concentration in the water (lake i < ii < iii). Increasing fertility itself is responsible for greater biomasses of phytoplankton and a consequent decrease in the depth of the euphotic zone.

The more nutrient-rich a water body is, the shallower its euphotic zone is likely to be (Figure 20.19b). This is not really a paradox. Water bodies with higher nutrient concentrations usually possess greater biomasses of phytoplankton that absorb light and reduce its availability at greater depth. (This is exactly analogous to the shading influence of the tree canopy in a forest, which may remove up to 98% of the radiant energy before it can reach the ground layer vegetation or, as we saw above, a stream bed.) Even quite shallow lakes, if sufficiently fertile, may be devoid of water weeds on the bottom because of shading by phytoplankton. The relationships shown in both parts of Figure 20.19 are derived from lakes but the pattern is qualitatively similar in ocean environments (Figure 20.20).

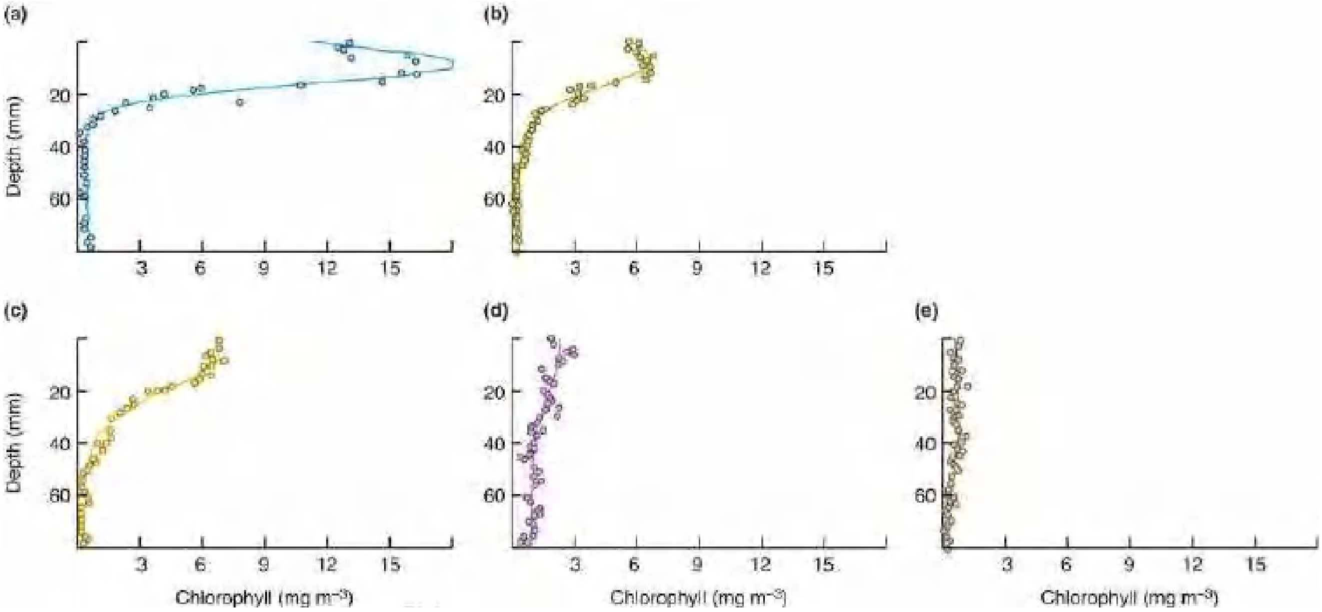


Figure 20.20 Examples of vertical chlorophyll profiles recorded in the ocean off the coast of Namibia. Example (a) is typical of locations associated with ocean upwelling: as cold upwelled water warms up, a surface phytoplankton bloom develops, reducing light penetration and thus productivity in deeper water. Example (b) illustrates how peak abundance can shift to deeper water as a surface bloom in an upwelling area depletes the nutrient concentrations there. The surface phytoplankton bloom in example (c) is less dramatic than in (a) (perhaps reflecting lower nutrient concentrations in the upwelling water); as a result, chlorophyll concentration remains relatively high to a greater depth. Examples (d) and (e) are for locations where nutrient concentrations are much lower.

Source: After Silulwane *et al.* (2001).

20.5 The fate of energy in ecosystems

Secondary productivity is defined as the rate of production of new biomass by heterotrophic organisms. Unlike autotrophs, heterotrophs cannot manufacture from simple molecules the complex, energy-rich compounds they need. They derive their matter and energy either directly by consuming autotroph biomass or indirectly from autotrophs by eating other heterotrophs. Autotrophs, the primary producers, comprise the first trophic level in a community; primary consumers occur at the second trophic level; secondary consumers (carnivores) at the third, and so on.

20.5.1 Patterns among trophic levels

there is a general positive relationship between primary and secondary productivity

Since secondary productivity depends on primary productivity, we should expect a positive relationship between the two variables in communities. Turning again to the stream study described in Section 20.4.1, recall that primary productivity declined dramatically during the summer when a canopy of tree leaves above the stream shaded out most of the incident radiation. A principal grazer of the algal biomass is the snail *Elimia clavaeformis*. Figure 20.21a shows how the growth rate of individual snails in the stream was also lowest in the summer; there was a statistically significant positive relationship between snail growth and monthly stream bed PAR (Hill *et al.*, 2001). Figure 20.21b–d illustrates the general relationship between primary and secondary productivity in aquatic and terrestrial examples. Secondary productivity by zooplankton, which principally consume phytoplankton cells, is positively related to phytoplankton productivity in a range of lakes in different parts of the world (Figure 20.21b). Note, however, that zooplankton production is consistently only about 10% of phytoplankton production. The productivity of heterotrophic bacteria in lakes and oceans also parallels that of phytoplankton (Figure 20.21c); they metabolise dissolved organic matter released from intact phytoplankton cells or produced as a result of ‘messy feeding’ by grazing animals. Figure 20.21d shows how the productivity of *Geospiza fortis* (one of Darwin’s finches), measured in terms of average brood size on an island in the Galápagos archipelago, is related to annual rainfall, itself an index of primary productivity.

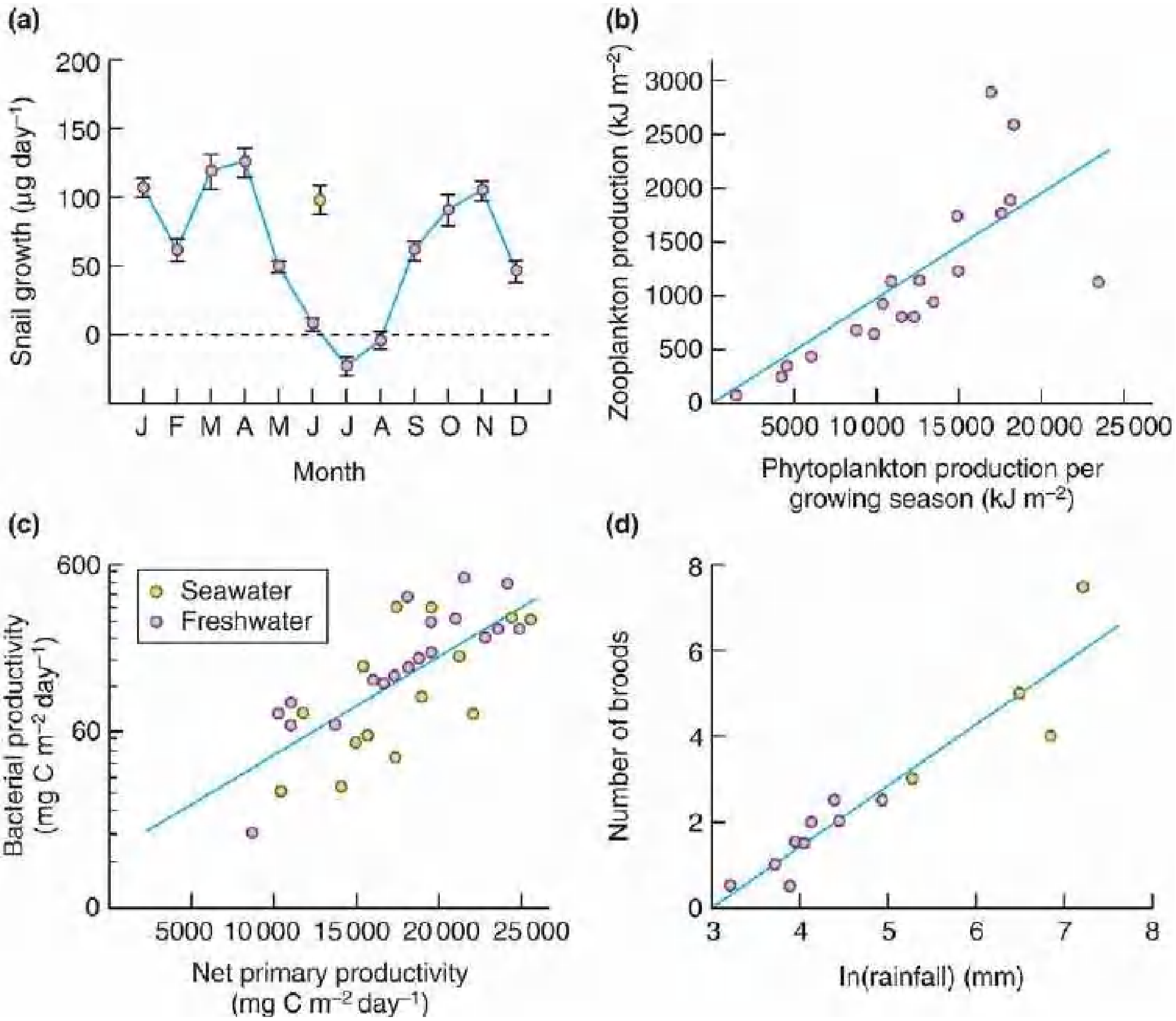


Figure 20.21 **Secondary productivity depends on primary productivity.** (a) Seasonal pattern of snail growth (mean increase in weight of individually marked snails during a month on the stream bed ± SE). The green circle represents growth at a nearby unshaded stream site in June. (b) Relationship between primary and secondary productivity for zooplankton in lakes. (c) Relationship between bacterial and phytoplankton productivity in freshwater and seawater. (d) A *Geospiza fortis* female is capable of producing up to 10 broods in a single year (each with 2–6 eggs), but the mean number of broods varies and is positively related to annual rainfall (and, thus, to primary productivity); the green circles are for particularly wet years when El Niño weather events occurred.

Source: (a) After Hill *et al.* (2001). (b) After Brylinsky & Mann (1973). (c) After Cole *et al.* (1988). (d) After Grant *et al.* (2000).

Eltonian pyramids of productivity and biomass in successive trophic levels

A general rule in both aquatic and terrestrial ecosystems is that secondary productivity by herbivores is approximately an order of magnitude less than the primary productivity on which it is based. This is a consistent feature of all grazer systems: that part of the trophic structure of a community that depends, at its base, on the consumption of *living* plant biomass (in the ecosystem context we use 'grazer' in a different sense to its definition in Chapter 9). It results in a pyramidal structure in which the productivity of plants provides a broad base on which a smaller productivity of primary consumers depends, with a still smaller productivity of secondary consumers above that. Trophic levels may also have a pyramidal structure when expressed in terms of biomass. Elton (1927) was the first to recognise this fundamental feature of community architecture, an idea that was later elaborated by Lindemann (1942).

another fundamental link between community and ecosystem ecology

Herbivore consumption has been found to scale nearly linearly with primary production in both terrestrial and aquatic ecosystems (Figure 20.22a). Thus, in different examples of a given ecosystem type, a more or less fixed proportion is consumed by herbivores irrespective of the level of plant productivity. But a markedly different pattern occurs in transfer from the second (herbivores) to the third (carnivores) trophic levels, where the proportion of prey consumed decreases as prey biomass increases. For example, the biomass of large carnivores (including lions, cheetahs, hyenas, etc.) in African savannas increases with an increase in herbivore prey biomass (including everything from dik-dik to buffalo), but less than proportionately (Figure 20.22b). Thus, from the least productive (dry Kalahari Desert) to the most productive (Ngorongoro Crater) sites there is three-fold less predator biomass per kilogram of prey. Put another way, the pyramid of predators (above) to prey (below) gets more bottom heavy as prey biomass increases. In their impressive meta-analysis, Hatton *et al.* (2015) show further that this same predator–prey relationship holds for other large carnivore–herbivore examples as well as for freshwater and marine studies involving zooplankton feeding on phytoplankton. This is a remarkably general finding. The explanation appears to be that where prey are abundant, density-dependent processes act to reduce overall prey productivity (e.g. lower rates of prey growth, nutritional quality, etc.) while other density-dependent factors come into play among the predators, such as mortality from disease and territoriality. The density-dependent processes discussed in earlier chapters dealing with populations seem, therefore, to play out also when a diverse array of populations are lumped together in the ecosystem context, providing a further important link between the community and ecosystem perspectives of ecology.

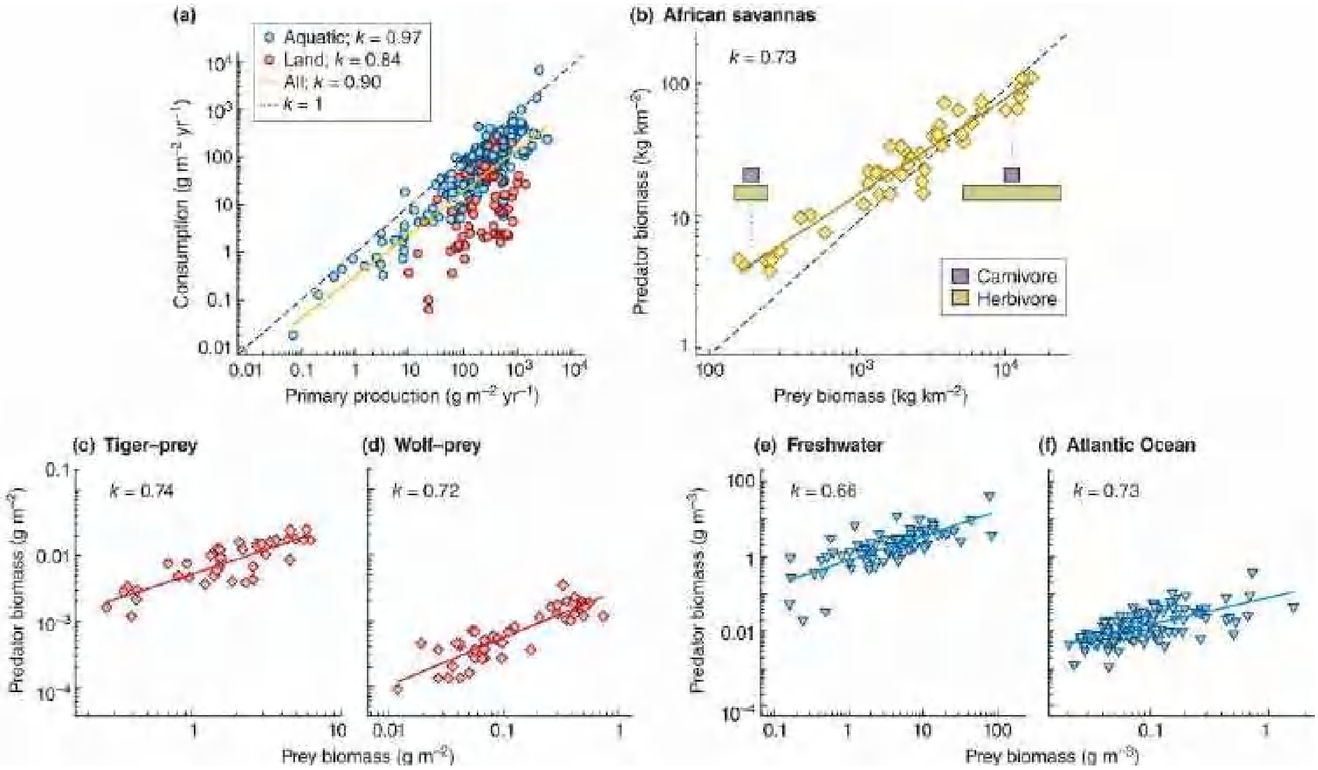


Figure 20.22 Relationships between herbivore consumption and primary production and between carnivore biomass and prey biomass. (a) Relationship between consumption by herbivores and primary production in a wide range of aquatic and terrestrial ecosystems worldwide. Plotted on logarithmic axes, the slope of the relationship *k* is close to 1, meaning that a more or less fixed proportion of plant production is consumed by herbivores irrespective of productivity level. The dashed line shows the relationship for *k* = 1. (b) The relationship between the biomasses of large carnivores and their herbivore prey in African savannas. Carnivore biomass increases with prey biomass but less than proportionately (*k* = 0.73), so that at high prey biomass the biomass pyramid is more bottom heavy. (c, d) The same relationship for other large carnivore–prey examples and (e, f) for freshwater and marine examples (zooplankton feeding on phytoplankton). In all cases, predator–prey biomass scaling is significantly less than *k* = 1, being approximately *k* = 0.75 or less. Compiled from more than 1000 studies by Hatton *et al.* (2015).

most primary productivity does not pass through the grazer system

The productivity at a trophic level is invariably less than that of the trophic level below on which they feed. Where has the missing energy gone? First, not all of the plant biomass produced is consumed alive by herbivores. Much dies without being grazed and supports the decomposer community (bacteria, archaea, fungi and detritivorous animals). Second, not all plant biomass eaten by herbivores (nor herbivore biomass eaten by carnivores) is assimilated and available for incorporation into consumer biomass. Some is lost in faeces, and this also passes to the decomposers. Third, not all energy that has been assimilated is actually converted to biomass. A proportion is lost as respiratory heat. This occurs both because no energy conversion process is ever 100% efficient (some is lost as unusable random heat, consistent with the second law of thermodynamics) and also because animals do work that requires energy, again released as heat. These three energy pathways occur at all trophic levels and are illustrated in Figure 20.23a.

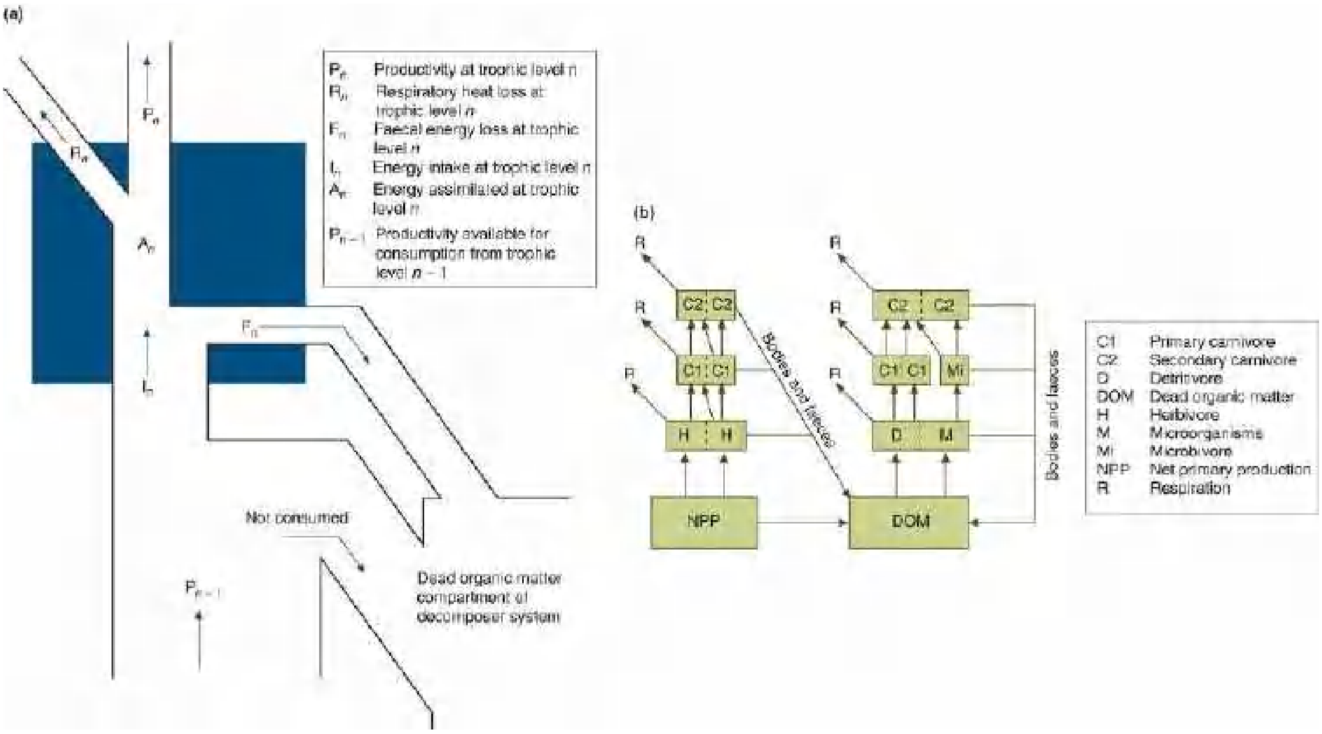


Figure 20.23 **Patterns in energy flow.** (a) Energy flow through a trophic compartment. The blue square represents the biomass of the compartment. (b) A generalised model of trophic structure and energy flow through a food web. Note that the decomposer and grazer systems in (b) do not remain distinct through the higher trophic levels – the same carnivores may consume both herbivores and detritivores.

Source: (b) After Heal & MacLean (1975).

20.5.2 Possible pathways of energy flow through a food web

alternative pathways that energy can trace through the community

Figure 20.23b provides a complete description of the trophic structure of a community. It consists of the grazer system pyramid of productivity, but with two additional elements of realism. Most importantly, it adds a *decomposer system* – this is invariably coupled to the grazer system in communities. Secondly, it recognises that there are subcomponents of each trophic level in each subsystem that operate in different ways. Thus a distinction is made between microbes and detritivores that occupy the same trophic level and utilise dead organic matter, and between consumers of microbes (microbivores) and of detritivores. Figure 20.23b, then, displays the possible routes that a joule of energy, fixed in net primary production, can take as it is dissipated on its path through the community. A joule of energy may be consumed and assimilated by a herbivore that uses part of it to do work and loses it as respiratory heat. Or it might be consumed by a herbivore and later assimilated by a carnivore that dies and enters the dead organic matter compartment. Here, what remains of the joule may be assimilated by a fungal hypha, then consumed by a soil mite, which uses it to do work, dissipating a further part of the joule as heat. At each consumption step, what remains of the joule may fail to be assimilated and may pass in the faeces to be dead organic matter, or it may be assimilated and respired, or assimilated and incorporated into the growth of body tissue (or the production of offspring – as in the case of broods of the bird in Figure 20.21d). The body may die and what remains of the joule may then enter the dead organic matter compartment, or it may be captured alive by a consumer in the next trophic level where it meets a further set of possible branching pathways. Ultimately, each joule will have found its way out of the community, dissipated as respiratory heat at one or more of the transitions in its path along the food chain. Whereas a molecule or ion may cycle endlessly through the food chains of a community, energy passes through just once.

The possible pathways in the grazer and decomposer systems are the same, with one critical exception – faeces and dead bodies are lost to the grazer system (and enter the decomposer system), but faeces and dead bodies from the decomposer system are simply sent back to the dead organic matter compartment at its base. This has a fundamental significance. The energy available as dead organic matter may finally be completely metabolised – and all the energy lost as respiratory heat – even if this requires several circuits through the decomposer system. The exceptions to this are situations where: (i) matter is exported out of the local environment to be metabolised elsewhere, for example detritus washed out of a stream; and (ii) local abiotic conditions are very unfavourable to decomposition processes, leaving pockets of incompletely metabolised high-energy matter, otherwise known as oil, coal and peat.

20.5.3 The importance of transfer efficiencies in determining energy pathways

> the relative importance of energy pathways depends on three transfer efficiencies: ...

The proportions of NPP that flow along each of the possible energy pathways depend on *transfer efficiencies* in the way energy is used and passed from one step to the next. A knowledge of the values of just three categories of transfer efficiency is all that is required to predict the pattern of energy flow. These are *consumption efficiency* (CE), *assimilation efficiency* (AE) and *production efficiency* (PE).

> ... consumption efficiency, ...

Consumption efficiency,

$$CE = I_n / P_{n-1} \times 100.$$

Repeated in words, consumption efficiency is the percentage of total productivity available at one trophic level (P_{n-1}) that is actually consumed ('ingested') by a trophic compartment 'one level up' (I_n). For primary consumers in the grazer system, consumption efficiency is the percentage of joules produced per unit time as NPP that finds its way into the guts of herbivores. In the case of secondary consumers, it is the percentage of herbivore productivity eaten by carnivores. The remainder dies without being eaten and enters the decomposer chain.

Various reported values for the consumption efficiencies of herbivores are shown in Figure 20.24. Most of the estimates are remarkably low, usually reflecting the unattractiveness of much plant material because of its high proportion of structural support tissue, but sometimes also as a consequence of generally low herbivore densities (because of the action of their natural enemies). The consumers of microscopic plants (microalgae growing on beds or free-living phytoplankton) can achieve greater densities and account for a greater percentage of primary production, because they have less structural tissue to deal with. Median values for consumption efficiency are less than 5% in forests, around 25% in grasslands and more than 50% in phytoplankton-dominated communities. We know much less about the consumption efficiencies of carnivores feeding on their prey, and any estimates are speculative. Vertebrate predators may consume 50–100% of production from vertebrate prey but perhaps only 5% from invertebrate prey. Invertebrate predators consume perhaps 25% of available invertebrate prey production.

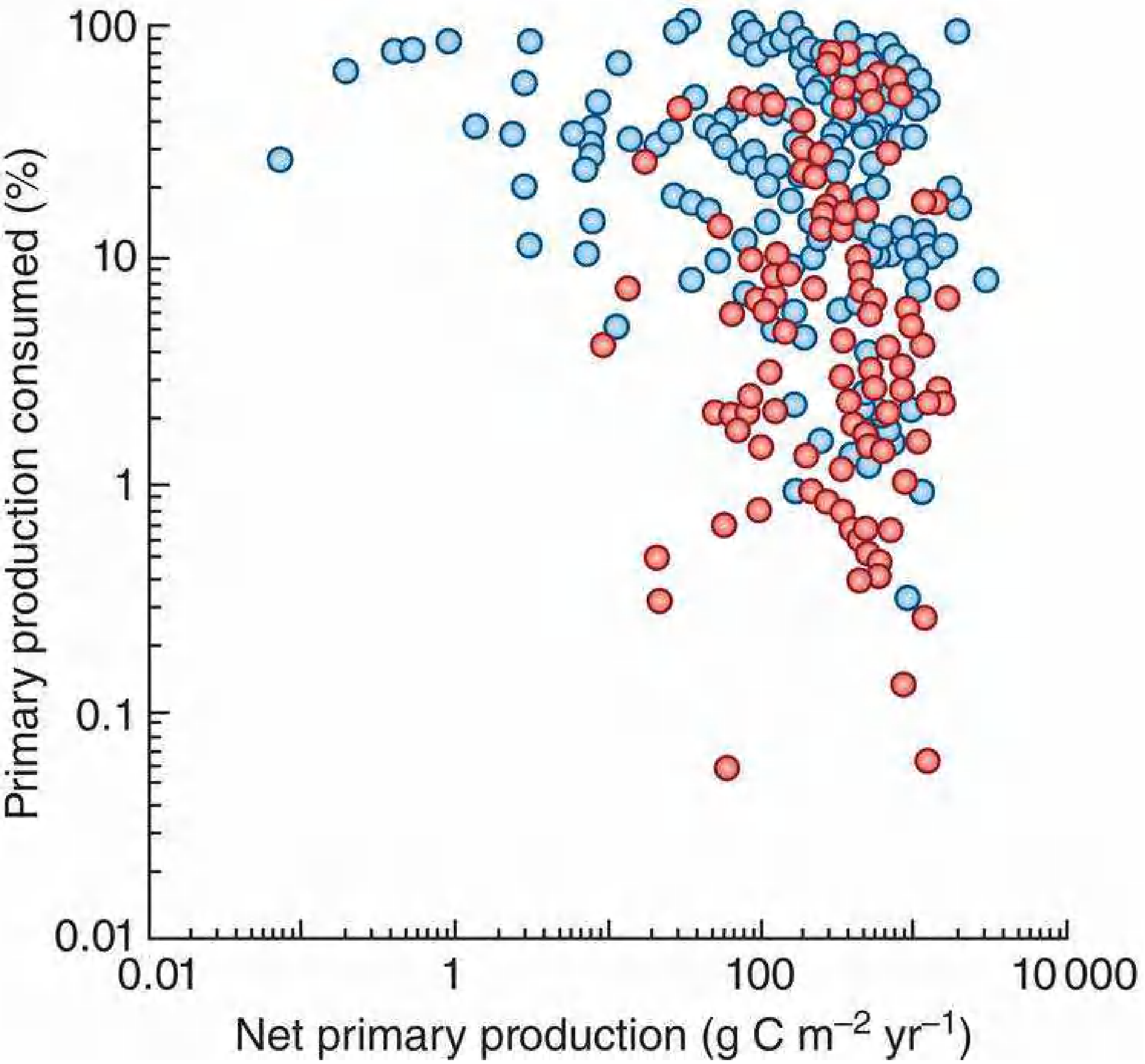


Figure 20.24 The percentage of net primary production (NPP) consumed by herbivores in relation to NPP in aquatic (blue circles) and terrestrial ecosystems (red circles). Compiled from data from a number of sources by Cebrian and Lartigue (2004).

Source: Cebrian & Lartigue (2004).

... assimilation efficiency ...

Assimilation efficiency,

$$AE = A_n / I_n \times 100.$$

Assimilation efficiency is the percentage of food energy taken into the guts of consumers in a trophic compartment (I_n) that is assimilated across the gut wall (A_n) and becomes available for incorporation into growth or to do work. The remainder is lost as faeces and enters the base of the decomposer system. An ‘assimilation efficiency’ is much less easily ascribed to microorganisms. Food does not enter an invagination of the outside world passing through the microorganism’s body (like the gut of a higher organism) and faeces are not produced. If bacteria and fungi effectively absorb 100% of the dead organic matter they digest externally, they would have an ‘assimilation efficiency’ of 100%, though of course they have paid a price for this in the form of the enzymes they have secreted into the environment.

Assimilation efficiencies are typically low for herbivores, detritivores and microbivores (20–50%) and high for carnivores (80% or more). In general, animals are poorly equipped to deal with dead organic matter (mainly plant material) and living vegetation, no doubt partly because of the very widespread occurrence of physical and chemical plant defenses, but mainly as a result of the high proportion of complex structural chemicals such as cellulose and lignin in their make-up. As Chapter 11 describes, however, many animals contain a symbiotic gut microflora that produces cellulase and aids in the assimilation of plant organic matter. In one sense, these animals have harnessed their own personal decomposer system. The way that plants allocate production to roots, wood, leaves, seeds and fruits influences their usefulness to herbivores. Seeds and fruits may be assimilated with efficiencies as high as 60–70%, and leaves with about 50% efficiency, while the assimilation efficiency for wood may be as low as 15%. The animal food of carnivores, detritivores such as vultures that consume animal carcasses, parasites and parasitoids poses less of a problem for digestion and assimilation. Indeed, Sanders *et al*. (2016) reported that the assimilation efficiency of the hyperparasitoid *Dendrocerus carpenteri* is higher (93%) when acting as a tertiary parasitoid (feeding on a parasitoid of a parasitoid of aphids) than as a secondary parasitoid (65%; feeding on a parasitoid of aphids), perhaps because nitrogen tends to be more concentrated at higher trophic levels and plant chemical defences have been removed further down the chain.

... and production efficiency ...

Production efficiency,

$$PE = P_n / A_n \times 100.$$

Production efficiency is the percentage of assimilated energy (A_n) that is incorporated into new biomass (P_n). The remainder is entirely lost to the community as respiratory heat. (Energy-rich secretory and excretory products, which have taken part in metabolic processes, may be viewed as production, P_n, and become available, like dead bodies, to the decomposers.)

Production efficiency varies mainly according to the metabolic type and size of the organisms concerned. Microorganisms, including protozoa, tend to have very high production efficiencies. They have short lives, small size and rapid population turnover. Invertebrates in general have high efficiencies (30–40%), losing relatively little energy in respiratory heat and converting more assimilate to production. Amongst the vertebrates, ectotherms (whose body temperature varies according to environmental temperature) have intermediate values for production efficiency (around 10%), whilst endotherms, with their high energy expenditure associated with maintaining a constant temperature, convert only 1–2% of assimilated energy into production. In general, efficiency of production increases with size in endotherms (since the smaller surface area to volume ratios of larger animals mean that they lose less of the heat they produce), whereas efficiency decreases very markedly in ectotherms (since those ratios mean that larger animals struggle to *absorb* enough heat). Hence, the small-bodied endotherms have the lowest efficiencies, with tiny insectivores (e.g. wrens and shrews) having the lowest production efficiencies of all.

... which combine to give trophic-level transfer efficiency

Trophic-level transfer efficiency,

$$\mathrm{TLTE} = P_n / P_{n-1} \times 100.$$

The overall trophic transfer efficiency from one trophic level to the next is simply CE × AE × PE. In the period after Lindemann's (1942) pioneering work, it was generally assumed that trophic transfer efficiencies were around 10%; indeed some ecologists referred to a 10% 'law'. However, while it is a reasonable 'rule of thumb', there is certainly no law of nature that results in precisely one-tenth of the energy that enters a trophic level transferring to the next. For example, a compilation of trophic studies from a wide range of freshwater and marine environments revealed that trophic-level transfer efficiencies varied between about 2% and 24%, although the mean was 10.13% (Figure 20.25). Transfer efficiencies in marine food webs tend to decline at higher trophic levels, with a mean of 13% from phytoplankton to zooplankton or benthic invertebrates, and 10% from zooplankton or benthic invertebrates to fish (Ware, 2000).

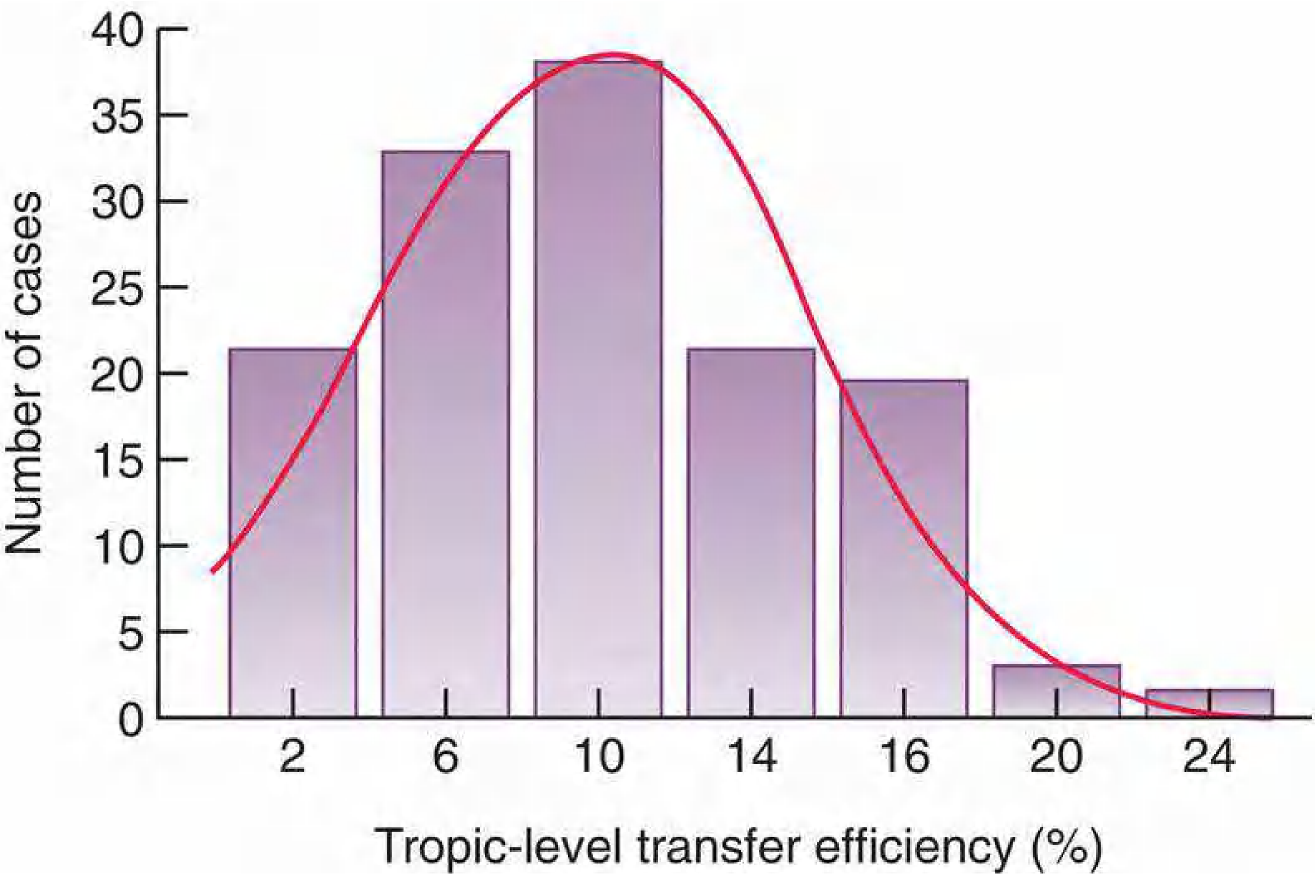


Figure 20.25 Frequency distribution of trophic-level transfer efficiencies in 48 trophic studies of aquatic communities. There is considerable variation among studies and among trophic levels. The mean is 10.13 % (SE = 0.49).

Source: After Pauly & Christensen (1995).

20.5.4 Energy flow: spatial and temporal variation

relative roles of grazer and decomposer systems in contrasting communities

Given accurate values for NPP in an ecosystem, and consumption, assimilation and production efficiencies for the various trophic groupings shown in the model in Figure 20.23b, it should be possible to predict and understand the relative importance of the different possible energy pathways. Perhaps not surprisingly, no study has incorporated all ecosystem compartments and all transfer efficiencies of the component species. However, some generalisations are possible when the gross features of contrasting systems are compared (Figure 20.26). First, the decomposer system is probably responsible for the majority of secondary production, and therefore respiratory heat loss, in every community in the world. The grazer system has its greatest role in plankton communities, where a large proportion of NPP is consumed alive and assimilated at quite a high efficiency. Even here, though, it is now clear that very high densities of heterotrophic bacteria in the plankton community subsist on dissolved organic molecules excreted by phytoplankton cells, perhaps consuming more than 50% of primary productivity as 'dead' organic matter in this way (Fenchel, 1987). The grazer system is highly variable in terrestrial communities because of low herbivore consumption and assimilation efficiencies in forests, but much higher consumption in grasslands, particularly where large herds of grazers still persist. Grazing contributes little to total energy flow in many small streams and ponds simply because primary productivity is so low. These aquatic systems depend for their energy base on dead organic matter produced in the terrestrial environment that falls or is washed or blown into the water (allochthonous production). The deep-ocean benthic community has a trophic structure very similar to that of streams and ponds (all can be described as heterotrophic communities), but even more extreme. In this case, the community lives in water too deep for photosynthesis to take place at all, but it derives its energy base from dead phytoplankton, bacteria, animals and faeces that sink from the autotrophic community in the euphotic zone above. From a different perspective, the ocean bed is equivalent to a forest floor beneath an impenetrable forest canopy.

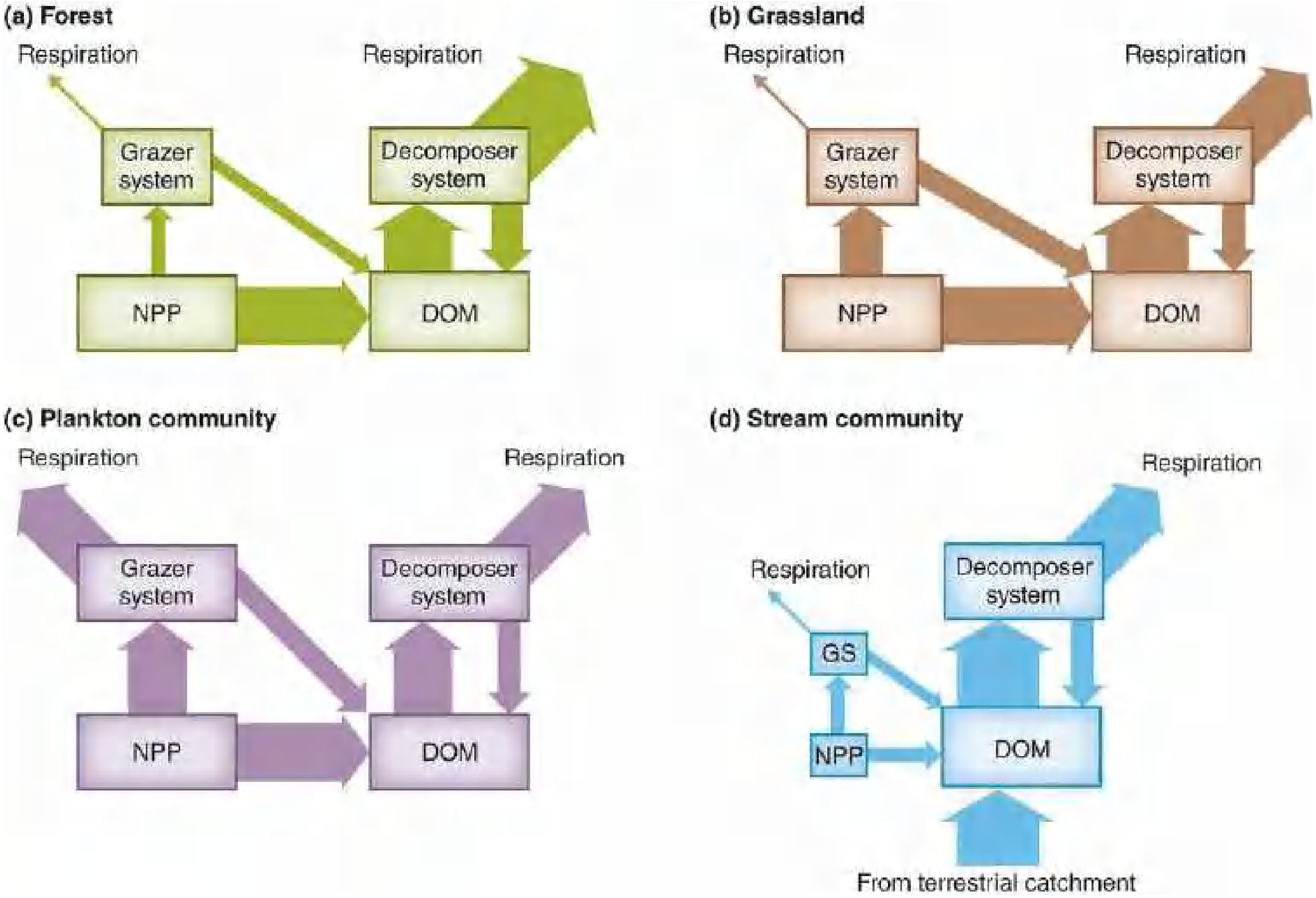


Figure 20.26 General patterns of energy flow. Energy flow for (a) a forest, (b) a grassland, (c) a marine plankton community, and (d) the community of a stream or small pond. The relative sizes of the boxes and arrows are proportional to the relative magnitudes of compartments and flows. DOM, dead organic matter; GS, grazer system; NPP, net primary production.

grazer consumption efficiencies are highest where plants have low C : N and C : P ratios

We can move from the relatively gross generalisations above to consider in Figure 20.27 a greater range of terrestrial and aquatic ecosystems (data compiled from over 200 published reports by Cebrian (1999)). Figure 20.27a first shows the range of values for NPP in a variety of terrestrial and aquatic ecosystems. Figure 20.27b re-emphasises how consumption efficiency by grazers is particularly low in ecosystems where plant biomass contains considerable support tissue and relatively low amounts of nitrogen and phosphorus (i.e. forests, shrublands and mangroves). Plant biomass not consumed by herbivores becomes detritus and contributes by far the largest proportion to the dead organic matter box in Figure 20.23b. Not surprisingly, the percentage of NPP destined to be detritus is highest in forests and lowest in phytoplankton and benthic microalgal communities (Figure 20.27c). Plant biomass from terrestrial communities is not only unpalatable to herbivores, it is also relatively more difficult for decomposers and detritivores to deal with. Thus, Figure 20.27d shows that a greater proportion of primary production accumulates as refractory detritus (persisting for more than a year) in forests, shrublands, grasslands and freshwater macrophyte meadows. Finally, Figure 20.27e shows the percentage of NPP that is exported out of the systems. The values are generally modest (medians of 20% or less) indicating that, in most cases, the majority of biomass produced in an ecosystem is consumed or decomposed there. The most obvious exceptions are mangroves and, in particular, macroalgal beds (which often inhabit rocky shores), where relatively large proportions of plant biomass are displaced and moved away by storm and tidal action.

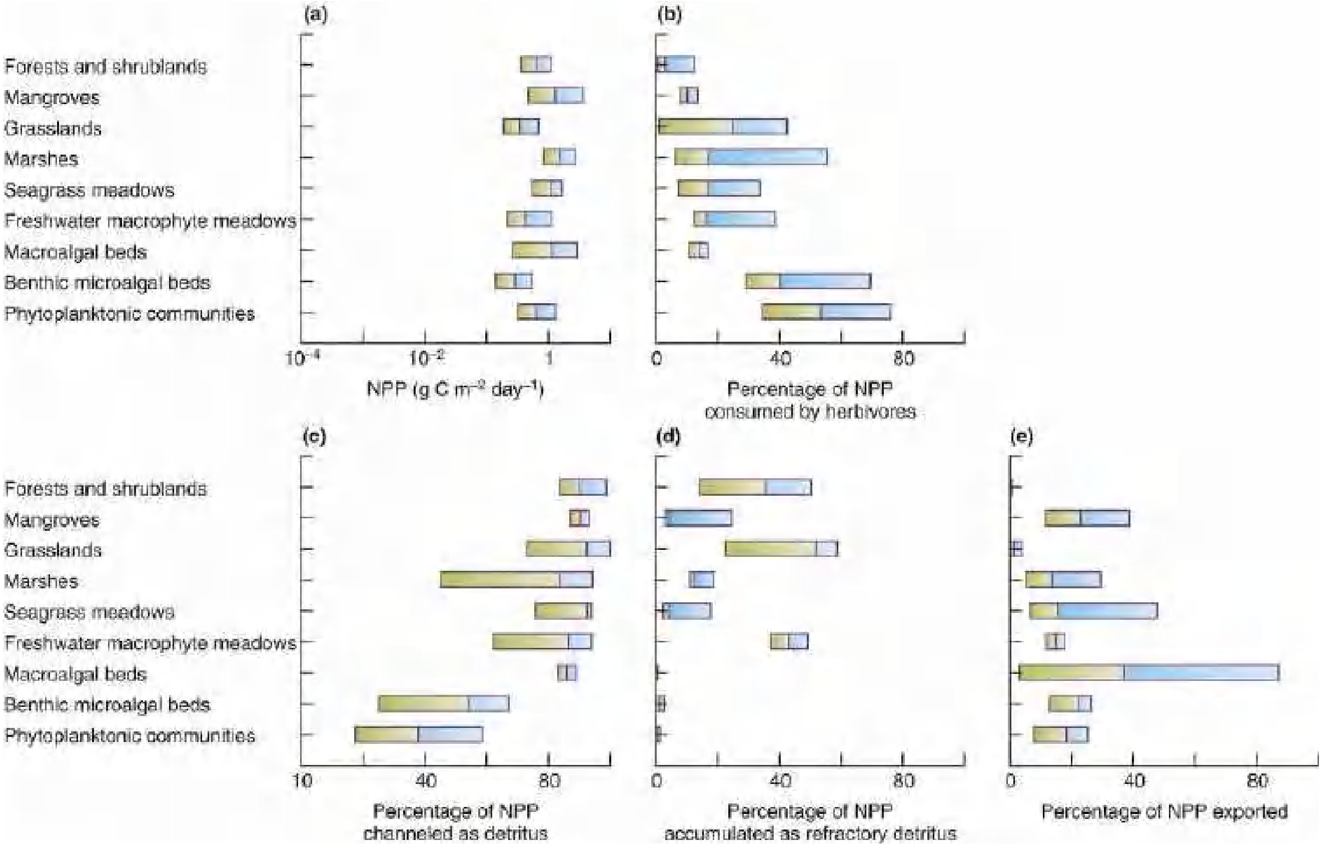


Figure 20.27 Patterns in the fate of NPP. Box plots showing for a range of ecosystem types: (a) net primary productivity (NPP), (b) percentage of NPP consumed by detritivores, (c) percentage of NPP channelled as detritus, (d) percentage of NPP accumulated as refractory detritus, and (e) percentage of NPP exported. Boxes encompass 25% and 75% quartiles and the central lines represent the median of a number of studies (with blue shading above the median, and green below, for clarity).

Source: After Cebrian (1999). (Also see Cebrian & Lartigue (2004).)

In general then, communities composed of plants whose stoichiometry represents a higher nutritional status (higher nitrogen and phosphorus concentrations, i.e. lower C : N and C : P) lose a higher percentage to herbivores, produce a smaller proportion of detritus, experience faster decomposition rates and, in consequence, accumulate less refractory detritus and have smaller stores of dead organic carbon (Cebrian & Lartigue, 2004).

temporal patterns in the balance between production and consumption of organic matter

The presentation of information in Figure 20.27 emphasises spatial patterns in the way energy moves through the world's ecosystems. However, we should not lose sight of the temporal patterns that exist in the balance between production and consumption of organic matter. Figure 20.28 shows how GPP, RE (the sum of autotrophic and heterotrophic respiration) and net ecosystem productivity (NEP, which equals GPP – RE) varied seasonally during five years of study of a boreal aspen (*Populus tremuloides*) forest in Canada. Note how NEP is negative (RE exceeds GPP and carbon stores are being used by the community) except in the summer months when GPP consistently exceeds RE. At this site, the cumulative annual values for NEP were always positive, indicating that more carbon is fixed than is respired each year and the forest is a carbon sink. However, this is not true for all ecosystems every year (Falge *et al.*, 2002).

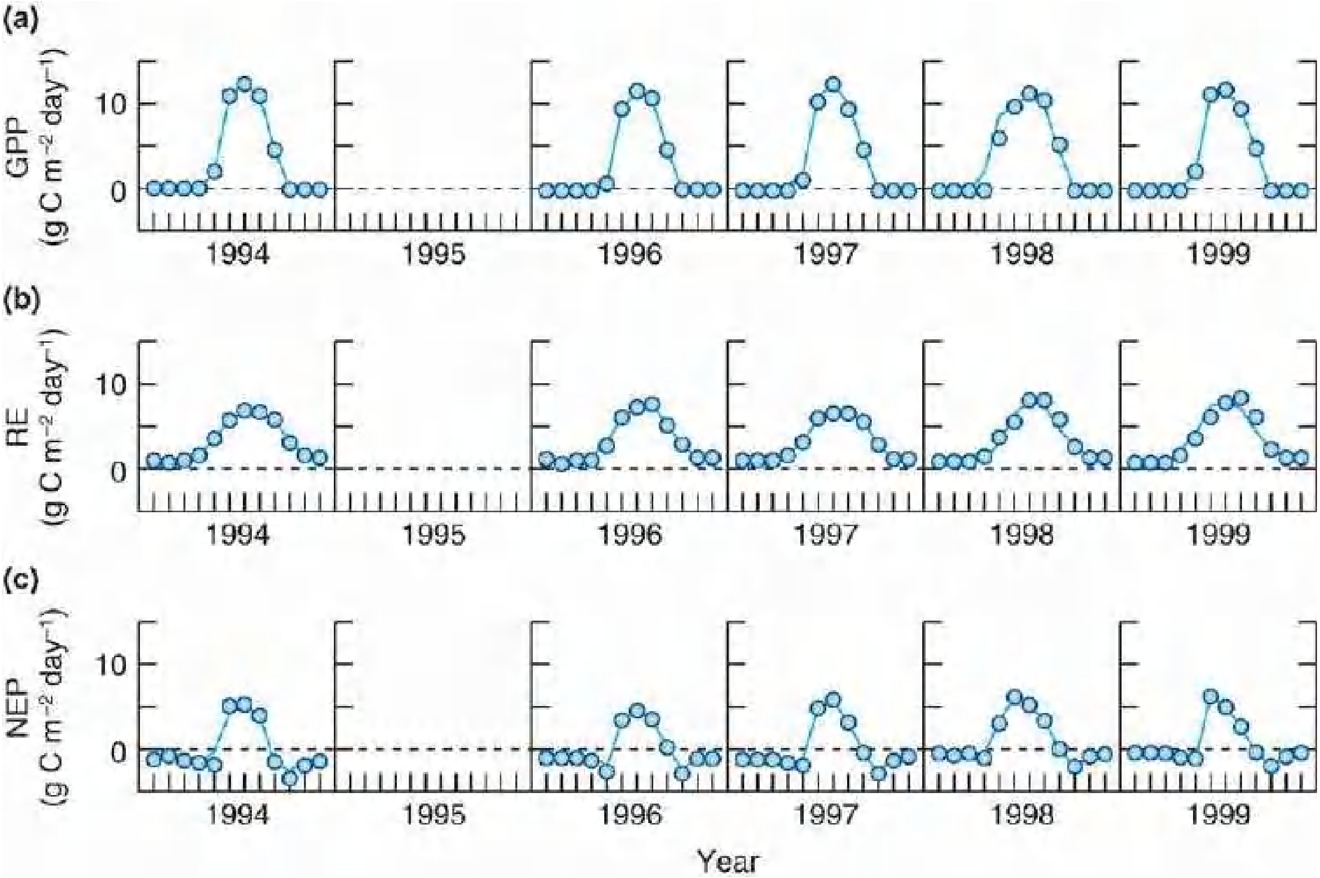


Figure 20.28 **Temporal patterns in ecosystem energetics**. Monthly mean values for: (a) gross primary productivity (GPP), (b) ecosystem respiration (RE), and (c) net ecosystem productivity (NEP) in a Canadian aspen forest. (No data were collected in 1995.)

Source: After Arain *et al.* (2002).

We also see interannual variation in these ecosystem fluxes, probably driven by variation in climate. Total annual GPP (the area under the GPP curves in Figure 20.28a) was highest in 1998 when the temperature was high (probably the result of an El Niño event – see below) and lowest in 1996 when the growing season temperatures were particularly low. Annual variations in GPP (e.g. 1419 g C m^{-2} in 1998, 1187 g C m^{-2} in 1996) were large compared with variations in RE (1132 g C m^{-2} and 1106 g C m^{-2}, respectively) because the occurrence of warm spring weather caused photosynthesis to increase faster than respiration. This led overall to higher values of NEP in warmer years (290 g C m^{-2} in 1998, 80 g C m^{-2} in 1996).

consequences of the ENSO for ecosystem energetics, an example of longer term cyclic climatic variation

This aspen forest is by no means the only ecosystem where annual variations in climate drive annual variations in energy flux. As noted earlier, some of this variation results from climatic cycles, such as the El Niño–Southern Oscillation (ENSO; see also Section 2.4.1). ENSO events happen sporadically but typically occur every 3–6 years. During such events, the temperature may be significantly higher in some locations and lower in others and, just as significantly, rainfall can be 4–10 times higher in some areas. El Niño correlates with dramatic changes in aquatic ecosystems (even leading to the collapse of fisheries; Jordan, 1991). Recent evidence shows that El Niño events can cause major changes on land too. Figure 20.29 shows the annual variation in caterpillar numbers on the Galápagos Islands in a standard census conducted in various years since 1977, plotted on the same graph as annual rainfall. The remarkably strong correlation comes about because of the dependence of caterpillar numbers on primary productivity, which itself is considerably higher in wet, ENSO years. We saw in Figure 20.21d how the total number of broods of the finch *Geospiza fortis* was much greater in the four ENSO years (open circles in that figure). This reflects the much greater production in very wet years of the seeds, fruits and caterpillars that they feed on. Not only do the finches increase the number of broods, but also the size of their clutches (from a mean of three to four eggs per brood) and the probability of successful rearing to the fledging stage.

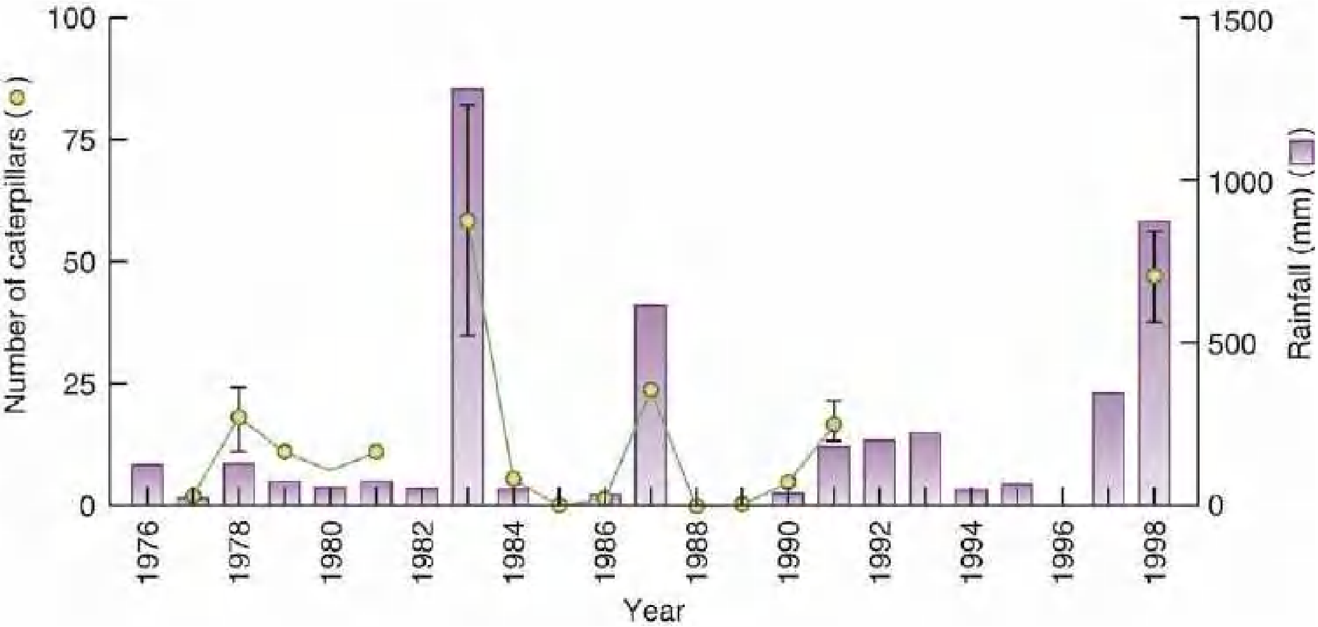


Figure 20.29 Galápagos caterpillars are more abundant in wet years. Annual variation in mean caterpillar numbers (± SE; green circles) in a standard census against a histogram of annual rainfall on the Galápagos island of Daphne Major.

Source: After Grant *et al.* (2000).

Our growing knowledge of the impact of ENSO events on energy flux through ecosystems suggests that the predicted directional changes (as opposed to cyclic variation) in extreme weather events expected as a result of human-induced global climate change will profoundly alter ecosystem processes in many parts of the world, a topic to which we will return in Chapter 22.

But next we turn to the flux of matter through ecosystems, recognising that the rate at which resources are supplied and used by autotrophs and heterotrophs depends fundamentally on the supply of nutrients (Chapter 21).

Chapter 21

The Flux of Matter through Ecosystems

21.1 Introduction

Chemical elements and compounds are vital for the processes of life. Living organisms expend energy to extract chemicals from their environment, they hold on to them and use them for a period, then lose them again. Thus, the activities of organisms profoundly influence the patterns of flux of chemical matter in the biosphere. Physiological ecologists focus their attention on how individual organisms obtain and use the chemicals they need (see Chapter 3). However, in this chapter, as in the last, we change the emphasis and consider the ways in which the biota on an area of land, or within a volume of water, accumulates, transforms and moves matter between the various components of the ecosystem. The area that we choose may be that of the whole globe, a continent, a river catchment or simply a square metre.

21.1.1 Relationships between energy flux and nutrient cycling

The great bulk of living matter in any community is water. The rest is made up mainly of carbon compounds (95% or more) and this is the form in which energy is accumulated and stored. The energy is ultimately dissipated when the carbon compounds are oxidised to carbon dioxide (CO_2) by the metabolism of living tissue or of its decomposers. Although we consider the fluxes of energy and carbon in different chapters, the two are intimately bound together in all biological systems.

Carbon enters the trophic structure of a community when a simple molecule, CO_2, is taken up in photosynthesis. If it becomes incorporated in net primary productivity, it is available for consumption as part of a molecule of sugar, fat, protein or, very often, cellulose. It follows exactly the same route as energy, being successively consumed, defaecated, assimilated and perhaps incorporated into secondary productivity somewhere within one of the trophic compartments. When the high-energy molecule in which the carbon is resident is finally used to provide energy for work, the energy is dissipated as heat (as we have discussed in Chapter 20) and the carbon is released again to the atmosphere as CO_2. Here, the tight link between energy and carbon ends.

energy cannot be cycled and reused; matter can …

Once energy is transformed into heat, it can no longer be used by living organisms to do work or to fuel the synthesis of biomass. (Its only possible role is momentary, in helping to maintain a high body temperature, with incidental consequences for temperature-dependent physiological processes.) The heat is eventually lost to the atmosphere and can never be recycled. In contrast, the carbon in CO_2 can be used again in photosynthesis. Carbon, and all nutrient elements (nitrogen, phosphorus, etc.) are available to plants as simple inorganic molecules or ions in the atmosphere (CO_2), or as dissolved ions in water (nitrate, phosphate, potassium, etc.). Each can be incorporated into complex organic carbon compounds in biomass. Ultimately, however, when the carbon compounds are metabolised to CO_2, the mineral nutrients are released again in simple inorganic form. Another plant may then absorb them, and so an individual atom of a nutrient element may pass repeatedly through one food chain after another. The relationship between energy flow and nutrient cycling is illustrated in Figure 21.1.

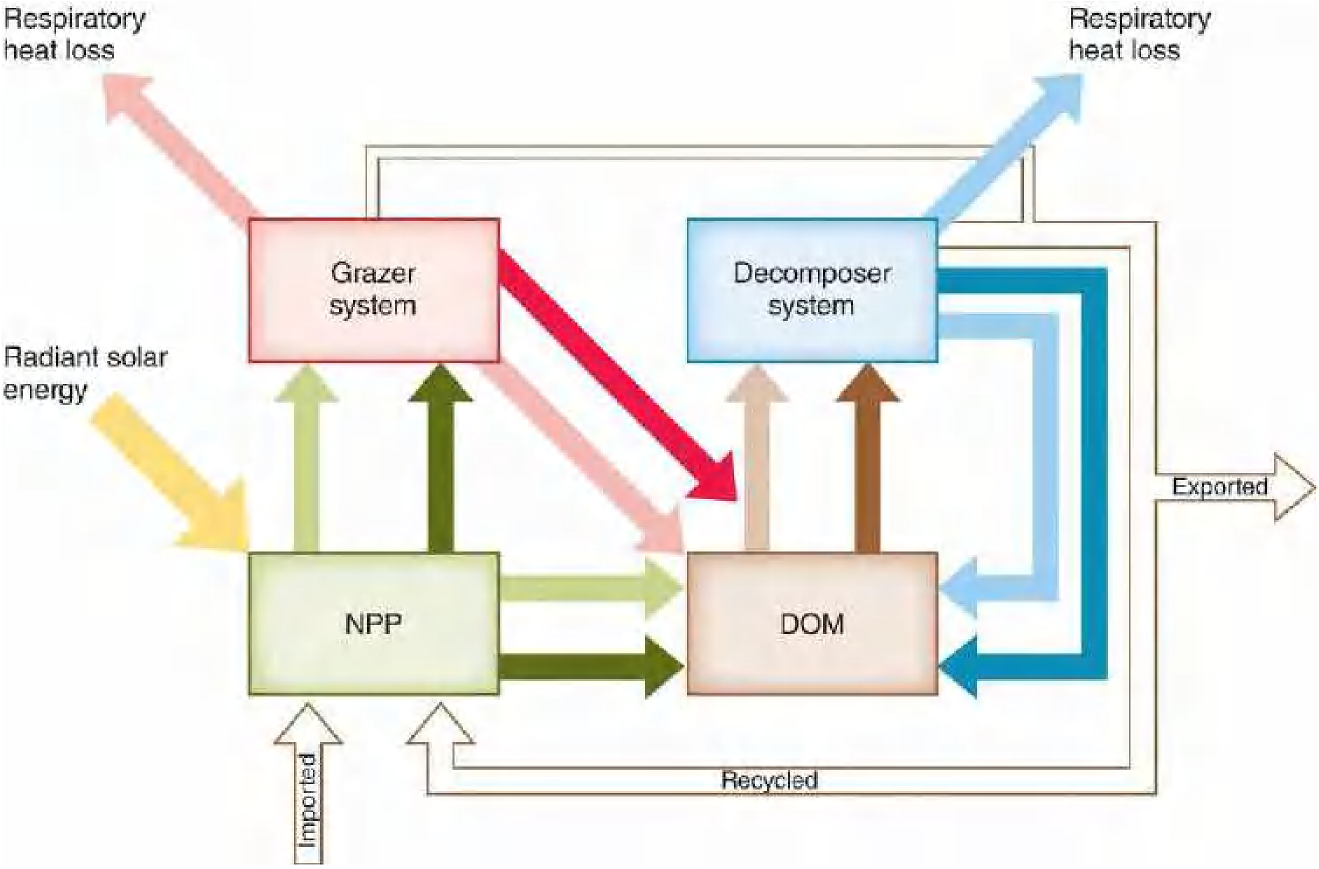


Figure 21.1 The relationship between energy flow and nutrient flux. Nutrients locked in organic matter (dark arrows) are distinguished from the free inorganic state (white arrows). Energy flow is depicted by pale arrows. Energy cannot be recycled but nutrients can. However, nutrient cycling is never perfect – some of the inorganic nutrient budget is exported from the ecosystem while further supplies may be imported from outside. DOM, dead organic matter; NPP, net primary production.

By its very nature, then, each joule of energy can be *used* only once, whereas chemical nutrients, the building blocks of biomass, simply change the form of molecule of which they are part (e.g. nitrate-N to protein-N to nitrate-N). They can be used again, and repeatedly recycled. Unlike the energy of solar radiation, nutrients are not in unalterable supply, and the process of locking some up into living biomass reduces the supply remaining to the rest of the community. If plants, and their consumers, were not eventually decomposed, the supply of nutrients would become exhausted and life on the planet would cease. The activity of heterotrophic organisms is crucial in bringing about nutrient cycling and maintaining productivity. The decomposer system plays an overwhelming role in the release of nutrients in their simple inorganic form, but some is also released from the grazer system (carbon in CO_2 respired by herbivores, carnivores and parasites).

... but nutrient cycling is never perfect

Not all nutrients released during decomposition are necessarily taken up again by plants. Nutrient recycling is never perfect and some nutrients are exported from land by runoff into streams (ultimately to lakes and oceans) and others, such as nitrogen and sulphur, that have gaseous phases, can be lost to the atmosphere. Moreover, a community receives additional supplies of nutrients that do not depend directly on inputs from recently decomposed matter – minerals dissolved in rain, for example, or derived from weathered rock.

21.1.2 Biogeochemistry and biogeochemical cycles

the ‘bio’ in biogeochemistry

We can conceive of pools of chemical elements existing in compartments. Some compartments occur in the *atmosphere* (carbon in CO_2, nitrogen as gaseous nitrogen, etc.), some in the rocks of the *lithosphere* (calcium as a constituent of calcium carbonate, potassium in feldspar) and others in the *hydrosphere* – the water in soil, streams, lakes or oceans (nitrogen in dissolved nitrate, phosphorus in phosphate, carbon in carbonic acid, etc.). In all these cases the elements exist in an inorganic form. In contrast, living organisms (the biota) and dead and decaying bodies can be viewed as compartments containing elements in an organic form (carbon in cellulose or fat, nitrogen in protein, phosphorus in adenosine triphosphate, etc.). Studies of the chemical processes occurring within these compartments and, more particularly, of the fluxes of elements between them, comprise the science of biogeochemistry.

Many geochemical fluxes would occur in the absence of life, if only because all geological formations above sea level are eroding and degrading. Volcanoes release sulphur into the atmosphere whether there are organisms present or not. On the other hand, organisms alter the rate of flux and the differential flux of the elements by extracting and recycling some chemicals from the underlying geochemical flow. The term biogeochemistry is apt.

biogeochemistry can be studied at different scales

The flux of matter can be investigated at a variety of spatial and temporal scales. Ecologists interested in the gains, uses and losses of nutrients by the community of a small pond or a hectare of grassland can focus on local pools of chemicals. They need not concern themselves with the contribution to the nutrient budget made by volcanoes or the possible fate of nutrients leached from land to eventually be deposited on the ocean floor. At a larger scale, we find that the chemistry of stream water is profoundly influenced by the biota of the area of land it drains (its catchment area; see Section 21.2.4) and, in turn, influences the chemistry and biota of the lake, estuary or sea into which it flows. We deal with the details of nutrient fluxes through terrestrial and aquatic ecosystems in Sections 21.2 and 21.3. Other investigators are interested in the global scale. With their broad brush they paint a picture of the contents and fluxes of the largest conceivable compartments – the entire atmosphere, the oceans as a whole, and so on. Global biogeochemical cycles will be discussed in Section 21.4.

21.1.3 Nutrient budgets

Nutrients are gained and lost by ecosystems in a variety of ways (Figure 21.2). We can construct a nutrient budget by identifying and measuring all the processes on the credit and debit sides of the equation. For some nutrients, in some ecosystems, the budget may be more or less in balance.

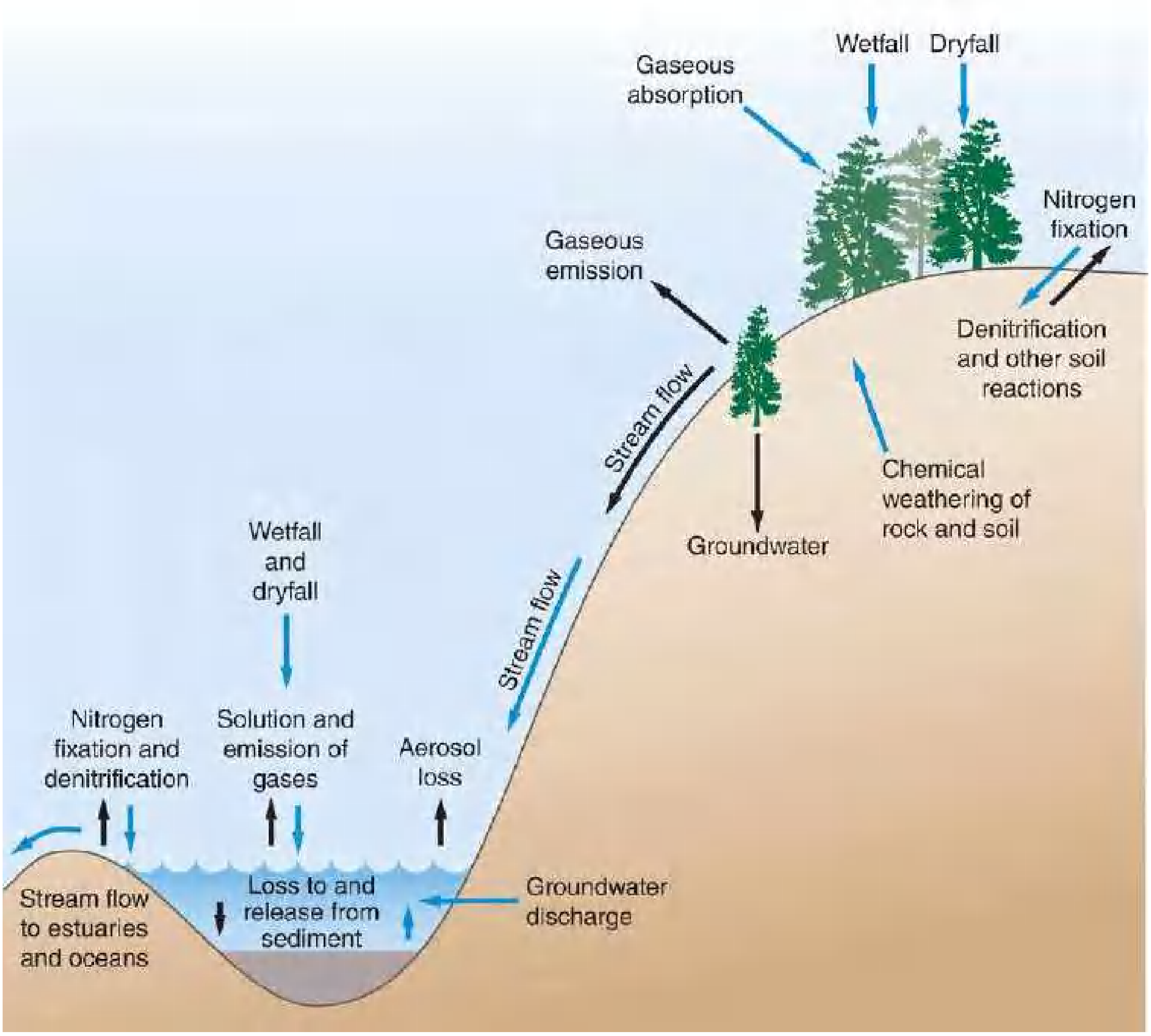


Figure 21.2 Components of the nutrient budgets of a terrestrial and an aquatic system. Note how the two communities are linked by stream flow, which is a major output from the terrestrial system and a major input to the aquatic one. Inputs are shown in colour and outputs in black.

inputs sometimes balance outputs … but inputs can exceed outputs …

In other cases, the inputs exceed the outputs and elements accumulate in the compartments of living biomass and dead organic matter. This is especially obvious for carbon during community succession when the total pool of organic carbon increases as primary successions proceed from their pioneer stages through to mature forest (Section 18.4.1).

… and vice versa

For phosphorus, on the other hand, outputs from terrestrial ecosystems exceed inputs because of the continual loss from the principal nutrient pool derived from bedrock (mostly as the mineral apatite) (Figure 21.3). However, the composition of the retained phosphorus changes, with organic phosphorus in plants increasing in early succession and organic phosphorus in the soil eventually becoming a dominant part of the budget. ‘Labile’ inorganic phosphorus is available to plants whereas the unavailable, ‘occluded’ phosphorus increases through time as phosphorus in the soil is physically encapsulated or surrounded by minerals such as iron and aluminium oxides (Yang & Post, 2011).

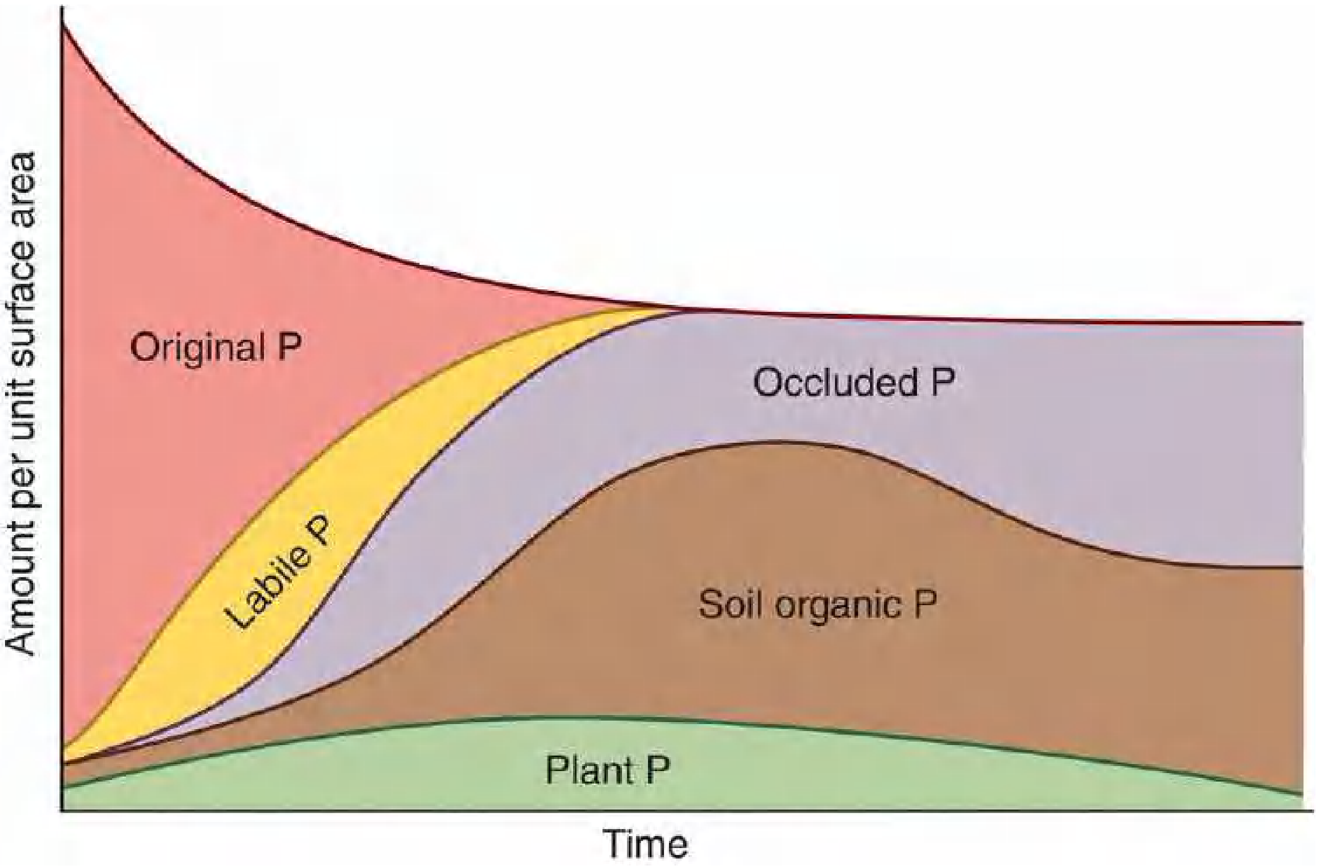


Figure 21.3 **A model of phosphorus (P) dynamics during primary succession.** The timescale may be decades to millennia, depending on the succession concerned. Amounts in the vegetation and soil are shown per unit surface area of ground. Original P is inorganic P in the system when succession begins. Labile P is the fraction of inorganic P that is readily available to plants. Occluded inorganic P is unavailable to plants. P in organic form occurs in both plants and soil.

Source: After Vitousek (2004).

nutrient losses from terrestrial to aquatic ecosystems

Acid rain (produced by oxides of sulphur and nitrogen from the burning of fossil fuels) can increase loss from terrestrial to aquatic ecosystems when the enhanced export of cations exceeds replenishment from weathering. Outputs from terrestrial ecosystems may also exceed inputs if the biota is disturbed by an event such as fire or large-scale deforestation or intensive agriculture. Taken overall, human activities have massively increased nutrient losses from terrestrial to aquatic ecosystems.

APPLICATION 21.1 Nutrient pollution of aquatic ecosystems

cultural eutrophication of lakes and rivers …

In areas of the globe with dense human populations and/or where industrial agriculture is prevalent, water bodies receive excess nutrients in the form of nitrogen and phosphorus from human and domestic animal sewage and synthetic fertilisers. Such cultural eutrophication of lakes generally fuels a change from low to very high primary productivity with a shift from macrophyte to phytoplankton domination, harmful algal blooms, frequent occurrences of low dissolved oxygen concentrations when the excessive biomass decomposes, and reductions in biodiversity including fish kills (Schindler, 2012). A similar chain of events occurs in rivers with long water retention times, such as the lower reaches of large rivers and impounded river sections, while short retention time rivers tend to become dominated by benthic algae (Hilton *et al*., 2006). The adverse human consequences include loss of cultural ecosystem services (aesthetic enjoyment, angling) and significant economic costs associated with loss of tourism and the need for enhanced water treatment.

… and estuaries and coastal marine environments

Not surprisingly, estuaries in developed areas of the world are also subject to cultural eutrophication, with significant economic consequences in lost fisheries, but the problem does not stop there. Evidence has accumulated of eutrophication-related, oxygen-depletion zones in coastal marine environments around the mouths of estuaries. These have been commonly referred to as ‘dead zones’, because of a greatly impaired ability to support life in bottom and near-bottom water, and more than 400 have been identified worldwide (Figure 21.4). The coastal dead zones occur adjacent to the most highly developed areas of land.

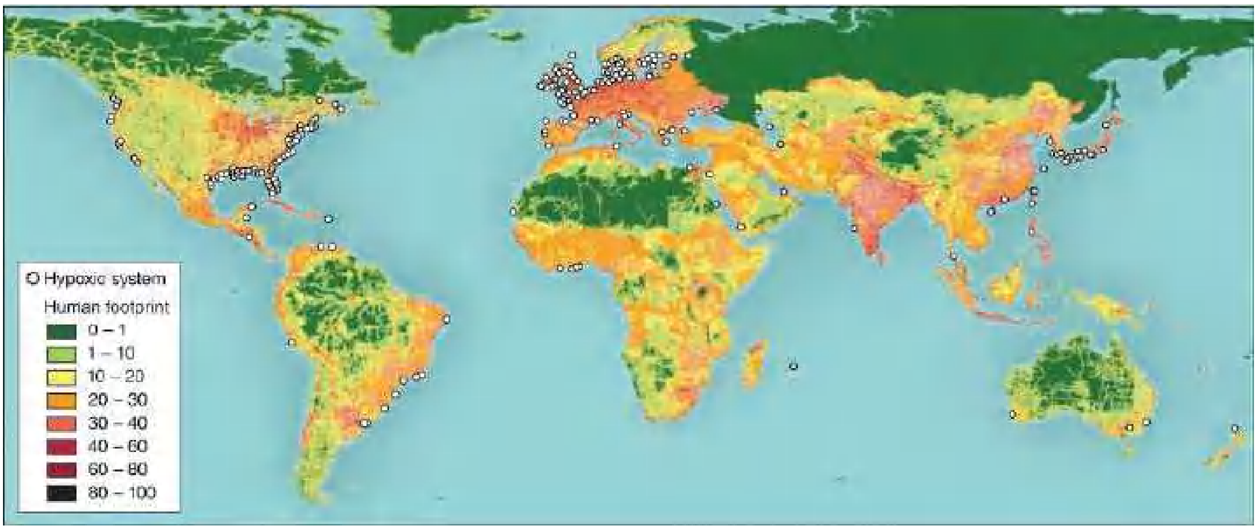


Figure 21.4 Global distribution of eutrophication-related coastal dead zones. Coastal dead zones are depicted by open circles. There is a close match with regional measures of the human footprint. The colour key is for Sanderson *et al.*'s (2002) human influence index, which takes into account human population density and access and the scale of power infrastructure, normalised and expressed as a percentage to reflect the continuum of human influence across each terrestrial biome. A data point from Stewart Island, New Zealand has been removed because the oxygen concentration was recorded below a salmon farm and is not related to catchment eutrophication (S. F. Thrush, pers. comm.).

Source: From Diaz & Rosenberg (2008).

The key to reducing dead zones is to treat sewage to remove nutrients (tertiary treatment) and to keep fertilisers on the land and out of the water. Diaz and Rosenberg (2008) suggest that returning to preindustrial levels of nutrient input would be unrealistic but argue that an appropriate management goal is to reduce nutrient inputs to levels that occurred in the mid-19th century before the spread of dead zones began.

21.2 Nutrient budgets in terrestrial communities

21.2.1 Inputs to terrestrial communities

nutrient inputs ... from the weathering of rock and soil ...

Weathering of parent bedrock and soil is generally the dominant source of nutrients such as calcium, iron, magnesium, phosphorus and potassium, which may then be taken up via the roots of plants. Mechanical weathering is caused by processes such as freezing of water and the growth of roots in crevices. However, much more important to the release of plant nutrients are chemical weathering processes. Of particular significance is carbonation, in which carbonic acid (H_2CO_3) reacts with minerals to release ions such as calcium and potassium. Simple dissolution of minerals in water also makes nutrients available from rock and soil, and so do hydrolytic reactions involving organic acids released by the ectomycorrhizal fungi (see Section 13.9.2) associated with plant roots (Landeweert *et al.*, 2001).

... from the atmosphere ...

Atmospheric CO_2 is the source of carbon to terrestrial communities. Similarly, gaseous nitrogen from the atmosphere provides most of the nitrogen content of communities. Several types of bacteria and archaea possess the enzyme nitrogenase and convert atmospheric nitrogen to soluble ammonium (NH_4^+) ions, which can then be taken up through the roots and used by plants. All terrestrial ecosystems receive some available nitrogen through the activity of free-living nitrogen-fixing bacteria (diazotrophs) such as *Azotobacter*, but communities that include plants such as legumes and alder trees (*Alnus* spp.), with their root nodules that hold symbiotic nitrogen-fixing bacteria (including *Rhizobium* and *Bradyrhizobium*; see Section 13.11.1), may receive a much larger amount of their nitrogen in this way. Associations between moss (*Sphagnum fuscum*) and free-living nitrogen-fixers in a harsh subarctic environment fixed 2.6 kg N ha^{-1} $year^{-1}$, which is only slightly more than the atmospheric input of nitrogen in precipitation (Rousk *et al.*, 2015). In a temperate woodland, on the other hand, more than 80 kg ha^{-1} $year^{-1}$ of nitrogen was supplied by biological nitrogen fixation to red alder trees (*Alnus rubra*), compared with 1–2 kg ha^{-1} $year^{-1}$ from rainfall (Bormann & Gordon, 1984). And nitrogen fixation by legume-based crops can be even more dramatic: values in the range 100–300 kg ha^{-1} $year^{-1}$ are not unusual (e.g. Li *et al.*, 2015).

... as wetfall and dryfall ...

Other nutrients from the atmosphere become available to communities as *wetfall* (in rain, snow and fog) or *dryfall* (settling of particles during periods without rain, and gaseous uptake). Rain is not pure water but contains chemicals derived from a number of sources: (i) trace gases, such as oxides of sulphur and nitrogen; (ii) aerosols produced when tiny water droplets from the oceans evaporate in the atmosphere and leave behind particles rich in sodium, magnesium, chloride and sulphate; and (iii) dust particles from fires, volcanoes and windstorms, often rich in calcium, potassium, sulphate and even phosphorus. Thus, by far the largest source of total atmospheric phosphorus is mineral dust resulting from the wind erosion of soils (Vet *et al.*, 2014) and a dust storm in Africa's Sahel Desert can deposit phosphorus as far away as the Amazon rainforest (Figure 21.5). The constituents of rainfall that serve as nuclei for raindrop formation make up the *rainout* component, whereas other constituents, both particulate and gaseous, are cleansed from the atmosphere as the rain falls – these are the *washout* component (Waring & Schlesinger, 1985). The nutrient concentrations in rain are highest early in a rainstorm, but fall subsequently as the atmosphere is progressively cleansed. Snow scavenges chemicals from the atmosphere less effectively than rain, but tiny fog droplets have particularly high ionic concentrations. Nutrients dissolved in precipitation mostly become available to plants when the water reaches the soil and can be taken up by the plant roots. However, some are absorbed by leaves directly.

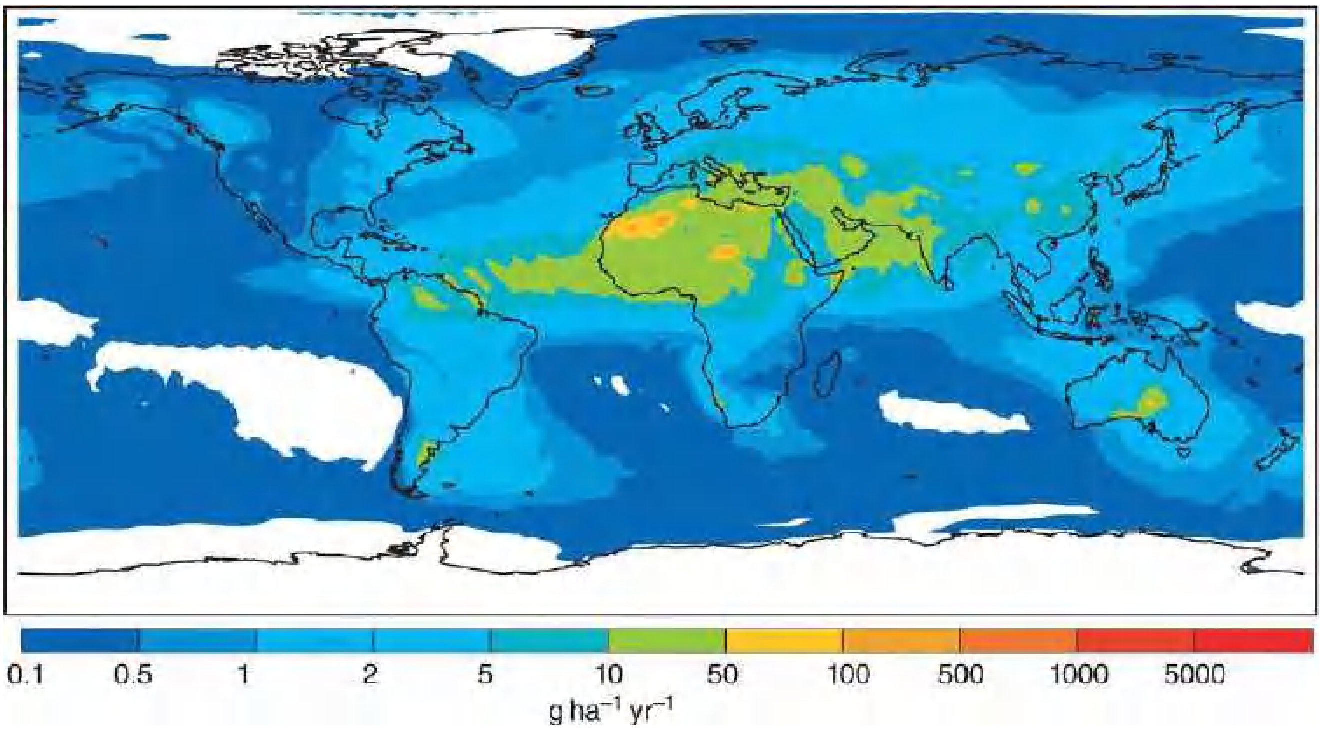
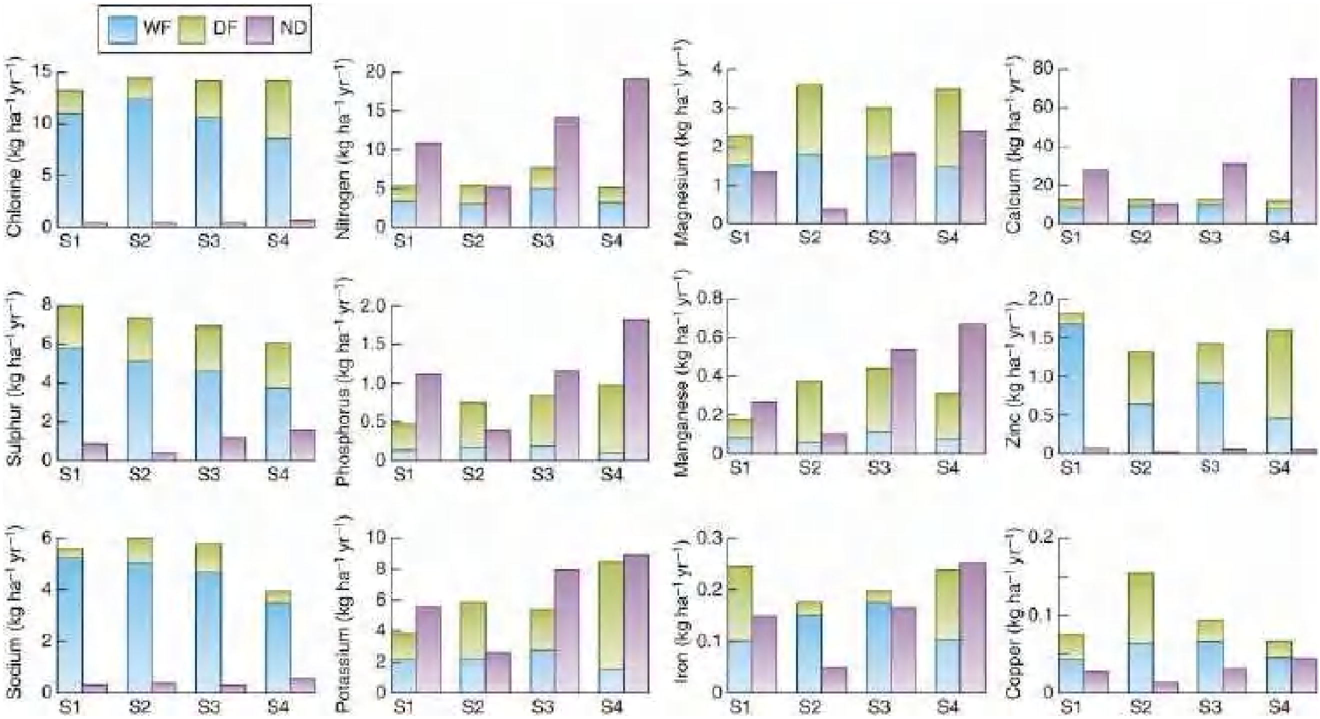


Figure 21.5 Simulated global map of total phosphorus from the atmosphere. Note how deposition rates of more than 100 g ha^{-1} year^{-1} are associated with dust storms originating in Africa and spreading across to South America.

Source: From Mahowald *et al*. (2008).

Dryfall can be a particularly important process in communities with a long dry season. In four Spanish oak forests (*Quercus pyrenaica*) situated along a rainfall gradient, for example, dryfall sometimes accounted for more than half of the atmospheric input to the tree canopy of magnesium, manganese, iron, phosphorus, potassium, zinc and copper (Figure 21.6). For most elements, the importance of dryfall was more marked in forests in drier environments. However, dryfall was not insignificant for forests in wetter locations. Figure 21.6 also plots for each nutrient the annual forest demand (annual increase in above-ground biomass multiplied by the mineral concentration in the biomass). Annual deposition of many elements in wetfall and dryfall was much greater than needed to satisfy demand (e.g. chlorine, sulphur, sodium, zinc). But for other elements, and especially for forests in drier environments, annual atmospheric inputs more or less matched demand (e.g. phosphorus, potassium, manganese, magnesium) or were inadequate (nitrogen, calcium).



[**Figure 21.6**](#) **Wetfall and dryfall along a rainfall gradient.** Annual atmospheric deposition as wetfall (WF) and dryfall (DF) compared with annual nutrient demand (ND; to account for above-ground tree growth) for four oak forests along a rainfall gradient (S1 wettest, S4 driest) in Spain.

Source: After Marcos & Lancho (2002).

While we may conceive of wetfall and dryfall inputs arriving vertically, part of the pattern of nutrient income to a forest depends on its ability to intercept horizontally driven airborne nutrients. This was demonstrated for mixed deciduous forests in New York State when the aptly named Weathers *et al.* (2001) showed that inputs of sulphur, nitrogen and calcium at the forest edge were 17–56% greater than in its interior. The widespread tendency for forests to become fragmented as a result of human activities is likely to have had unexpected consequences for their nutrient budgets because more fragmented forests have a greater proportion of edge habitat.

... from hydrological inputs ...

Stream water plays a major role in the output of nutrients from terrestrial ecosystems (see Section 21.3). However, in a few cases, stream flow can provide a significant input to terrestrial communities when, after flooding, material is deposited in floodplains.

... and from human activities

Last, and by no means least, human activities contribute significant inputs of nutrients to many communities. For example, the amounts of CO_2 and oxides of nitrogen and sulphur in the atmosphere have been increased by the burning of fossil fuels, and the concentrations of nitrate and phosphate in stream water have been raised by agricultural practices and sewage disposal. These changes have far-reaching consequences, which will be discussed later.

21.2.2 Outputs from terrestrial communities

nutrients can be lost ...

A particular nutrient atom may be taken up by a plant that is then eaten by a herbivore which then dies and is decomposed, releasing the atom back to the soil from where it is taken up through the roots of another plant. In this manner, nutrients may circulate within the community for many years. Alternatively, the atom may pass through the system in a matter of minutes, perhaps without interacting with the biota at all. Whatever the case, the atom will eventually be lost through one of the variety of processes that remove nutrients from the system (see Figure 21.2). These processes constitute the debit side of the nutrient budget equation.

... to the atmosphere ...

Release to the atmosphere is one pathway of nutrient loss, and is of particular significance in carbon and nitrogen budgets. In some communities there is an approximate annual balance in the carbon budget; the carbon fixed by photosynthesising plants is balanced by the carbon released to the atmosphere as CO_2 from the respiration of plants, microorganisms and animals, but this is by no means always the case (Section 21.2.3). Other gases are released through the activities of anaerobic bacteria. Methane (CH_4) is a well-known product of the soils of bogs, swamps and floodplain forests, produced by bacteria in the waterlogged, anoxic zone of wetland soils. Some CH_4 is also produced by bacteria in the anaerobic guts of animals, including termites. However, the net flux of CH_4 to the atmosphere depends on the rate at which it is produced in relation to its rate of consumption by aerobic bacteria in shallower, unsaturated soil horizons, or within termite mounds (Jamali *et al*., 2011), such that most does not reach the atmosphere. Other pathways of carbon loss are important in particular instances. For example, fire can turn a very large proportion of a community's carbon into CO_2 in a very short time. Substantial losses of nutrients also occur when foresters or farmers harvest and remove their trees and crops. And, notably, where intensive agriculture is practiced, the anaerobic guts of cattle and other ruminants release large amounts of CH_4 to the atmosphere.

We have seen that many types of free-living and symbiotic nitrogen-fixing bacteria convert atmospheric N_2 into ammonium ions that can be taken up and converted to organic form by plants, which are eaten by animals and decomposed by microbes (Section 21.2.1). On the debit side, other bacteria, in environments largely devoid of oxygen, are constantly converting nitrate into molecular N_2 in the process of denitrification. A third critical process in the nitrogen cycle is nitrification, in which bacteria convert ammonium into nitrate. Besides the nitrogen that returns to the atmosphere as a result of denitrification, both nitrification and denitrification produce intermediate compounds, such as the gases N_2O, NO and NO_2, some of which leak to the atmosphere (Figure 21.7).

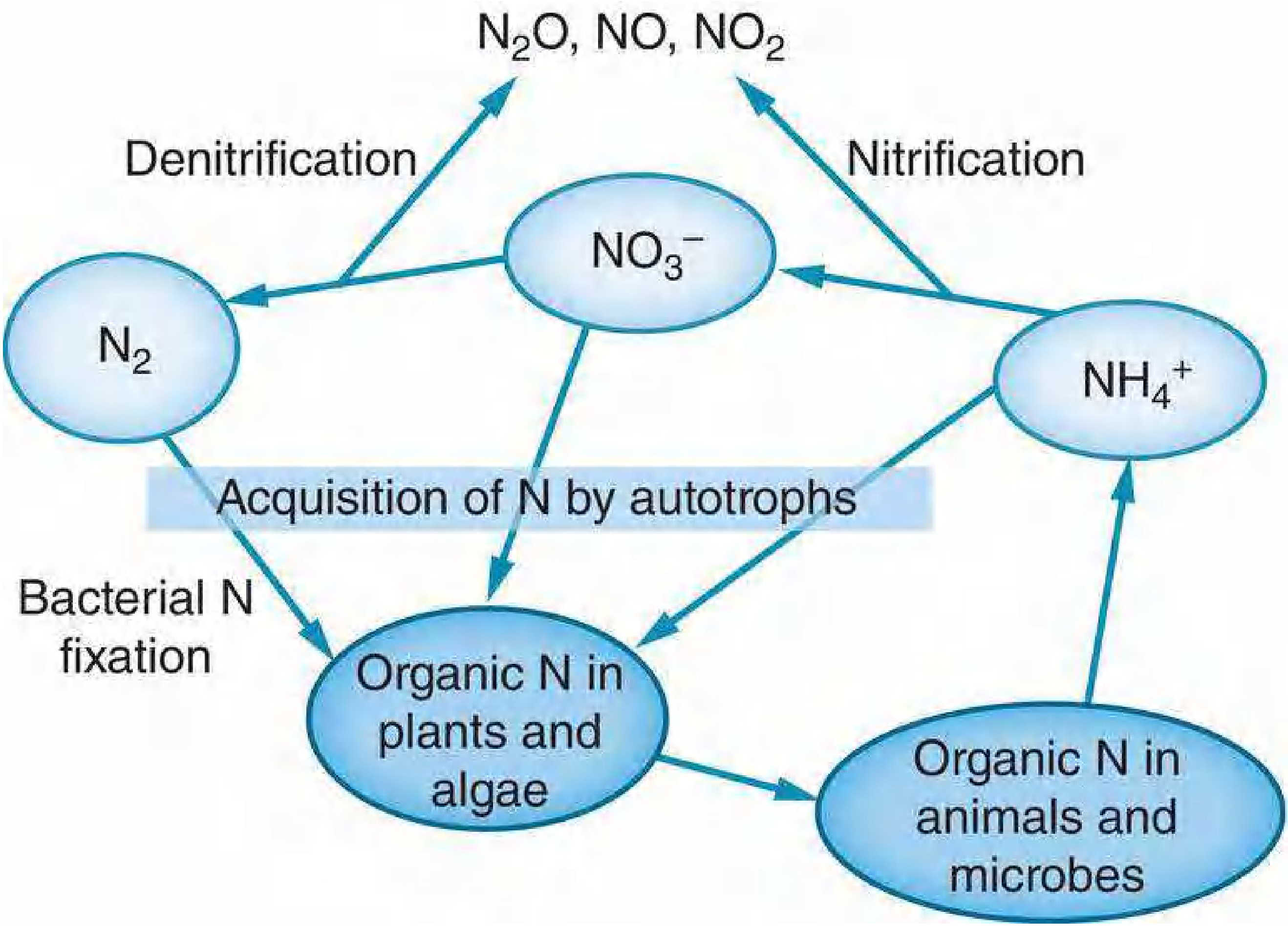


Figure 21.7 **A simplified model of nitrogen transformations in ecosystems.** Light blue ovals represent sources of nitrogen for the biota. Dark blue ovals represent pools of nitrogen acquired by the biota.

Source: After Howarth (2002).

... and to groundwater and streams

For many elements, the most important pathway of loss is in stream flow. The water that drains from the soil of a terrestrial community, via the groundwater into a stream, carries a load of nutrients that is partly dissolved and partly particulate. With the exception of iron and phosphorus, which are not mobile in soils, the loss of plant nutrients is predominantly in solution. Particulate matter in stream flow occurs both as dead organic matter (mainly tree leaves) and as inorganic particles. After rainfall, or snowmelt, the water draining into streams is generally more dilute than during dry periods, when the concentrated waters of soil solution make a greater contribution. However, the effect of high volume more than compensates for lower concentrations in wet periods. Thus, total loss of nutrients is usually greatest in years when rainfall and stream discharge are high. In regions where the bedrock is permeable, losses occur not only in stream flow but also in water that drains deep into the groundwater. This may discharge into a stream or lake after a considerable delay and at some distance from the terrestrial community.

21.2.3 Carbon inputs and outputs may vary with forest age

inputs and outputs are by no means always in balance

Law *et al.* (2001) compared patterns of carbon storage and flux in a young (clear cut 22 years previously) and an old forest (not previously logged, trees from 50 to 250 years old) of ponderosa pine (*Pinus ponderosa*) in Oregon, USA. Their results, which relate to a secondary succession (see Section 18.4.1) in which the early stages retain large stores of organic matter from previous vegetation, are summarised in Figure 21.8.

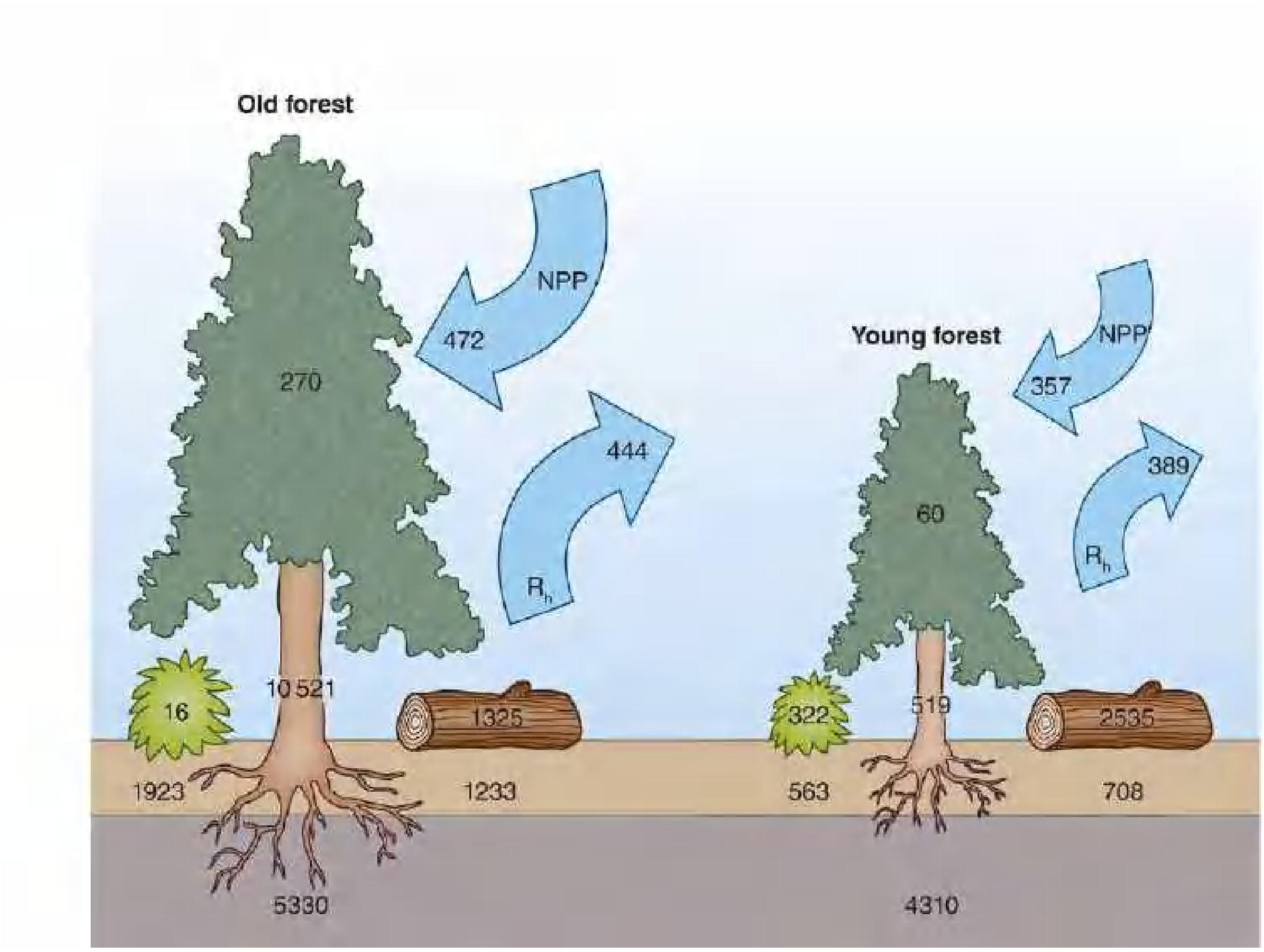


Figure 21.8 Annual carbon budgets for an old and a young ponderosa pine forest. Carbon storage figures are in $\mathrm{g\,C\,m^{-2}}$ while net primary productivity (NPP) and heterotrophic respiration (R_h) are in $\mathrm{g\,C\,m^{-2}\,year^{-1}}$ (arrows). The numbers above ground represent carbon storage in tree foliage, in the remainder of forest biomass, in understory plants, and in dead wood on the forest floor. The numbers just below the ground surface are for tree roots and litter. The numbers deepest underground are for soil carbon

. *Source*: After Law *et al.* (2001).

Total ecosystem carbon content (vegetation, detritus and soil) of the old forest was about twice that of its young counterpart. There were notable differences in percentage carbon stored in living biomass (61% in old, 15% in young) and in dead wood on the forest floor (6% in old, 26% in young). These differences reflect the influence of soil organic matter and woody debris in the young forest derived from the prelogged period of its history. As far as living biomass is concerned, the old forest contained more than 10 times as much as the young forest, with the biggest difference in the wood component of tree biomass.

Below-ground primary productivity differed little between the two forests but because of a much lower above-ground net primary productivity (ANPP) in the young forest, total net primary productivity (NPP) was 25% higher in the old forest. Shrubs accounted for 27% of ANPP in the young forest, but only 10% in the shadier old forest. In both forests, respiration from the soil community accounted for 77% of total heterotrophic respiration.

> the old forest was a net sink for carbon (input greater than output) …

Heterotrophic respiration (decomposers, detritivores and other animals) was somewhat lower in the old forest than NPP, indicating that this forest is a net sink for carbon. In the young forest, however, heterotrophic respiration exceeded NPP making this site a net source of CO_2 to the atmosphere. In a similar study in six successional stands (5 to more than 90 years old) in a humid, temperate mixed forest in Virginia, USA, the switch from carbon source to carbon sink occurred 5–19 years after abandonment of the previously agricultural land (Wang & Epstein, 2013).

> … whereas the young forest was a net carbon source (output greater than input)

These results provide a good illustration of the pathways, stores and fluxes of carbon in forest communities. They also serve to emphasise that inputs and outputs are not always in balance in ecosystems.

APPLICATION 21.2 Managing forests to mitigate climate warming

Whether or not a forest is a net source or sink for carbon is not just of academic interest. We are experiencing dramatic climate warming caused by fossil-fuel-generated CO_2 in the atmosphere, but some mitigation is provided by the retention or planting of forests that act as carbon sinks by sequestering carbon in biomass, litter and soil. An increasing fraction of the world's forests are intensively managed to meet human needs for wood, fibre and other ecosystem services. Plantation foresters use species and genotypes that maximise merchantable wood volume, but traditionally they have not been interested in whether their plantations are also likely to be carbon sinks. It is true that managed forests will assimilate more carbon in biomass than agricultural or many other land uses that they replace, but we should also ask about the relative extent to which managed and unmanaged forests enhance carbon sequestration by absorbing more carbon than is respired over the long term.

In their meta-analysis, Noormets *et al.* (2015) reviewed published data for more than 3000 forests globally, of which nearly 900 were managed. Average tree age in the managed forests was about 50 years younger than in unmanaged forests, and the latter contained about twice the carbon stocks of their managed counterparts (Table 21.1). Nevertheless, both categories had similar gross and net primary productivities. Managed forests, however, had higher net ecosystem productivities (NEP) (NPP minus total ecosystem respiration RE; see Section 20.1.2), indicating that these highly productive plantation forests were accumulating carbon in biomass, litter and soil, faster than the unmanaged forests. Note, however, that both managed and unmanaged forests are carbon sinks (positive NEPs) and thus can mitigate the effects of increased carbon release.

Table 21.1 Global mean (± SE) carbon pools and fluxes in managed and unmanaged forests. *Source:* From Noormets *et al.* (2015), based on the database of Bond-Lamberty & Thomson (2010).

Carbon pools and fluxes	Managed forests	Unmanaged forest
Aboveground biomass carbon ($g\ m^{-2}$)	3465 ± 1104	8870 ± 1042
Belowground biomass carbon ($g\ m^{-2}$)	821 ± 249	1463 ± 178
Litter carbon ($g\ m^{-2}$)	1164 ± 366	1764 ± 258
Mineral soil carbon ($g\ m^{-2}$)	6246 ± 1749	11356 ± 1305
Mean tree age (year)	21 ± 3	68 ± 3
GPP ($g\ C\ m^{-2}\ yr^{-1}$)	1989 ± 169	2187 ± 159
NPP ($g\ C\ m^{-2}\ yr^{-1}$)	674 ± 75	595 ± 32
NEP ($g\ C\ m^{-2}\ yr^{-1}$)	444 ± 84	300 ± 84
Total detritus production ($g\ C\ m^{-2}\ yr^{-1}$)	377 ± 43	491 ± 35
Rh_{soil} ($g\ C\ m^{-2}\ yr^{-1}$)	499 ± 40	458 ± 40
Total detritus production minus Rh_{soil} ($g\ C\ m^{-2}\ yr^{-1}$)	−122	+33

A significant finding was that in contrast to unmanaged forests, soils in managed forests were losing more carbon annually (respiration by soil organisms) than they were regaining in detritus input. Thus, maximising merchantable wood volume may have significant carbon costs for the soil pool, with implications for the overall mitigation potential of forest planting. Noormets *et al.* (2015) conclude that tomorrow's managed forests should strike a balance that maximises as many benefits as possible including the amount of carbon stock retained in the forest, rather than being focused only on merchantable volume. They emphasise the need for greater understanding of the mechanisms underlying (i) the incorporation of carbon into long-lived soil pools and (ii) the priming of decomposition of old soil carbon by new inputs.

More generally, the best way to manage both plantation and natural forests to mitigate climate change is far from clear (Belassen & Luyssaert, 2014). Firstly, we now know that not only harvested, but also unharvested forests have in recent decades been absorbing more carbon than they have been releasing, counter to the previous conventional wisdom that mature forest would be in carbon equilibrium (with biomass maintained at a more or less constant level). It seems that forest growth globally has been accelerated by the fertilising effect of higher atmospheric CO_2 concentrations and higher nitrogen levels associated with emissions from fossil-fuel burning and agricultural and forestry practices. Another factor is that while trees absorb CO_2, use of their timber can reduce other sources of CO_2 emission when it replaces fossil fuels or carbon-intensive materials such as concrete and steel, making the decision to harvest or not harvest more complex. The biggest unknown, however, is the effect of predicted future increases in CO_2 concentrations: will these lead to increases in photosynthesis and provide for an enhanced carbon sink? Or will, instead, increased ecosystem respiration come to dominate, releasing more carbon than is absorbed? If forests remain net absorbers, Belassen and Luyssaert (2014) argue for preserving mature forests, because the decomposing residues of forest harvesting quickly add to carbon emissions, and increased use of wood products would fail to compensate. Conversely, if mature forests become carbon sources, increased harvesting could benefit mitigation (if not biodiversity!). Certainly, future forest harvesting should be targeted to uses that save most tonnes of CO_2 per cubic metre harvested (e.g. by replacing steel and cement).

21.2.4 Importance of nutrient cycling in relation to inputs and outputs

closed and open ecosystems

Internal recycling accounts for the bulk of nitrogen and phosphorus taken up in food chains and mineralised from organic matter. Typically, recycled amounts are an order of magnitude greater than amounts entering or leaving terrestrial ecosystems (Vitousek & Matson, 2009). For these nutrients, the ecosystems are comparatively *closed*. It is clear, on the other hand, that many nutrients arrive in ecosystems at rates that far exceed demands from the vegetation (e.g. chlorine, sulphur, sodium, zinc in the case of the Spanish oak forests in Figure 21.6). For these nutrients the ecosystems are comparatively *open*, with rates of exchange across ecosystem boundaries exceeding rates of internal recycling.

the movement of water links terrestrial and aquatic communities

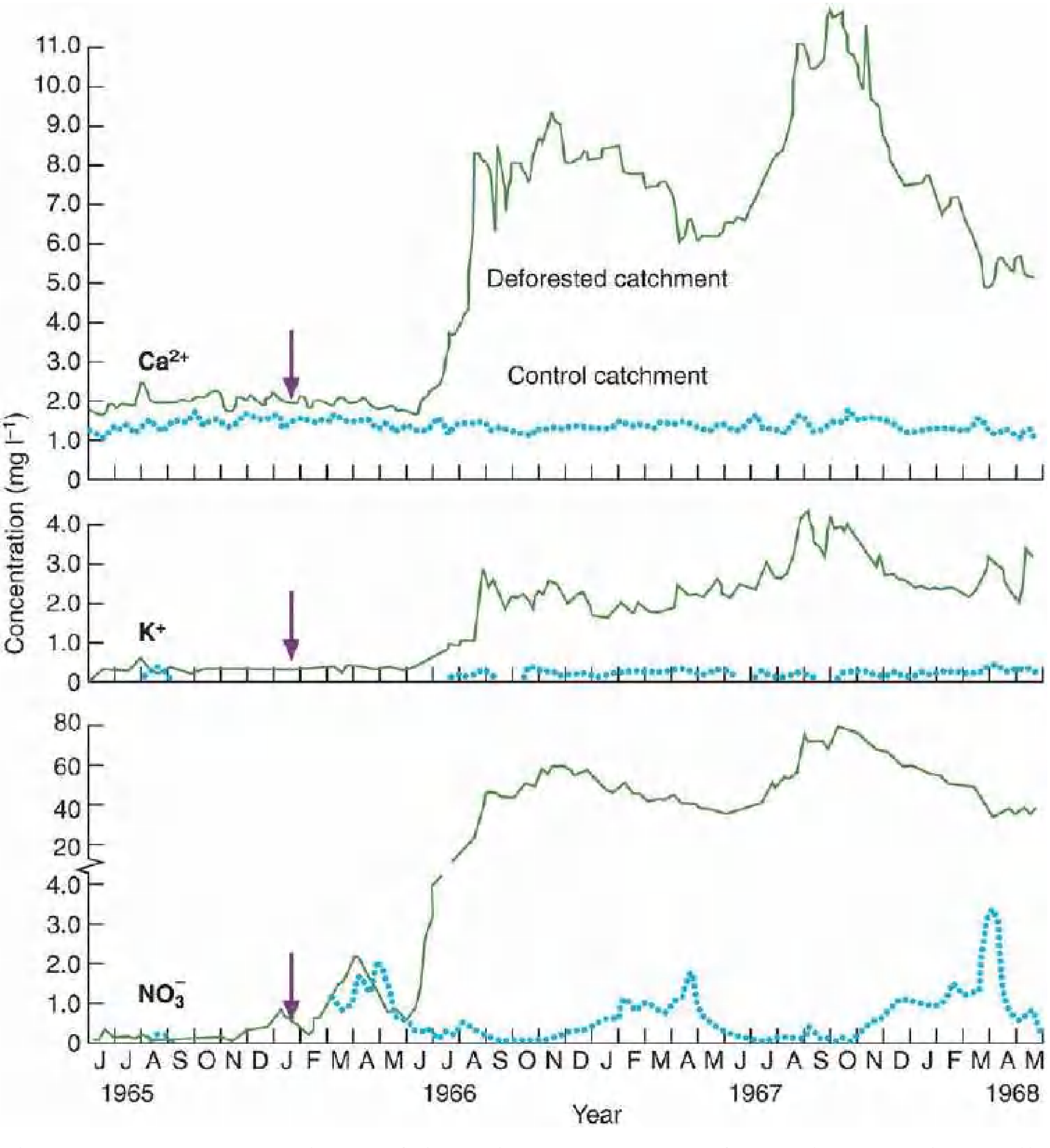
Because many nutrient losses from terrestrial communities are channelled through streams, a comparison of the chemistry of stream water with that of incoming precipitation can reveal a lot about the differential uptake and cycling of chemical elements by the terrestrial biota. A classic study of this question was carried out by Likens and his associates in the Hubbard Brook Experimental Forest, an area of temperate deciduous forest drained by small streams in the White Mountains of New Hampshire, USA. The catchment area – the extent of terrestrial environment drained by a particular stream – was taken as the unit of study because of the role that streams play in nutrient export. Six small catchments were defined and their outflows were monitored. A network of precipitation gauges recorded the incoming amounts of rain, sleet and snow. Chemical analyses of precipitation and stream water made it possible to calculate the amounts of various nutrients entering and leaving the system. In most cases, the output of chemical nutrients in stream flow was greater than their input from rain, sleet and snow. The source of the excess chemicals is parent rock and soil, which are weathered and leached at a rate of about 70 g m^{-2} $year^{-1}$.

at Hubbard Brook inputs and outputs were small compared with internal cycling

The inputs and outputs of nutrients to the Hubbard Brook forest were generally quite small in comparison with the amounts held in biomass and recycled within the system. Nitrogen was added not only in precipitation (6.5 kg ha^{-1} $year^{-1}$) but also through nitrogen fixation by microorganisms (14 kg ha^{-1} $year^{-1}$). (Note that denitrification by other microorganisms, releasing nitrogen to the atmosphere, will also have been occurring but was not measured.) The export in streams of only 4 kg ha^{-1} $year^{-1}$ – only 0.1% of the total nitrogen standing crop held in living and dead forest organic matter – emphasises how securely nitrogen is held and cycled within the forest biomass. However, despite bigger net losses to the forest of other nutrients, their export was still low in relation to the amounts bound in biomass. In other words, relatively efficient recycling was the norm.

deforestation uncouples cycling and leads to a loss of nutrients

In a large-scale experiment, all the trees were felled in one of the Hubbard Brook catchments and herbicides were applied to prevent regrowth. The overall export of dissolved inorganic nutrients from the disturbed catchment then rose to 13 times the normal rate (Figure 21.9). Two phenomena were responsible. Firstly, the enormous reduction in transpiring surfaces (leaves) led to 40% more precipitation passing through the groundwater to be discharged to the streams, and this increased outflow caused greater rates of leaching of chemicals and weathering of rock and soil. Secondly, and more significantly, deforestation effectively broke the within-system nutrient cycling by uncoupling the decomposition process from the plant uptake process. In the absence of nutrient uptake in the spring, when the deciduous trees would have started production, the inorganic nutrients released by decomposer activity were available to be leached in the drainage water.



[Figure 21.9](#) Concentrations of ions in streamwater from the experimentally deforested catchment and a control catchment at Hubbard Brook. The timing of deforestation is indicated by the black arrows. Note that the 'nitrate' axis has a break in it.

Source: After Likens & Borman (1975).

The biggest effect of deforestation was on nitrate-N, emphasising the normally efficient cycling to which inorganic nitrogen is subject. The output of nitrate in streams increased 60-fold after the disturbance. Other biologically important ions were also leached faster as a result of the uncoupling of nutrient cycling mechanisms (potassium: a 14-fold increase; calcium: a seven-fold increase; magnesium: a five-fold increase). However, the loss of sodium, an element of lower biological significance, showed a much less dramatic change following deforestation (a 2.5-fold increase). Presumably it is cycled less efficiently in the forest and so uncoupling had less effect.

21.3 Nutrient budgets in aquatic communities

Aquatic ecosystems contrast with their terrestrial counterparts in terms of nutrient sources, fluxes and sinks. Thus, aquatic systems receive much of their supply of nutrients from their terrestrial catchment areas in stream inflow (see Figure 21.2). In streams and rivers, and also in lakes with a stream outflow, export in outgoing stream water is a major factor. By contrast, in lakes without an outflow (or where this is small relative to the volume of the lake), and also in oceans, nutrient accumulation in permanent sediments is often the major export pathway.

21.3.1 Streams

nutrient ‘spiralling’ in streams

We noted, in the case of Hubbard Brook, that nutrient cycling within the forest was great in comparison to nutrient exchange through import and export. By contrast, only a small fraction of available nutrients take part in biological interactions in stream and river communities. The majority flows on, as particles or dissolved in the water, to be discharged into a lake or the sea. Nevertheless, some nutrients do cycle from an inorganic form in stream water to an organic form in biota to an inorganic form in stream water, and so on. But because of the inexorable transport downstream, the displacement of nutrients is better represented as a spiral (Elwood *et al*., 1983), where fast phases of inorganic nutrient displacement alternate with periods when the nutrient is locked in biomass at successive locations downstream (Figure 21.10). Bacteria, fungi and microscopic algae, growing on the substratum of the stream bed, are mainly responsible for the uptake of inorganic nutrients from stream water in the biotic phase of spiralling. Nutrients, in organic form, pass on through the food web via invertebrates that graze and scrape microbes from the substratum (grazer–scrapers – see Figure 11.5). Ultimately, decomposition of the biota releases inorganic nutrient molecules and the spiral continues.

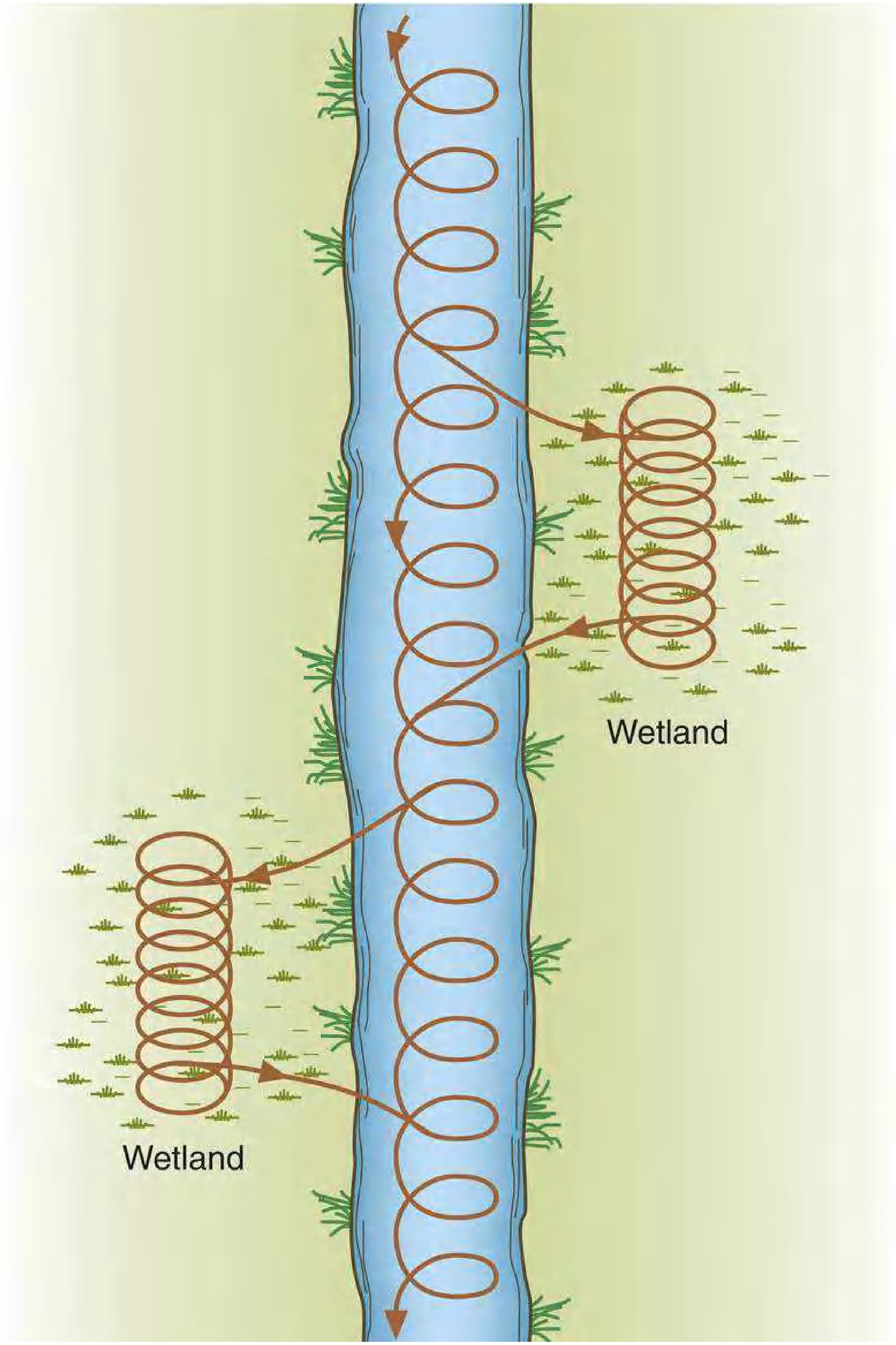

Wetland
Wetland

Figure 21.10 **Nutrient spiralling in a river channel and adjacent wetland areas.**

Source: After Ward (1988).

an alien fish produces a tighter spiral in a New Zealand stream

Sometimes an individual taxon can influence the tightness of nutrient spiralling. Additions of a tracer containing the stable isotope ^{15}N (which occurs naturally but at tiny concentrations) were used to investigate nitrogen dynamics in two New Zealand streams, one with native fish (*Galaxias depressiceps*) and the other with brown trout (*Salmo trutta*) (Simon *et al*., 2004). In New Zealand streams, the alien brown trout initiate a trophic cascade (see Chapter 17.1.1), reducing invertebrate grazing activity, which, in turn, leads to increased algal biomass accumulation and higher rates of photosynthesis per unit area of stream bed. The downstream pattern of ^{15}N uptake (concentrations measured at increasing distances downstream of the site of ^{15}N addition) showed that demand by benthic algae for nitrate (NO_3) from stream water was, as predicted, higher in the trout stream. As a result, the distance between each loop in the river spiral illustrated in Figure 21.10 (the uptake length, calculated as the inverse of the fractional rate of nutrient uptake per unit stream length) was an order of magnitude less in the trout stream (58 m) than the native fish stream (547 m).

The concept of nutrient spiralling is equally applicable to wetlands, such as backwaters, marshes and alluvial forests, which occur in the floodplains of rivers. However, in these cases spiralling can be expected to be much tighter because of reduced water velocity (Prior & Johnes, 2002).

21.3.2 Lakes

nutrient flux in lakes: important roles for plankton and lake position

In lakes, it is usually the phytoplankton and their consumers, the zooplankton, that play the key roles in nutrient cycling. However, most lakes are interconnected with each other by rivers, and standing stocks of nutrients are determined only partly by processes within the lakes. Their position with respect to other water bodies in the landscape can also have a marked effect on nutrient status. This is well illustrated for a series of lakes connected by a river that ultimately flows into Toolik Lake in Arctic Alaska (Figure 21.11a). The main reason for the downstream increase in magnesium and calcium was increased weathering (Figure 21.11b). This comes about because a greater proportion of the water entering downstream lakes has been in intimate contact with the parent rock for longer. Put another way, the higher concentrations reflect the larger catchment areas that feed the downstream lakes. The pattern for calcium and magnesium may also partly reflect progressive evaporative concentration with longer residence times of water in the system as well as material processing by the biota in streams and lakes as the water moves downstream. The nutrients that generally limit production in lakes, nitrogen and phosphorus, were in very low concentrations and could not be measured reliably. However, the downstream decrease in productivity that was observed (Figure 21.11c) suggests that the available nutrients were consumed by the plankton in each lake and this consumption was sufficient to lower the nutrient availability in successive lakes downstream. The downstream decrease of nitrogen, phosphorus and carbon in particulate matter (Figure 21.11d) simply reflects the lower downstream rates of primary productivity. More broadly, this downstream decline in productivity is unusual. In less pristine conditions, productivity is more likely to increase in a downstream direction, partly because of the addition of more nutrients from larger catchment areas but also because of increasing human inputs in lowland areas through fertiliser application and sewage.

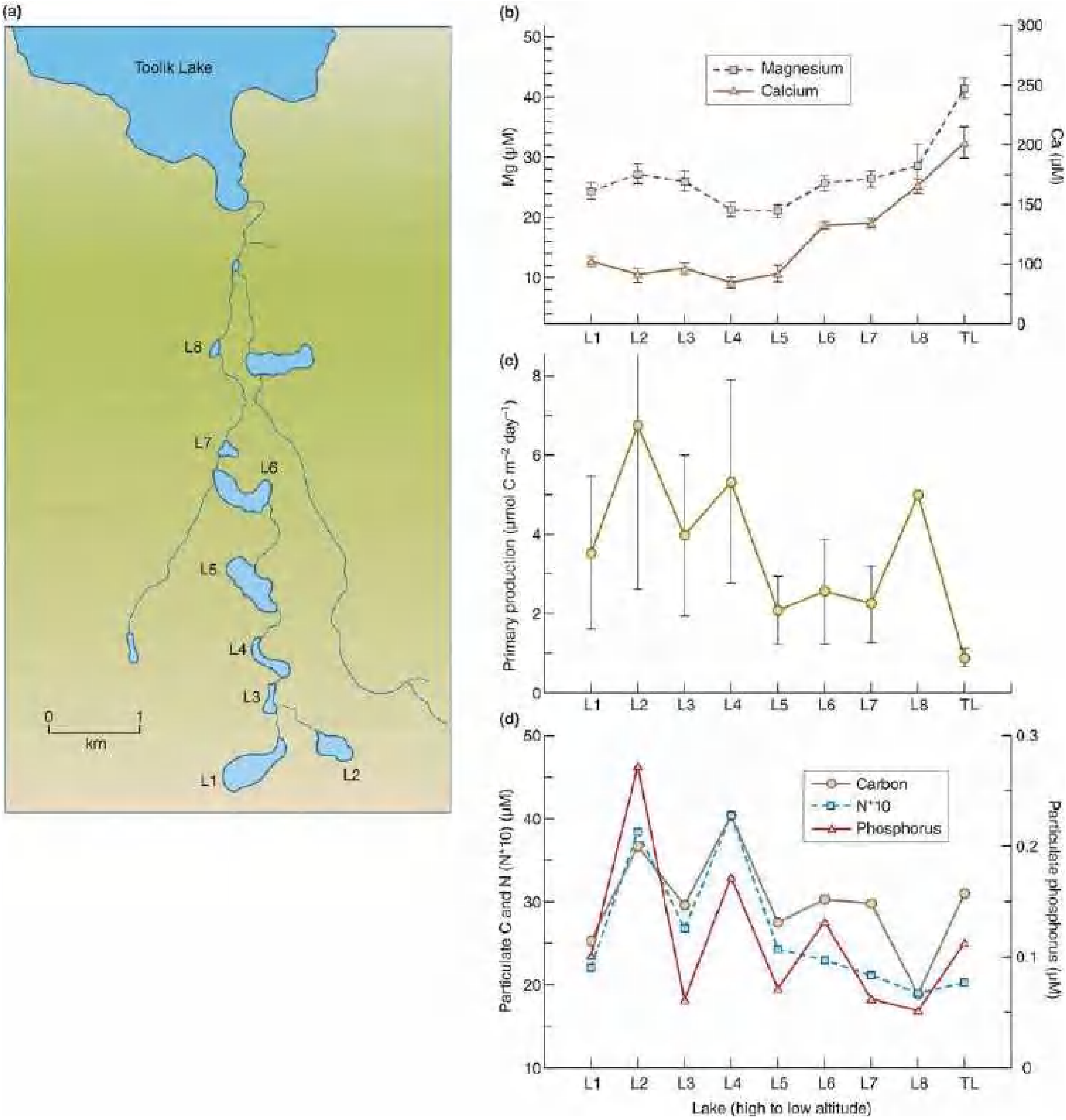


Figure 21.11 The position of a lake with respect to other water bodies in the landscape can have a marked effect on nutrient status. (a) Spatial arrangement of eight small lakes (L1–L8) interconnected by a river that flows into Toolik Lake in arctic Alaska. (b) Mean values, averaged over all sampling occasions during 1991–97 (± SE), for magnesium (Mg) and calcium (Ca) concentrations in the study lakes. (c) Pattern in primary productivity down the lake chain. (d) Mean values for carbon (C), nitrogen (N) and phosphorus in particulate form.

Source: After Kling *et al.* (2000).

saline lakes lose water only by evaporation, and have high nutrient concentrations

Many lakes in arid regions, lacking a stream outflow, lose water only by evaporation. The waters of these endorheic lakes (internal flow) are thus more concentrated than their freshwater counterparts, being particularly rich in sodium (with values up to 30 000 mg l^{-1} or more) but also in other nutrients such as phosphorus (up to 7000 μg l^{-1} or more). Saline lakes should not be considered oddities; globally, they are just as abundant in terms of numbers and volume as freshwater lakes (Williams, 1988). They are usually very fertile and have dense populations of blue–green algae (for example, *Spirulina platensis*), and some, such as Lake Nakuru in Kenya, support huge aggregations of plankton-filtering flamingoes (*Phoeniconaias minor*). No doubt, the high level of phosphorus is due in part to the concentrating effect of evaporation. In addition, there may be a tight nutrient cycle in lakes such as Nakuru in which continuous flamingo feeding and the supply of their excreta to the sediment creates circumstances where phosphorus is continuously regenerated from the sediment to be taken up again by phytoplankton (Moss, 1989).

21.3.3 Estuaries

nutrient flux in estuaries: planktonic and benthic organisms ...

In estuaries, both planktonic organisms (as in lakes) and benthic organisms (as in rivers) are significant in nutrient flux. Hughes *et al.* (2000) introduced tracer levels of the isotope ^{15}N into the water of an estuary in Massachusetts, USA, to study how nitrogen derived from the catchment area is used and transformed in the estuarine food web. They focused their study on the upper, low salinity part of the estuary where water derived from the river catchment first meets the saline influence of tidal seawater. The planktonic centric diatom *Actinocyclus normanii* turned out to be the primary vector of nitrogen to some benthic organisms (large crustaceans) and particularly pelagic organisms (planktonic copepods and juvenile fishes). Certain components of the sedimentary biota received a small proportion of their nitrogen via the centric diatom (10–30%; e.g. pennate diatoms, harpacticoid copepods, oligochaete worms, bottom-feeding fishes such as mummichog (*Fundulus heteroclitus*), and sand shrimps). But many others obtained almost all their nitrogen from a pathway based on plant detritus. The patterns of nitrogen flow through this estuarine food web are shown in Figure 21.12. The relative importance of nutrient fluxes through the grazer and decomposer systems can be expected to vary from estuary to estuary.

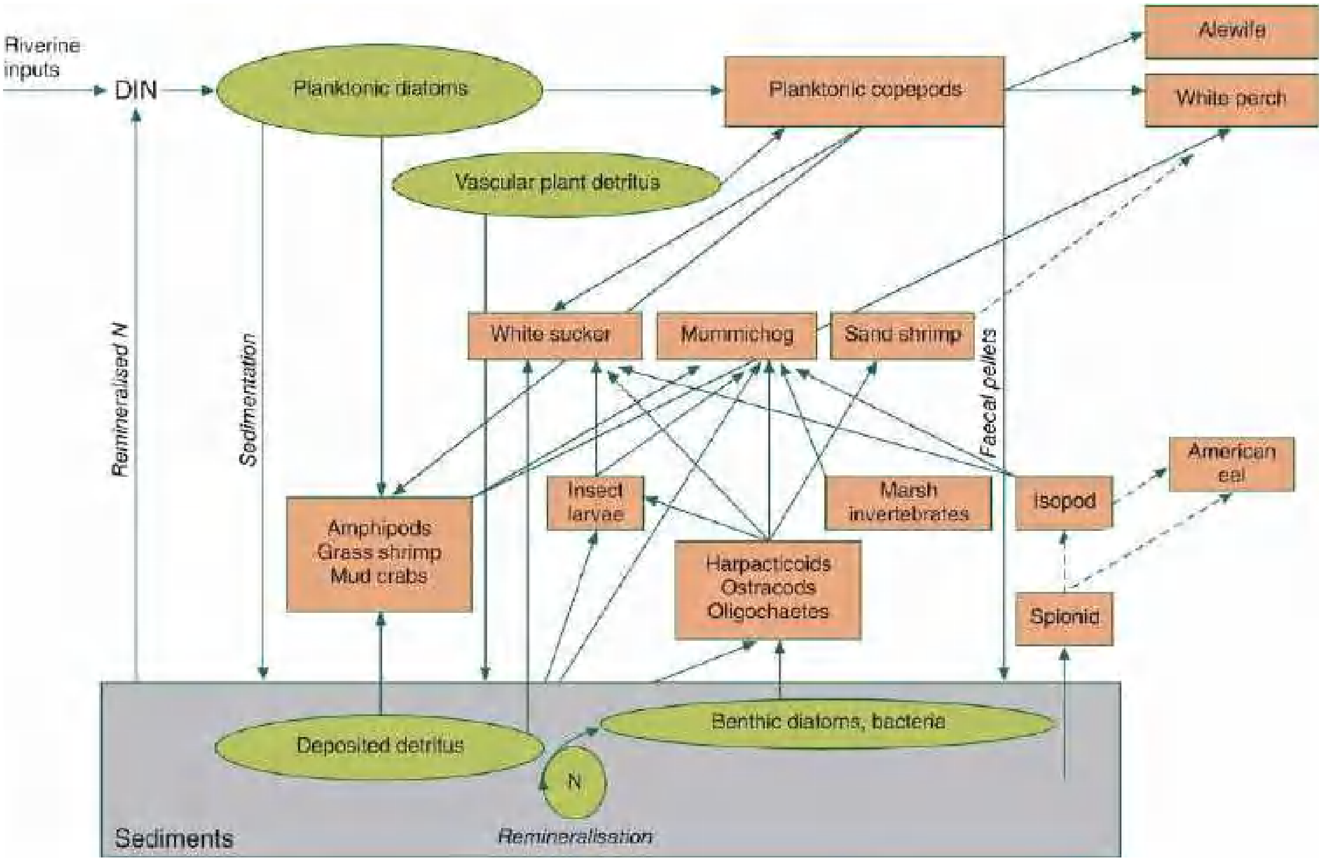
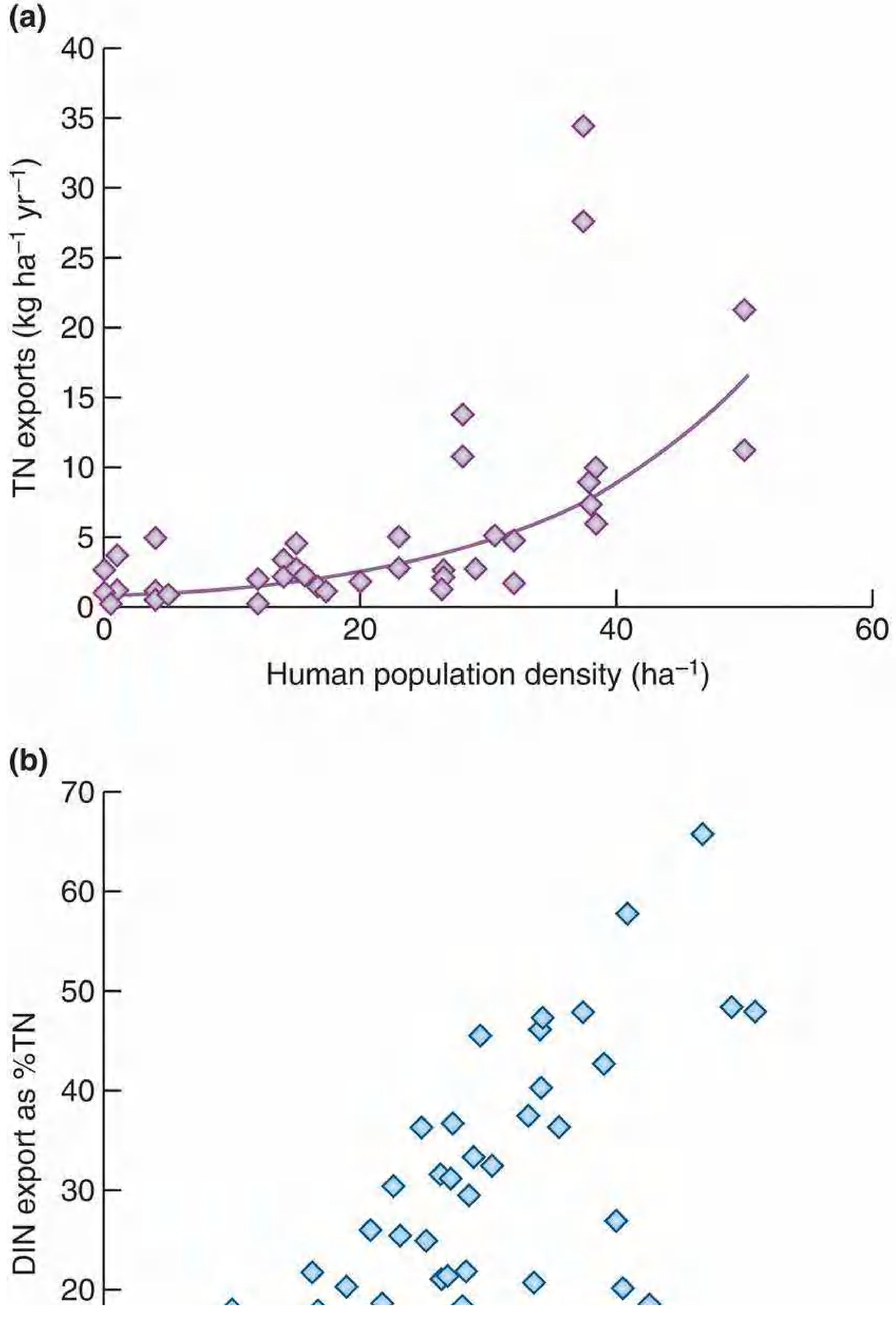


Figure 21.12 Nitrogen flux in an estuary. Conceptual model of nitrogen (N) flux through the food web of the upper Parker River estuary, Massachusetts, USA. Dashed arrows indicate suspected pathways. DIN, dissolved inorganic nitrogen.

Source: After Hughes *et al.* (2000).

… and human activities

The chemistry of estuarine (and coastal marine) water is strongly influenced by features of the catchment area through which the rivers have been flowing, and human activities play a major role in determining the nature of the water supplied. In a revealing comparison, van Breeman (2002) describes the forms of nitrogen in water at the mouths of rivers in North and South America. In the North American case, where the river flows through a largely forested region but has been subject to considerable human impact (fertiliser input, logging, acid precipitation, etc.), nitrogen was almost exclusively exported to estuaries and the sea in inorganic form (only 2% organic). In contrast, a pristine South American river, subject to very little human impact, exported 70% of its nitrogen in organic form. In Australian rivers too, pristine forested catchments export little nitrogen or phosphorus, and the predominant form of nitrogen is organic. However, as human population density increases (greater agricultural runoff and sewage) and forests are cleared (less tight retention of nutrients), the export to river mouths of both nitrogen and phosphorus increases and the predominant form of nitrogen changes to inorganic (Figure 21.13).


(a)
TN exports (kg ha^{-1} yr^{-1})
0
5
10
15
20
25
30
35
40
0
20
40
60
Human population density (ha^{-1})
(b)
DIN export as %TN
20
30
40
50
60
70

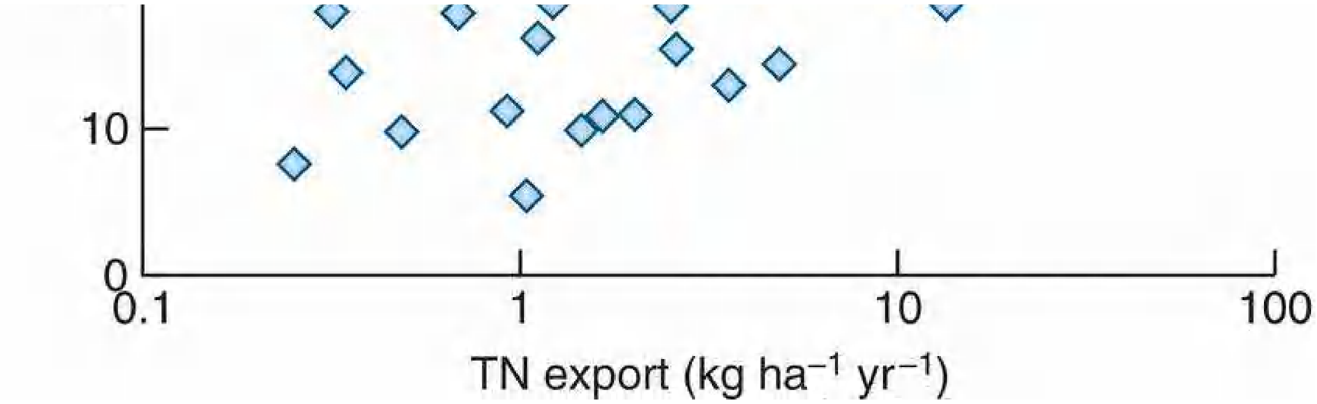


Figure 21.13 The export to river mouths of nitrogen increases with human population density. (a) Export of total nitrogen (TN) in relation to population density in 24 catchment areas near Sydney, Australia. (b) Rivers with low TN export rates (more pristine) contain nitrogen predominantly in organic form and the percentage of TN that is inorganic increases with TN. DIN, dissolved inorganic nitrogen.

Source: After Harris (2001).

21.3.4 Continental shelf regions of the oceans

coastal regions of oceans are influenced by their terrestrial catchment areas …

The nutrient budgets of coastal regions of oceans, like estuaries, are strongly influenced by the nature of catchment areas that supply the water, via rivers, to the sea. Concentrations of nitrogen or phosphorus may limit productivity in these areas as in other water bodies, but a further human-induced effect on the chemistry of river water has special significance for planktonic communities in the oceans. Today, a large proportion of the world’s rivers have been dammed or diverted (for hydroelectric generation, irrigation and human water supply). Associated with damming is the loss of upper soils and vegetation through inundation, loss of soil through shoreline erosion, and underground channelling of water through tunnels. These effects reduce the contact of water with vegetated soil and, therefore, reduce weathering. Figure 21.14 illustrates the patterns of export of dissolved silicate, an essential component of the cell walls of planktonic diatoms in the sea, for a dammed river and a freely flowing river in Sweden. The export of silicate was dramatically lower in the dammed case.

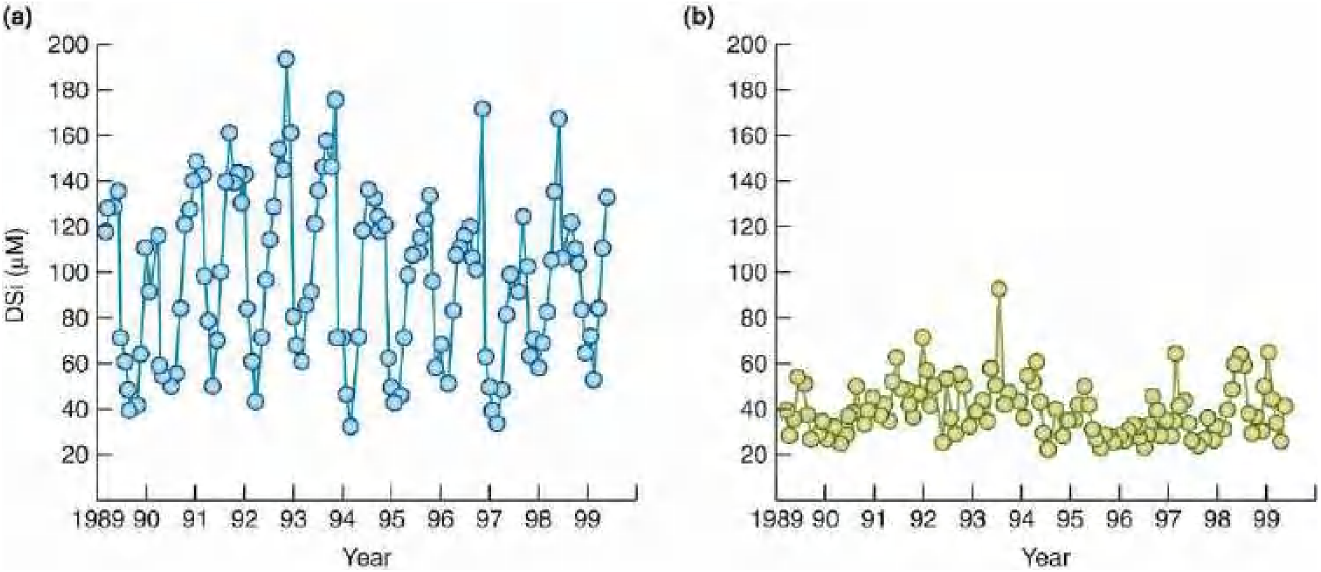


Figure 21.14 The effect of river damming on silicate concentration at the river mouth. Dissolved silicate (DSi) concentrations at the river mouths of (a) the non-dammed River Kalixalven and (b) the dammed River Lulealven.

Source: Humborg *et al.* (2002).

... by local upwelling ...

Another important mechanism of nutrient enrichment in coastal regions is local upwelling, bringing high nutrient concentrations from deep to shallow water where they fuel primary productivity, often producing phytoplankton blooms. Three categories of upwelling have been described and studied off the east coast of Australia: (i) wind-driven upwellings in response to seasonal north and north-easterly breezes; (ii) upwelling driven by the encroachment of the East Australian Current (EAC) onto the continental shelf; and (iii) upwelling caused by the separation of the EAC from the coast. Figure 21.15 provides examples of the distribution of nitrate concentrations associated with each mechanism. Wind-driven upwellings (generally considered to be the dominant mechanism globally) are not persistent or massive in scale. The highest nitrate concentrations are generally associated with encroachment upwellings, while separation upwellings are the most widespread along the coast of New South Wales.

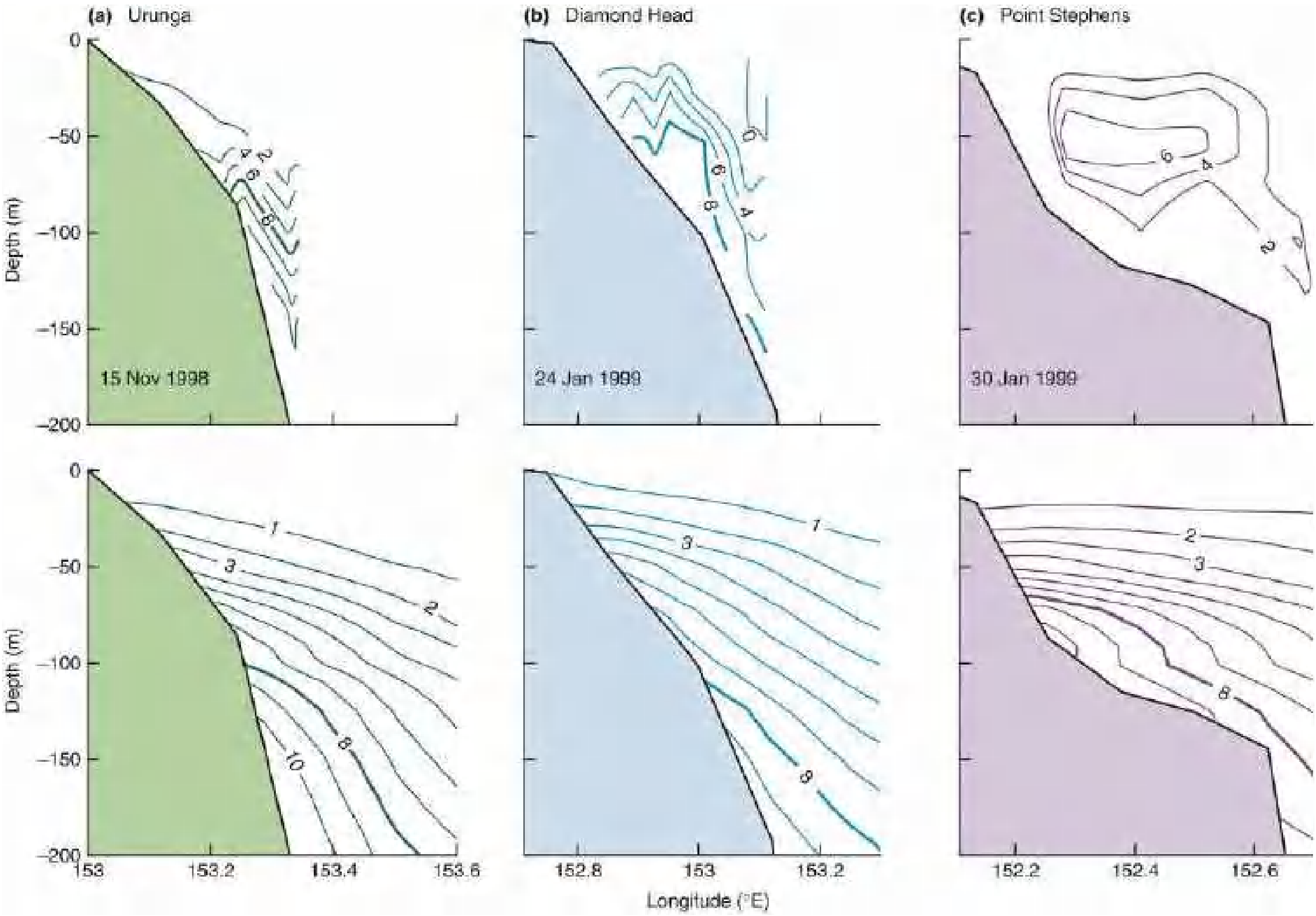


Figure 21.15 Contours of nitrate concentration during upwelling events along the New South Wales coast. (a) Urunga (wind-driven), (b) Diamond Head (encroachment-driven), and (c) Point Stephens (separation-driven). Shaded areas are land. The upper graph in each case shows the mean nitrate concentrations during upwelling, and the lower graphs show the concentrations that can be taken as characteristic of these sites in the absence of an upwelling event. Maximum concentration is 10 μmol l^{-1}. The contour interval is 1 or 2 μmol l^{-1} and the thick lines represent 8 μmol l^{-1}.

Source: After Roughan & Middleton (2002).

… and by atmospheric deposition

Atmospheric deposition is a further source of nutrients to continental shelf regions. The relative importance of catchment inputs, local upwelling and atmospheric deposition for a key nutrient such as nitrogen depends on human activities in the catchment, the physics of the continental shelf and the level of atmospheric pollution. Both the North Sea in Europe and the north-eastern coast of the USA (Maine to Virginia) receive an appreciable supply of upwelling nutrients, while the Gulf of Mexico is more isolated from the deep ocean and receives a much smaller proportion of its nitrogen from this source (Table 21.2). The Gulf of Mexico receives the bulk of its nitrogen from catchment sources, particularly from the Mississippi River that drains a huge area of the USA. Catchment sources are also substantial for the North Sea, whose catchment area is smaller than that of the Gulf of Mexico but where human activity is particularly intense. In contrast, the north-eastern coastal shelf of the USA receives much less nitrogen from catchment runoff, mainly because the land area draining into this region is much smaller. In all three cases, atmospheric deposition makes up a more minor component of nitrogen inputs.

Table 21.2 Sources of nitrogen for three continental shelf areas. Atmospheric deposition is that occurring directly onto the ocean surface. Atmospheric deposition onto land is incorporated into runoff from the terrestrial catchment. 1 teragram (Tg) = 10^{12} g.

Source: After Boyer & Howarth (2008).

	Terrestrial catchment1 teragram (Tg yr^{-1})	**Atmospheric deposition (Tg yr^{-1})**	**Deep ocean (Tg yr^{-1})**
Gulf of Mexico	2.1	0.3	0.15
NE coast of USA	0.3	0.2	1.5
North Sea	1.0	0.6	1.3

APPLICATION 21.3 Constructing wetlands to reduce nitrate runoff to coastal seas

To alleviate problems of coastal eutrophication, and ocean dead zones more generally (see Section 21.1.3), improvements are required to the treatment of wastewater and the management of terrestrial catchment areas to reduce runoff of nutrients from agriculture. The plants and microorganisms in natural wetlands (swamps, marshes, etc.) and in riparian zones along the banks of streams, extract dissolved nutrients as the water filters through the soil. However, most wetlands around the world have been drained and riparian vegetation removed as agriculture has intensified. Sometimes these can be restored, but another approach is to construct artificial wetlands, planted with appropriate vegetation and flow-controlled to maximise nutrient removal.

As a result of his global meta-analysis of 200 examples, Kadlec (2012) concluded that a constructed wetland might be expected to remove 1250–3750 kg ha^{-1} $year^{-1}$ of nitrogen, representing 40–90% of nitrate flowing through it (Kadlec, 2012). Uptake of nitrogen into plant biomass can account for 4–17% of incoming nitrate, but much more is converted by denitrifying bacteria into gaseous nitrogen and channelled to the atmosphere. We saw in Section 21.2.2 that denitrification produces N_2O as an intermediate compound, some of which could leak to the atmosphere. This is a serious consideration because N_2O is a major greenhouse gas (with a much more pronounced effect than CO_2) but Kadlec (2012) calculates the likely effect to be minor.

Whether constructed wetlands will have a significant effect on the nitrate that eventually enters coastal seas depends, of course, not only on removal rates of individual wetlands but also on how many there are and the total area of the catchment that they comprise. Modelling of catchments in Sweden, the USA, China and elsewhere indicates that a significant contribution to water quality improvement (say a reduction of 30–40% in nitrate entering the sea) would generally need 2–7% of the whole catchment to be converted to wetland (Verhoeven *et al*., 2006). Other ecosystem service benefits are associated with constructed wetlands, including increased biodiversity, aesthetic enjoyment and recreational opportunities. However, constructed wetlands are expensive (per kg nitrogen removed) and will generally need to be combined with other practices, such as optimal timing and rate of application of manure/fertiliser, interplanting of crops to minimise nutrient loss, replacing winter crops by spring crops, planting riparian buffers along streams and rivers and retiring some areas of intensive agriculture. Many nations are responsible for eutrophication of the Baltic Sea (Figure 21.16), with associated algal blooms, dead zones and reductions in fish stocks. As part of a multinational effort, Sweden, for example, aims to reduce its nitrogen inputs to the Baltic by 5800 tonnes per year (32% of the anthropogenic load) (Swedish EPA, 2009). The Swedish plan involves a broad suite of actions, including constructed wetlands.

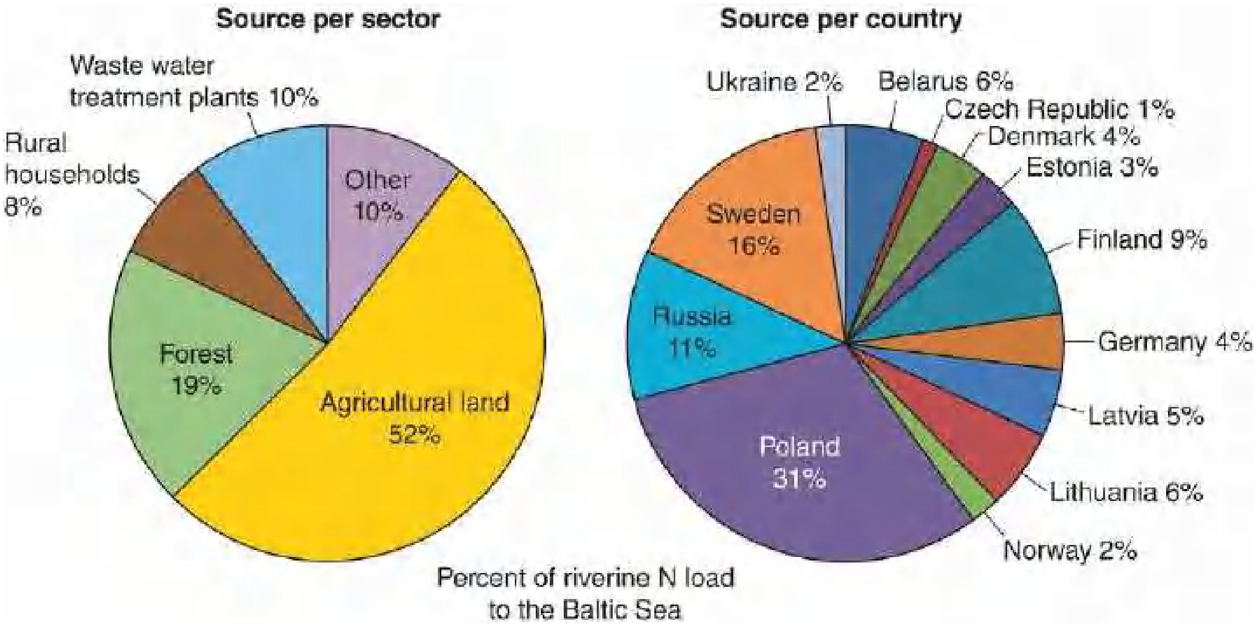


Figure 21.16 Source of catchment inputs of nitrogen (N) to the Baltic Sea according to societal sector and country.

Source: From Arheimer *et al.* (2012).

21.3.5 Open oceans

We can view the open ocean as the largest of all endorheic 'lakes' – a huge basin of water supplied by the world's rivers and losing water only by evaporation.

the open ocean: an important role for plankton

We considered biologically mediated transformations of carbon in terrestrial ecosystems in Section 21.2.3 (Figure 21.8). Figure 21.17 illustrates the same thing but for the open ocean. The main transformers of dissolved inorganic carbon (essentially CO_2) are the small phytoplankton, which recycle CO_2 in the euphotic zone, and the larger plankton, which generate the majority of the carbon flux in particulate and dissolved organic form to the deep ocean. Of the generally small proportion of carbon fixed near the surface that finds its way to the ocean bed, some is consumed by deep-sea biota, some is remineralised into dissolved inorganic form by decomposers, and some becomes buried in the sediment.

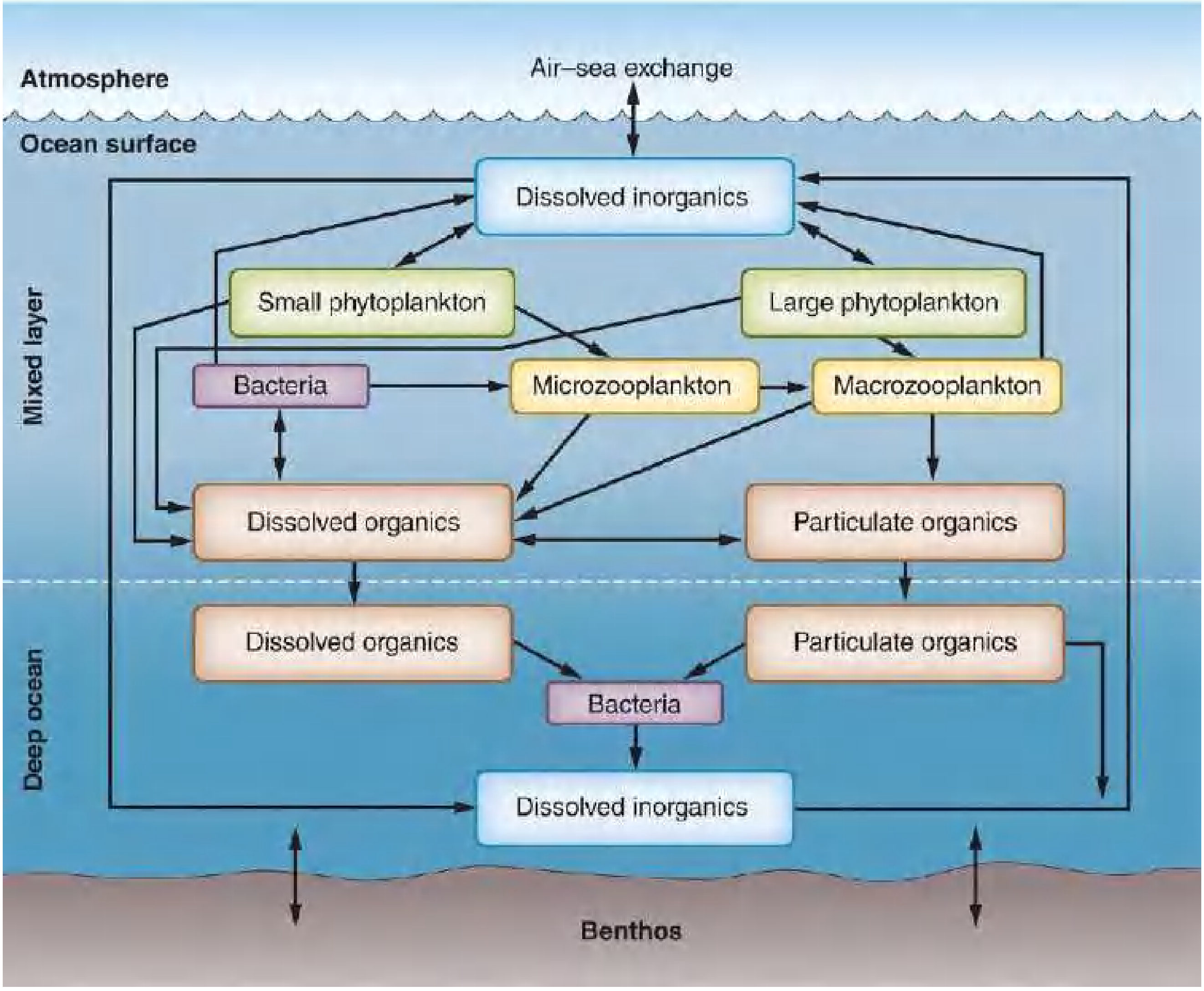


Figure 21.17 Biologically mediated transformations of carbon in the open ocean.

Source: After Fasham *et al*. (2001).

The physical processes responsible for resupplying nutrient-rich water from the ocean's dark depths to the sunlit upper ocean are critical for sustaining marine life. Two processes are particularly important globally: wintertime convective mixing when temperature-dependent density stratification breaks down (in the North Atlantic, for example) and upwelling in ocean divergence zones, such as those in equatorial regions. More recently it has become evident that non-winter storms can also produce upward nutrient pulses that might supply up to 30% as much nutrients as are yielded by wintertime convective mixing (Palter, 2015).

plankton may follow a seasonal pattern

Just as we saw in terrestrial ecosystems, marked seasonal and interannual differences in nutrient flux and availability can be detected in the open ocean. Thus, Figure 21.18a shows how chlorophyll *a* varied during the spring bloom at a site in the North Atlantic, reflecting a succession of dominant phytoplankton species. Large diatoms bloomed first, consuming almost all the available silicate (Figure 21.18b). Subsequently, a bloom of small flagellates used up the remaining nitrate.

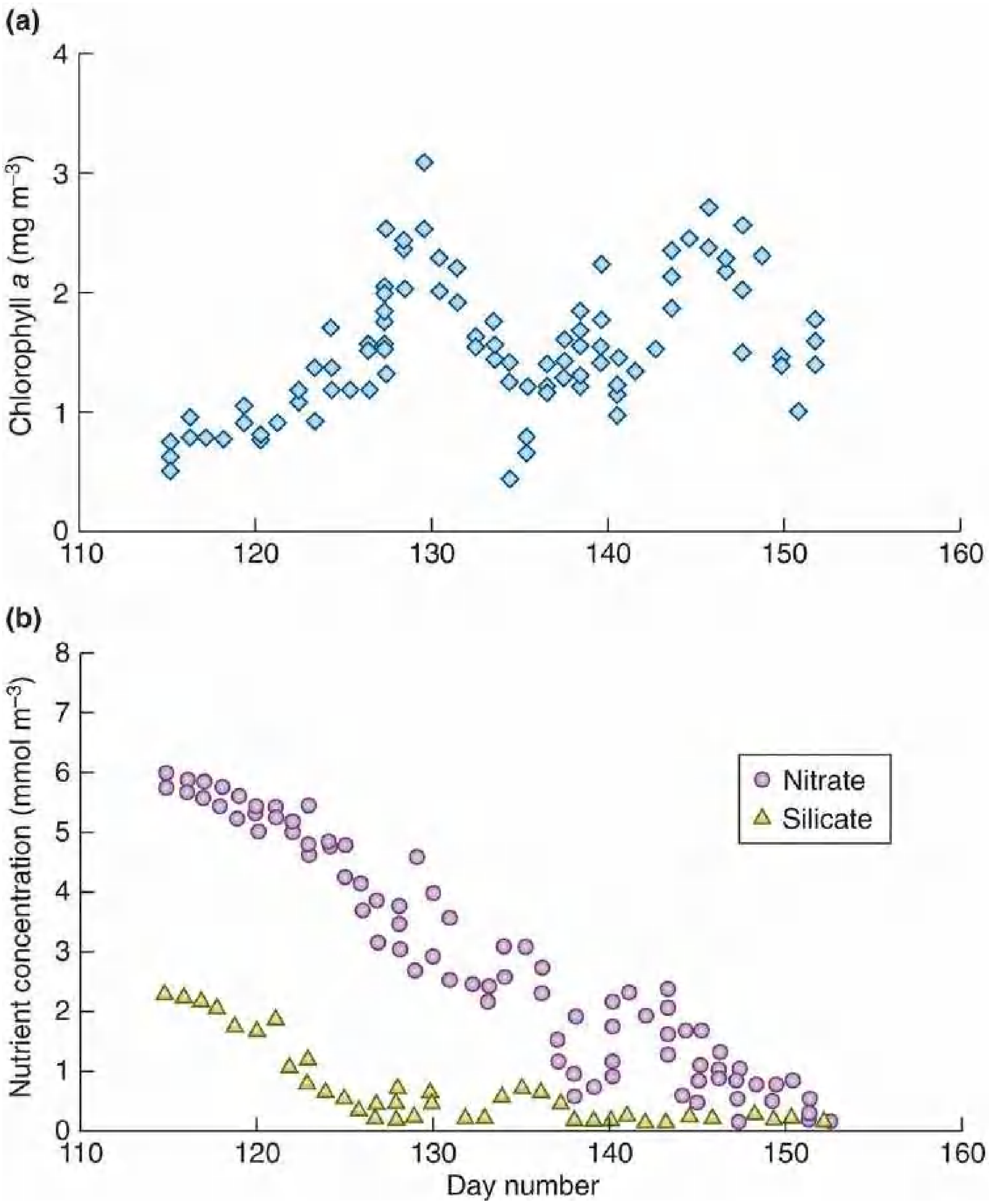


Figure 21.18 Patterns in (a) chlorophyll *a* and (b) silicate and nitrate concentrations during a spring bloom in the North Atlantic. Day number is days since 1 January.

Source: After Fasham *et al*. (2001).

iron as a factor limiting ocean primary productivity?

About 30% of the world's oceans have long been known to have low productivity despite high concentrations of nitrate. The hypothesis that this paradox was due to the iron limitation of phytoplankton productivity has been tested in locations as different as the equatorial Pacific, sub-Arctic Pacific and the open polar Southern Ocean (Boyd *et al*., 2012). Large infusions of dissolved iron at sites covering areas as large as 1000 km^2 have led to dramatic increases in primary productivity, detectable from satellite images, and to decreases in nitrate and silicate as these were taken up during algal production (the results are expressed as nitrate removal for two experiments in Figure 21.19). Bacterial productivity tripled within a few days in both cases, and rates of herbivory by micrograzers (flagellates and ciliates) also increased, but less so in the polar situation (where dominance by a grazer-resistant, highly silicified diatom probably suppressed grazing). The metazoan community, dominated by copepods, showed relatively little change in either situation.

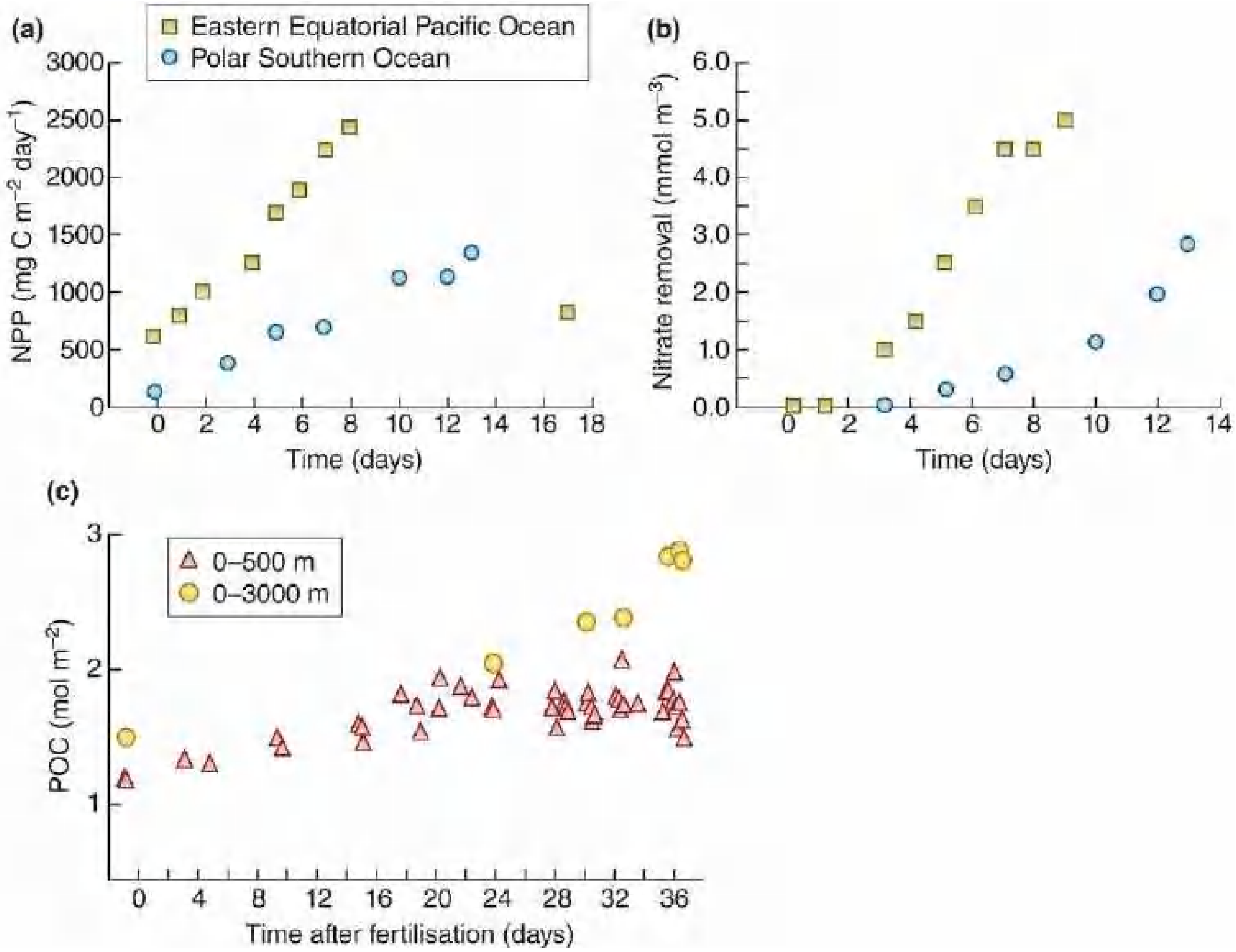


Figure 21.19 Adding dissolved iron to the ocean leads to dramatic increases in primary productivity and decreases in nutrient concentrations. (a) Rates of depth-integrated net primary production (NPP) after iron addition at sites in the eastern equatorial Pacific Ocean and polar Southern Ocean. (b) Nitrate removal during the time course of the two experiments. Note that silicate followed similar patterns. (c) Depth integrated particulate organic carbon (POC) during a 5-week iron-fertilisation experiment in the Southern Ocean. Each data point is for a vertical profile from 0 to 500 m depth or from 0 to 3000 m depth. Note the marked increases in POC with time, particularly at greater depth.

Source: (b) After Boyd (2002). (c) After Smetacek *et al.* (2012).

It is an intriguing thought that blooms in productivity might sometimes be caused by long-distance wind transport of land-derived, iron-rich particles, with consequences even for productivity of important fisheries, including stocks of both Japanese and Californian sardines (*Sardinops sagax*) (Qiu, 2015). This would mirror, but on a very different scale, the high productivity associated with inputs of land-derived, nutrient-rich water from rivers.

APPLICATION 21.4 Could fertilising the ocean with iron reduce global warming?

It is clear that iron fertilisation in certain ocean areas can increase phytoplankton primary production and thus soak up extra CO_2. But could such a *geoengineering* approach effectively mitigate the global warming caused by anthropogenic CO_2 production? This depends on the depths at which organic carbon in phytoplankton is remineralised back to CO_2 by heterotrophic organisms; to have a mitigating effect, the organic carbon must be transported to the deep ocean where it cannot readily re-exchange with the atmosphere. In a five-week experiment (the longest iron enrichment experiment to date) within a self-contained eddy in the Antarctic Circumpolar Current, Smetacek *et al*. (2012) tracked sinking particles from the surface to the seabed (Figure 21.19c). They concluded that at least half the phytoplankton bloom biomass (which peaked in week 4) sank below 1000 m and much of this is likely to have been deposited on the seafloor with a carbon sequestration time of centuries. Their results confirm the potential for iron fertilisation as a mitigation measure, but there are many unknowns, particularly relating to possible unintended consequences. These include the production of the potent greenhouse gas nitrous oxide, oxygen depletion in midwater when phytoplankton decompose, and stimulation of toxic algal blooms (Buesseler, 2012). Moreover, the ocean's capacity for carbon sequestration in low-iron ocean areas is a fraction of anthropogenic CO_2 production and such carbon sequestration is not permanent (Buesseler, 2012).

21.4 Global biogeochemical cycles

Nutrients are moved over vast distances by winds in the atmosphere and by the moving waters of streams and ocean currents. There are no boundaries, either natural or political. It is appropriate, therefore, to conclude this chapter by moving to an even larger spatial scale to examine global biogeochemical cycles. The earth system as a whole, including the biosphere, atmosphere, hydrosphere and lithosphere, is virtually closed, with the exception of negligible inputs via meteorites. This is why biogeochemical cycling plays such a critical role (Loreau, 2010a).

21.4.1 Hydrological cycle

The hydrological cycle is simple to conceive (although its elements are by no means always easy to measure) (Figure 21.20). The principal source of water is the oceans. Radiant energy makes water evaporate into the atmosphere, winds distribute it over the surface of the globe, and precipitation brings it down to earth (with a net movement of atmospheric water from oceans to continents), where it may be stored temporarily in soils, freshwater bodies, groundwater and icefields. Loss occurs from the land through evaporation and transpiration or as liquid flow through stream channels and groundwater aquifers, eventually to return to the sea. The major pools of water occur in the oceans (96.5% of the total for the biosphere), the ice and snow of polar ice caps, glaciers, snowfields and permafrost (1.75%), deep in the groundwater (1.69%) and in lakes, reservoirs, wetlands and rivers (0.015%). The proportion that is in transit at any time is very small – water draining through the soil, flowing along rivers and present as clouds, vapour and precipitation in the atmosphere constitutes only about 0.08% of the total. However, this small percentage plays a crucial role, both by supplying the requirements for survival of living organisms and for community productivity, and because so many chemical nutrients are transported with the water as it moves.

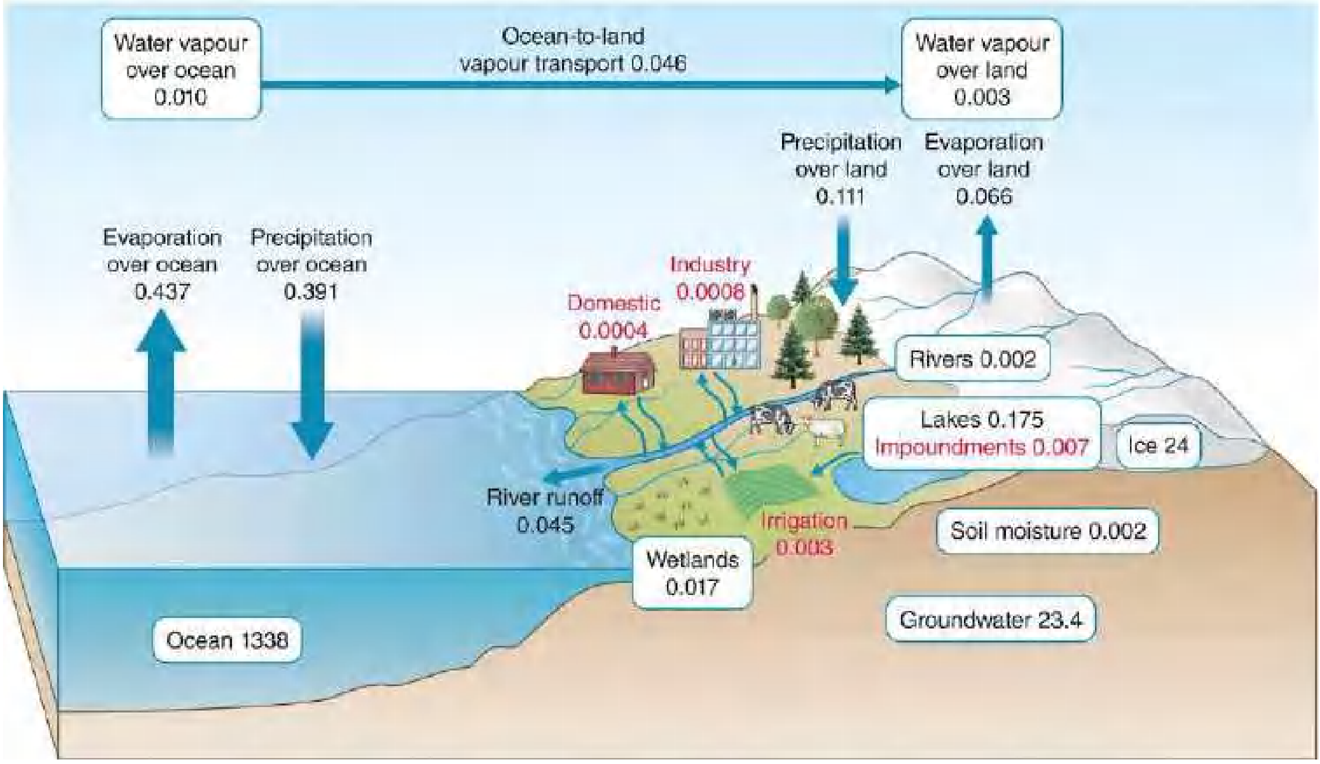


Figure 21.20 **The hydrological cycle.** Fluxes are shown as arrows (× $10^6\,km^3\,yr^{-1}$) and reservoirs in boxes (× $10^6\,km^3$). Some significant anthropogenic uses are shown in red.

Source: Redrawn from Chapin *et al.* (2011), based on estimates in Carpenter & Biggs (2009).

> plants live between two counterflowing movements of water

The hydrological cycle would proceed whether or not a biota was present. However, terrestrial vegetation can modify to a significant extent the fluxes that occur. Plants live between two counterflowing movements of water. One moves within the plant, proceeding from the soil into the roots, up through the stem and out from the leaves as transpiration. The other is deposited on the canopy as precipitation from where it may evaporate or drip from the leaves or flow down the stem to the soil. In the absence of vegetation, some of the incoming water would evaporate from the ground surface but the rest would enter the stream flow (via surface runoff and groundwater discharge). Vegetation can intercept water at two points on this journey, preventing some from reaching the stream and causing it to move back into the atmosphere by: (i) catching some in foliage from where it may evaporate; and (ii) preventing some from draining from the soil water by taking it up in the transpiration stream.

APPLICATION 21.5 Flood risk, groundwater exploitation and climate change

We have seen on a small scale how cutting down the forest in a catchment in Hubbard Brook can increase the throughput to streams of water together with its load of dissolved and particulate matter. It is small wonder that large-scale deforestation around the globe, usually to create new agricultural land, can lead to the loss of topsoil, nutrient impoverishment and increased severity of flooding. Future patterns of deforestation and afforestation need to be planned with flood risk in mind.

The turnover time to replenish atmospheric, river and soil water pools can be measured on scales from days to months. Groundwater aquifers, on the other hand, have average turnover times of hundreds of years. Many of these are already being overexploited to provide water for human use and some may not replenish under current conditions as they did in the past. Groundwater resources need much more careful management in the future.

Another major perturbation to the hydrological cycle will be global climate change resulting from human activities (see Section 22.2). The predicted temperature increase, with its concomitant changes to wind and weather patterns, can be expected to affect the hydrological cycle by causing melting of polar caps and glaciers, by changing patterns of precipitation and by influencing the details of evaporation, transpiration and stream flow. For example, many continental centres of agriculture, such as the Ukraine and mid-western USA, may be especially prone to drought, and future grain production may need to shift polewards to areas currently too cold to support intensive production (FAO, 2016).

process-based ecosystem models

A challenge for ecologists in many fields is to bridge the divide from the scale at which underlying biological processes operate (for example, the uptake and loss of water by individual plants) and the scale at which we judge the health of whole ecosystems (for example, whether regions or whole continents have sufficient water to sustain current levels of productivity). Models that aim to do so must reflect, on the one hand, the realities of plant physiology. But on the other hand, they cannot, for example, simply incorporate, one by one, the individual responses of the many plants in a region, from various species, in various habitats, and so on. Like all models, they must be simple – but not so simple as to be unhelpful. Process-based ecosystem models seek to bridge this kind of divide for the circulation of water, along with carbon, nitrogen and other elements. An example of such a model, formulated for forest catchments and developed using data from Hubbard Brook, is PnET-BGC (Figure 21.21; Gbondo-Tugbawa *et al.*, 2001). Its details are less important than the general principles behind its construction – namely that the complexities of the functioning of a large, multispecies community are reduced to a limited number of ‘compartments’ (in this case, 13 – wood, snow, soil solution, and so on) and the flow paths that connect them: gross photosynthesis, snowmelt, wood decay, and so on. A series of submodels then captures the behaviour of flow paths, or within compartments, utilising input parameters (temperature, wet and dry deposition each month, etc.) to generate outputs, either into other compartments or flow paths, or from the system as a whole.

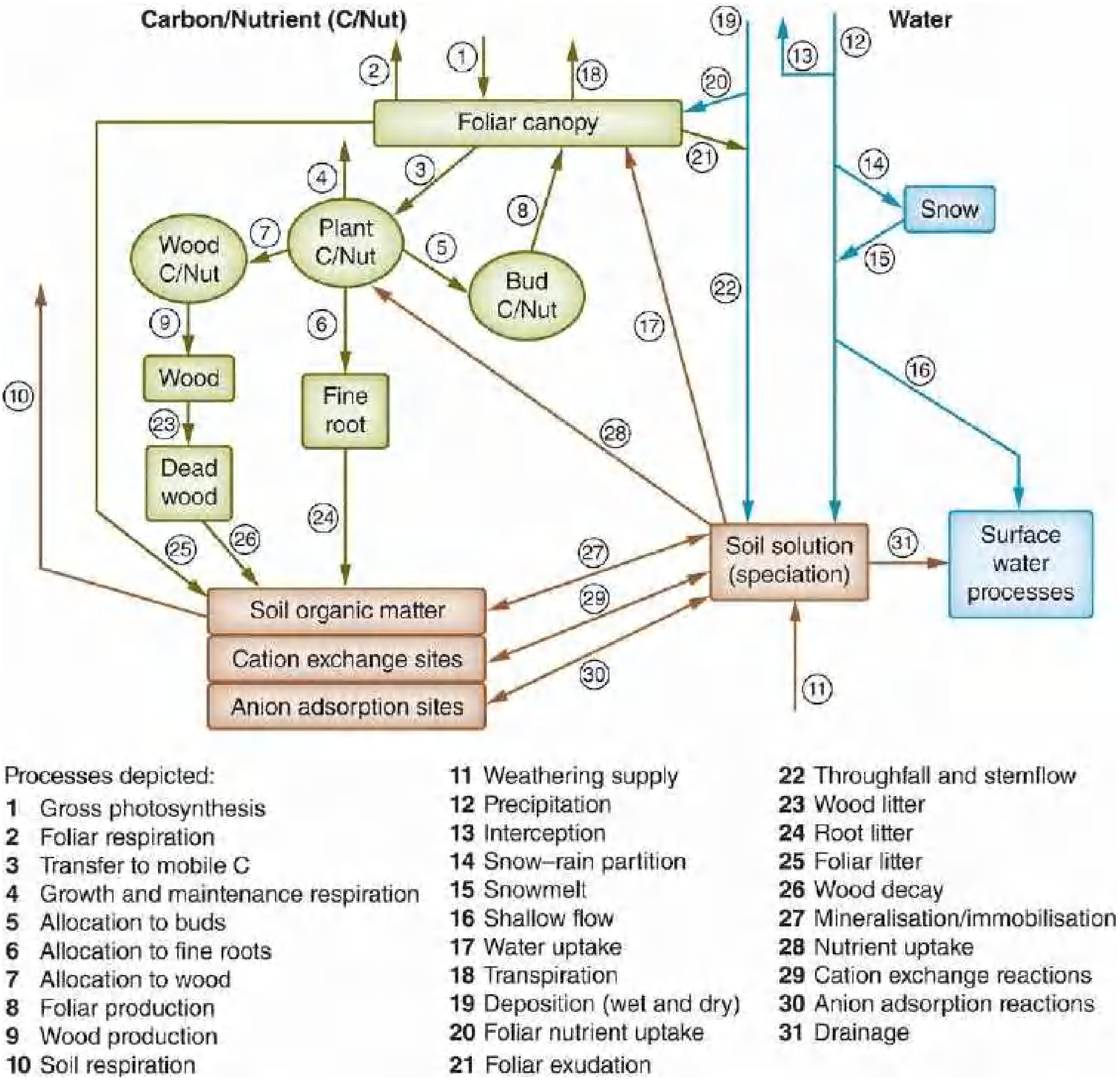


Figure 21.21 A process-based ecosystem model for the Hubbard Brook forest. The structure of the PnET-BGC process-based ecosystem model is illustrated, showing the compartments and flow paths of water, carbon and plant nutrients.

Source: After Gbondo-Tugbawa *et al.* (2001).

APPLICATION 21.6 Ecohydrological responses to predicted climate change

The model in Figure 21.21 allowed Campbell *et al*. (2011) to use Hubbard Brook as a test bed for investigating how the pattern of surface water supply may respond to future climate change. Application of PnET-BGC suggests that evapotranspiration may be expected to increase in future as conditions become warmer and wetter, but that this is likely to be largely offset by the increased precipitation, to the extent that stream flow may be little changed. However, a subsequent extension of the model to a range of forest sites in north-eastern USA – northern hardwoods, spruce-fir and central hardwoods, and encompassing a range of soil conditions and historical land disturbance – suggests that Hubbard Brook may not be typical (Figure 21.22). The extent of future climate change is of course uncertain, and so for each of seven sites, eight scenarios were simulated: four atmosphere–ocean general circulation models each run with both high and low projected future emissions of greenhouse gases (see Section 22.2). Hubbard Brook (a northern hardwood forest) was again predicted to experience little change in stream flow, but for Cone Pond, a spruce-fir forest, the prediction was a massive increase in stream flow, associated with the much lower transpiration demand of conifers, while for Fernow Forest, a central hardwood forest, a decrease seems possible. As the authors of the study note, in the face of such variability, dominant vegetation and soil type are important in determining hydrological responses to future climate change.

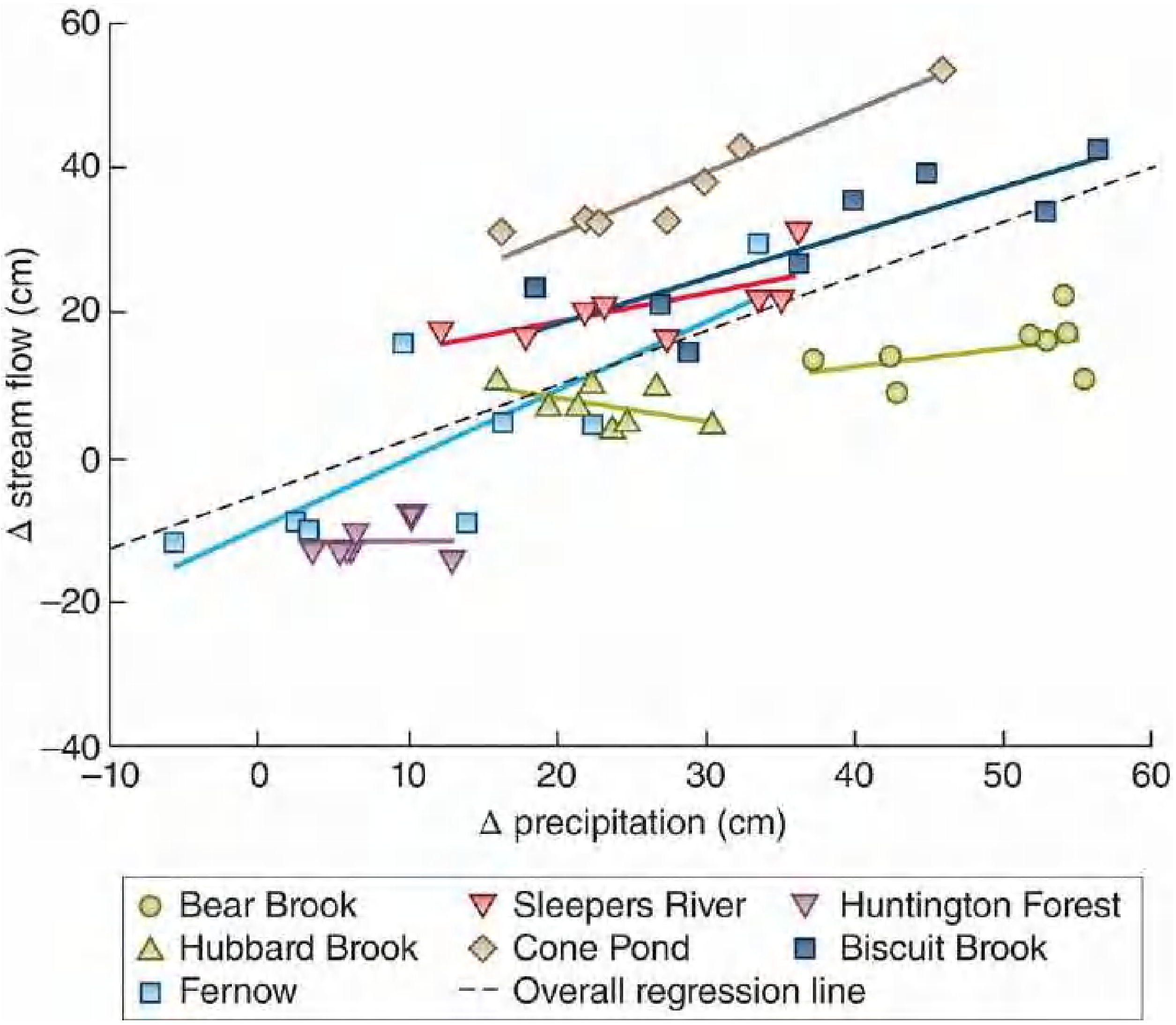


Figure 21.22 **Projected effects of climate change in different forests.** The relationships for seven forest sites in north-eastern USA, as indicated, between the projected change in stream flow and annual precipitation, from eight climate change scenarios (the eight data points for each site), comparing the period 2070–2100 to a reference period of 1970–2000. Relationships were significant (and positive) for Biscuit Brook, Cone Pond, Fernow and Sleepers River.

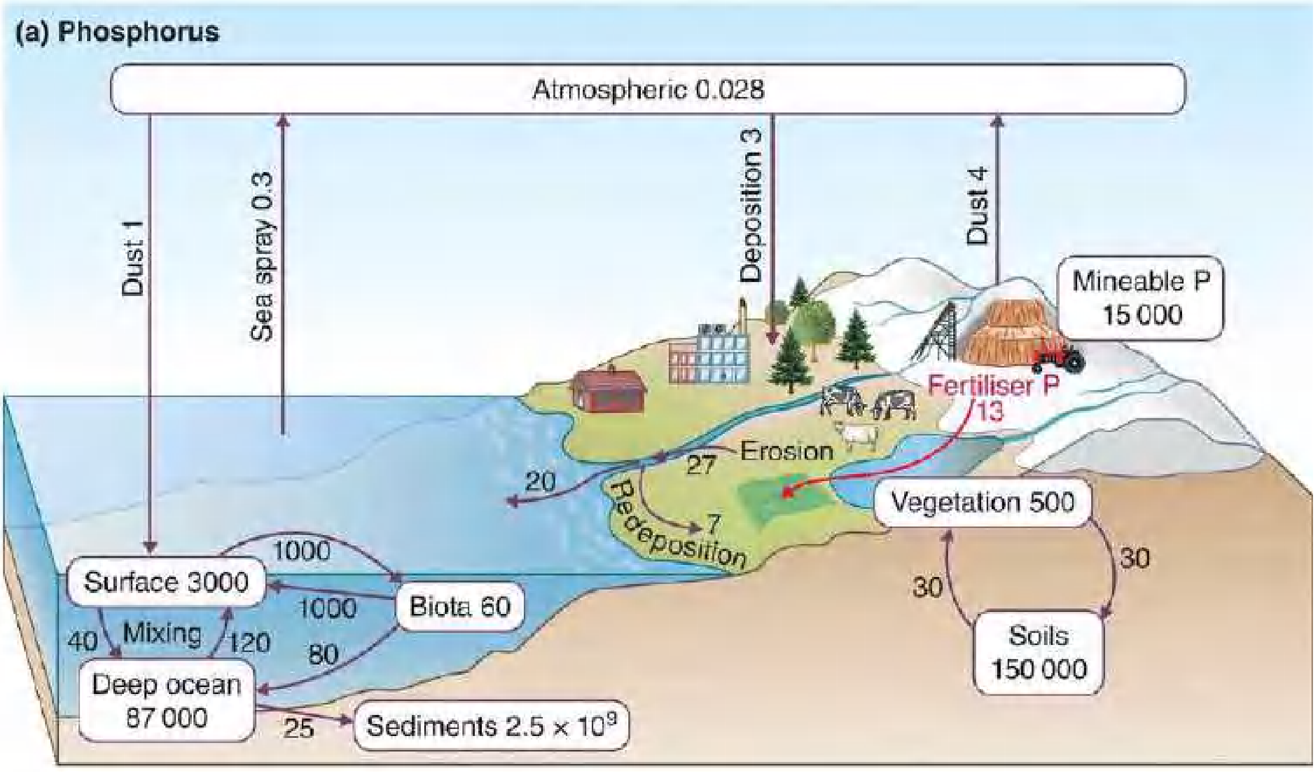
Source: After Pourmokhtarian *et al.* (2017).

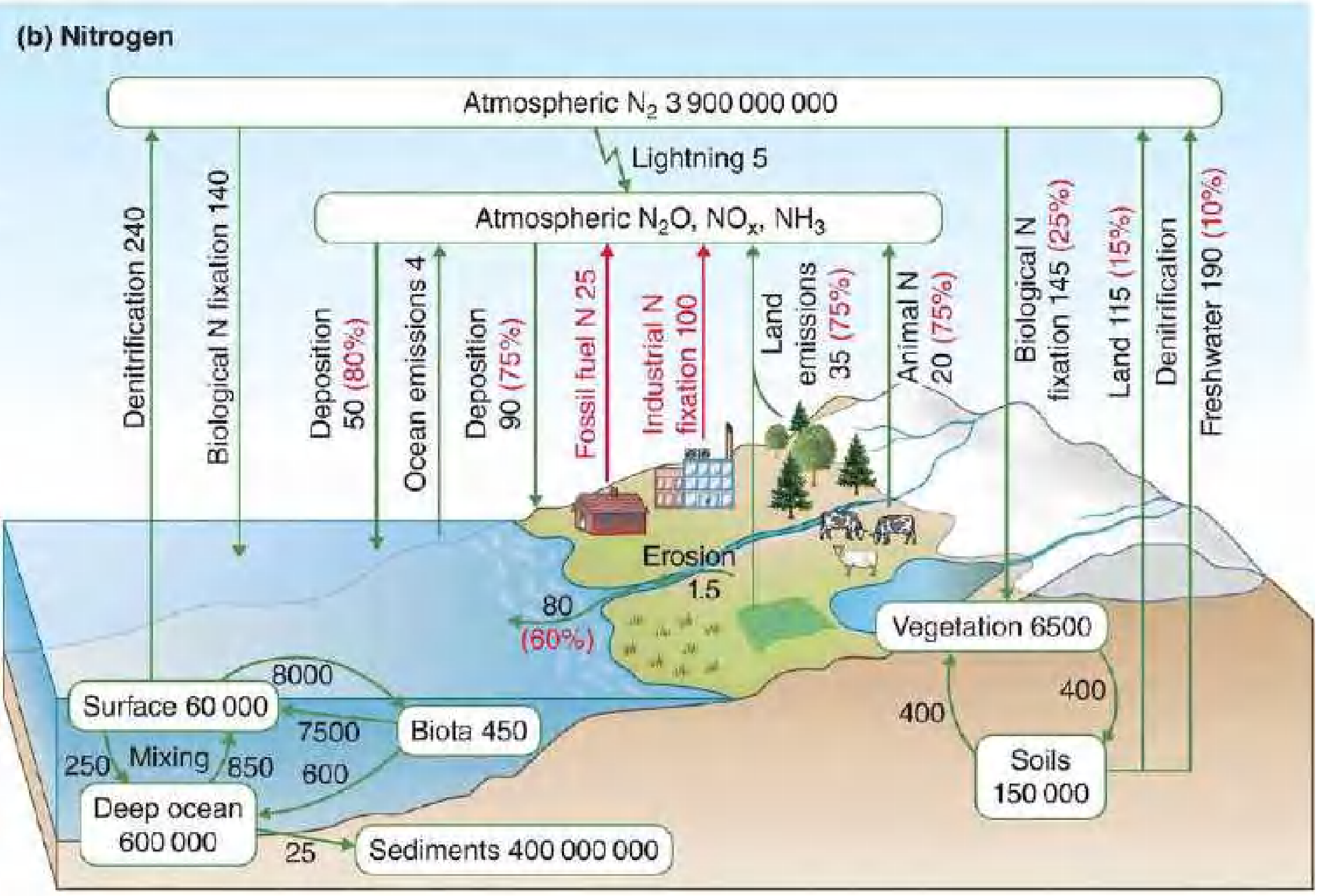
21.4.2 Phosphorus cycle

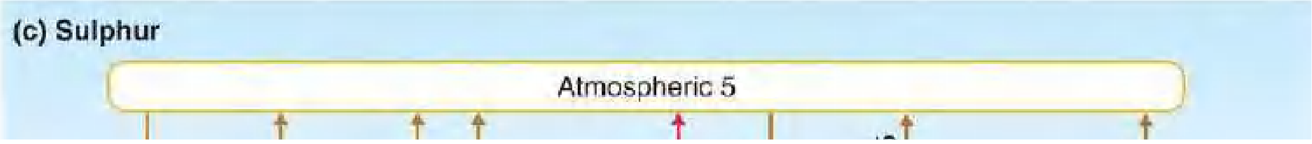
The biotas of both terrestrial and aquatic habitats obtain some of their nutrient elements predominantly via the weathering of rock. This is the case for phosphorus. Carbon and nitrogen, on the other hand, derive mainly from the atmosphere – the first from CO_2 and the second from gaseous nitrogen. Sulphur derives from both atmospheric and lithospheric sources. Here we consider phosphorus, and in the following sections, nitrogen, sulphur and carbon in turn, and ask how human activities upset the global biogeochemical cycles of these biologically important elements.

phosphorus derives mainly from the weathering of rocks

The principal stocks of phosphorus occur in the water of the soil, rivers, lakes and oceans and in rocks and ocean sediments. The phosphorus cycle may be described as an ‘open’ cycle because of the general tendency for mineral phosphorus to be carried from the land inexorably to the oceans, mainly in rivers, but also to smaller extents in groundwater, or via volcanic activity and atmospheric fallout, or through abrasion of coastal land. The cycle may alternatively be termed a ‘sedimentary cycle’ because ultimately phosphorus becomes incorporated in ocean sediments (Figure 21.23a). We can unravel an intriguing story that starts in a terrestrial catchment area. A typical phosphorus atom, released from the rock by chemical weathering, may enter and cycle within the terrestrial community for years, decades or centuries before it is carried via groundwater into a stream, where it takes part in the nutrient spiralling described in Section 21.3.1. Within a short time of entering the stream (weeks, months or years), the atom is carried to the ocean. It then makes, on average, about 100 round trips between the surface and deep waters, each lasting perhaps 1000 years. During each trip, it is taken up by organisms that live at the ocean surface, before eventually settling into the deep again. On average, on its 100th descent (after 10 million years in the ocean) it fails to be released as soluble phosphorus, but instead enters the bottom sediment in particulate form. Perhaps 100 million years later, the ocean floor is lifted up by geological activity to become dry land. Thus, our phosphorus atom will eventually find its way back via a river to the sea, and to its existence of cycle (biotic uptake and decomposition) within cycle (ocean mixing) within cycle (continental uplift and erosion).


(a) Phosphorus
Atmospheric 0.028
Dust 1
Sea spray 0.3
Deposition 3
Dust 4
Mineable P 15 000
Fertiliser P 13
Erosion
27
20
Redeposition
7
Vegetation 500
30
30
Soils 150 000
Surface 3000
1000
1000
Biota 60
40
Mixing
120
80
Deep ocean 87 000
25
Sediments 2.5 × 10^9


(b) Nitrogen
Atmospheric N$_2$ 3 900 000 000
Lightning 5
Atmospheric N$_2$O, NO$_x$, NH$_3$
Denitrification 240
Biological N fixation 140
Deposition 50 (80%)
Ocean emissions 4
Deposition 90 (75%)
Fossil fuel N 25
Industrial N fixation 100
Land emissions 35 (75%)
Animal N 20 (75%)
Biological N fixation 145 (25%)
Land 115 (15%)
Denitrification
Freshwater 190 (10%)
Erosion 1.5
80 (60%)
Vegetation 6500
400
400
Soils 150 000
Surface 60 000
8000
7500
Biota 450
250
Mixing
850
600
Deep ocean 600 000
25
Sediments 400 000 000


(c) Sulphur
Atmospheric 5

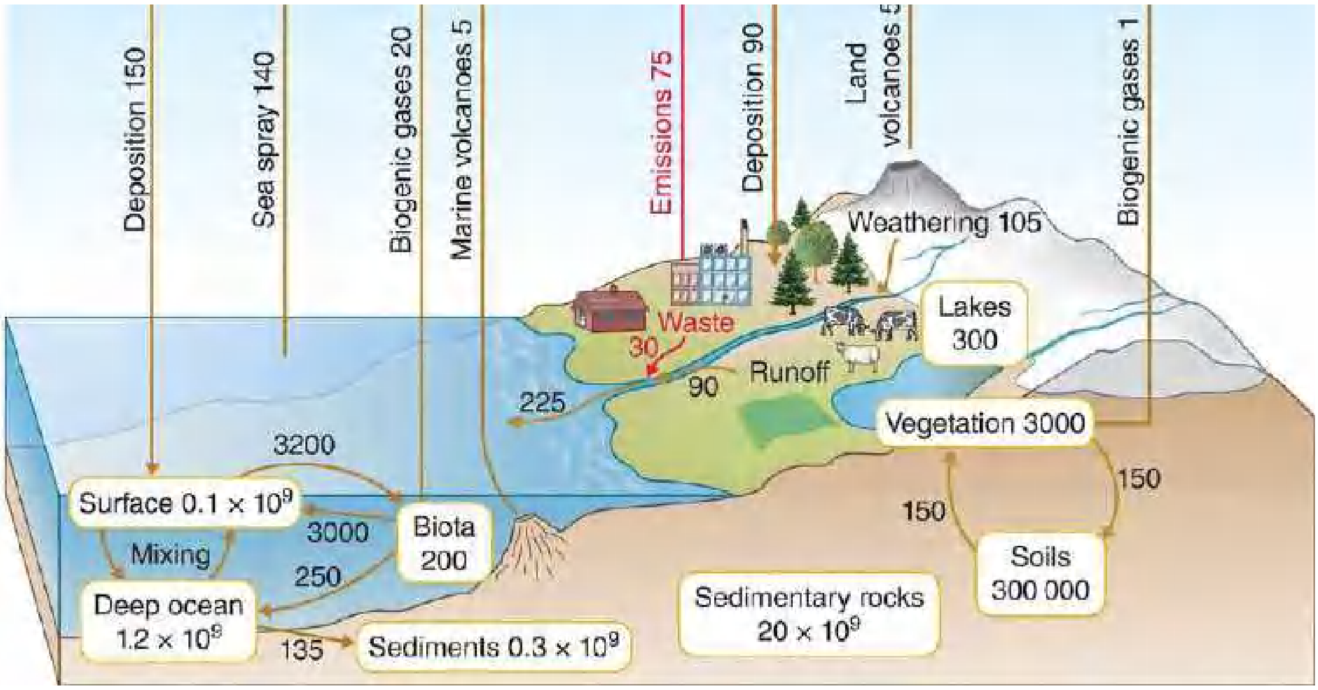


Figure 21.23 **Global cycles of (a) phosphorus, (b) nitrogen and (c) sulphur.** Major pools are shown in boxes (Tg) and fluxes as arrows (Tg year^{-1}). Anthropogenic contributions to fluxes are shown in brackets in red or as red arrows where the human contribution is 100%.

Source: Redrawn from Chapin *et al.* (2011), based on estimates from a number of sources.

APPLICATION 21.7 Human activities contribute the majority of phosphorus in inland waters

Human activities affect the phosphorus cycle in a number of ways. Marine fishing transfers about 50 Tg of phosphorus from the ocean to the land each year. Since the total oceanic pool of phosphorus is around 90 000 Tg, this reverse flow has negligible consequences for the ocean compartment. However, phosphorus from the fish catch will eventually move back through the rivers to the sea and, thus, fishing contributes indirectly to increased concentrations in inland and coastal waters. Between 10 and 15 Tg of phosphorus are dispersed annually as fertiliser (some derived from the marine fish catch) and a further 2 or 3 Tg as an additive to domestic detergents. Much of the former reaches the aquatic system as agricultural runoff, whereas the latter arrives in domestic sewage. In addition, deforestation and many forms of land cultivation increase erosion in catchment areas and contribute a large proportion of the estimated 27 Tg year^{-1} of phosphorus in runoff, and some more as wind-dispersed dust. All told, the rate of phosphorus fluxes globally is now about three times higher than before the industrial and agricultural revolutions.

The consequences of phosphorus enrichment of waterways and coastal seas (usually in concert with nitrate pollution) were discussed in Section 21.1.3. Eutrophication effects are localised, in the sense that only those waters draining the relevant catchment areas are affected. But the problem is pervasive and worldwide.

21.4.3 Nitrogen cycle

the nitrogen cycle has an atmospheric phase of overwhelming importance

The atmospheric phase is predominant in the global nitrogen cycle, in which nitrogen fixation and denitrification by microbial organisms are by far the most important (Figure 21.23b). Atmospheric nitrogen is also fixed by lightning discharges during storms and reaches the ground as nitric acid dissolved in rainwater, but only about 1–2% of fixed nitrogen derives from this pathway. Organic forms of nitrogen are also widespread in the atmosphere, some of which results from the reaction of hydrocarbons and oxides of nitrogen in polluted air masses. In addition, amines and urea are naturally injected as aerosols or gases from terrestrial and aquatic ecosystems; and a third source consists of bacteria and pollen (Neff *et al.*, 2002). The magnitude of the nitrogen flux in stream flow from terrestrial to aquatic communities may be relatively small, but it is by no means insignificant for the aquatic systems involved. This is because nitrogen is one of the two elements (along with phosphorus) that most often limits plant growth. Finally, there is a small annual loss of nitrogen to ocean sediments.

APPLICATION 21.8 Humans impact on the nitrogen cycle in diverse ways

Human activities have a variety of far-reaching effects on the nitrogen cycle. Deforestation, and land clearance in general, leads to substantial increases in nitrate flux in stream flow and N_2O losses to the atmosphere (see Section 21.2.2). In addition, technological processes yield fixed nitrogen as a by-product of internal combustion and in the production of fertilisers. The agricultural practice of planting legume crops, with their root nodules containing nitrogen-fixing bacteria, contributes further to nitrogen fixation. In fact, the amount of fixed nitrogen produced by these human activities is of the same order of magnitude as that produced by natural nitrogen fixation. The production of nitrogenous fertilisers (projected to rise to 165 Tg year^{-1} by 2050; Galloway *et al.*, 2004) is of particular significance because an appreciable proportion of fertiliser added to land finds its way into streams, lakes and coastal seas. The artificially raised concentrations of nitrogen contribute to the process of cultural eutrophication of water bodies.

Human activities impinge on the atmospheric phase of the nitrogen cycle too. For example, fertilisation of agricultural soils leads to increased runoff as well as an increase in denitrification, and the handling and spreading of manure in areas of intensive animal husbandry releases substantial amounts of ammonia to the atmosphere. Atmospheric NO_x, produced by the burning of fossil fuels, is converted, within days, to nitric acid, which contributes, together with NH_3, to the acidity of precipitation within and downwind of industrial regions. Sulphuric acid is the other culprit.

21.4.4 Sulphur cycle

the sulphur cycle has atmospheric and lithospheric phases of similar magnitude

In the global phosphorus cycle we have seen that the lithospheric phase is predominant (Figure 21.23a), whereas the nitrogen cycle has an atmospheric phase of overwhelming importance (Figure 21.23b). Sulphur, by contrast, has atmospheric and lithospheric phases of similar magnitude (Figure 21.23c).

Three natural biogeochemical processes release sulphur to the atmosphere: (i) the formation of the volatile compound dimethylsulphide (DMS) (by enzymatic breakdown of an abundant compound in phytoplankton – dimethylsulphonioproprionate); (ii) anaerobic respiration by sulphate-reducing bacteria; and (iii) volcanic activity. Total biological release of sulphur to the atmosphere is estimated to be 21 Tg year^{-1}, and of this more than 90% is in the form of DMS. Most of the remainder is produced by sulphur bacteria that release reduced sulphur compounds, particularly H_2S, from waterlogged bog and marsh communities and from marine communities associated with tidal flats. Volcanic production provides a further 10 Tg year^{-1} to the atmosphere. A reverse flow from the atmosphere involves oxidation of sulphur compounds to sulphate, which returns to earth as both wetfall and dryfall.

The weathering of rocks provides about half the sulphur draining off the land into rivers and lakes, the remainder deriving from atmospheric sources. On its way to the ocean, a proportion of the available sulphur (mainly dissolved sulphate) is taken up by plants, passed along food chains and, via decomposition processes, becomes available again to plants. However, in comparison to phosphorus and nitrogen, a much smaller fraction of the flux of sulphur is involved in internal recycling in terrestrial and aquatic communities. Finally, there is a continuous loss of sulphur to ocean sediments, mainly through abiotic processes such as the conversion of H_2S, by reaction with iron, to ferrous sulphide (which gives marine sediments their black colour).

APPLICATION 21.9 Sulphur and acid rain

The combustion of fossil fuels is the major human perturbation to the global sulphur cycle (coal contains 1–5% sulphur and oil contains 2–3%). The SO_2 released to the atmosphere is oxidised and converted to sulphuric acid in aerosol droplets, mostly less than 1 μm in size. Anthropogenic releases of sulphur to the atmosphere comprise more than 20% of the total, but whereas natural inputs are spread fairly evenly over the globe, most human inputs are concentrated in and around industrial zones where they can contribute up to 90% of the total. Concentrations decline progressively downwind from sites of production, but they can still be high at distances of several hundred kilometres. Thus, one nation can export its SO_2 to other countries; concerted international political action is required to alleviate the problems that arise.

Water in equilibrium with CO_2 in the atmosphere forms dilute carbonic acid with a pH of about 5.6. However, the pH of acid precipitation (rain or snow) can average well below 5.0, and values as low as 2.4 have been recorded in Britain, 2.8 in Scandinavia and 2.1 in the USA. The emission of SO_2 often contributes most to the acid rain problem, though together NO_x and NH_3 account for 30–50% of the problem (Mooney *et al.*, 1987 ; Sutton *et al.*, 1993).

We saw previously how a low pH can drastically affect the biotas of streams and lakes (Chapter 2). Acid rain (see Section 2.8) has been responsible for the extinction of fish in thousands of lakes, particularly in Scandinavia. In addition, a low pH can have far-reaching consequences for forests and other terrestrial communities. It can affect plants directly, by breaking down lipids in foliage and damaging membranes, or indirectly, by increasing leaching of some nutrients from the soil and by rendering other nutrients unavailable for uptake by plants. It is important to note that some perturbations to biogeochemical cycles arise through indirect, 'knock-on' effects on other biogeochemical components. For example, alterations in the sulphur flux in themselves are not always damaging to terrestrial and aquatic communities, but the effect of sulphate's ability to mobilise metals such as aluminium, to which many organisms are sensitive, may indirectly lead to changes in community composition.

Provided that governments show the political will to reduce emissions of SO_2 and NO_x (for example, by making use of techniques already available to remove sulphur from coal and oil), the acid rain problem should be controllable. Indeed, reductions in sulphur emissions have occurred in various parts of the world while they have increased in others.

21.4.5 Carbon cycle

opposing forces of photosynthesis and respiration drive the global carbon cycle

Photosynthesis and respiration are the two opposing processes that drive the global carbon cycle. It is predominantly a gaseous cycle, with CO_2 as the main vehicle of flux between the atmosphere, hydrosphere and biota (Figure 21.24).

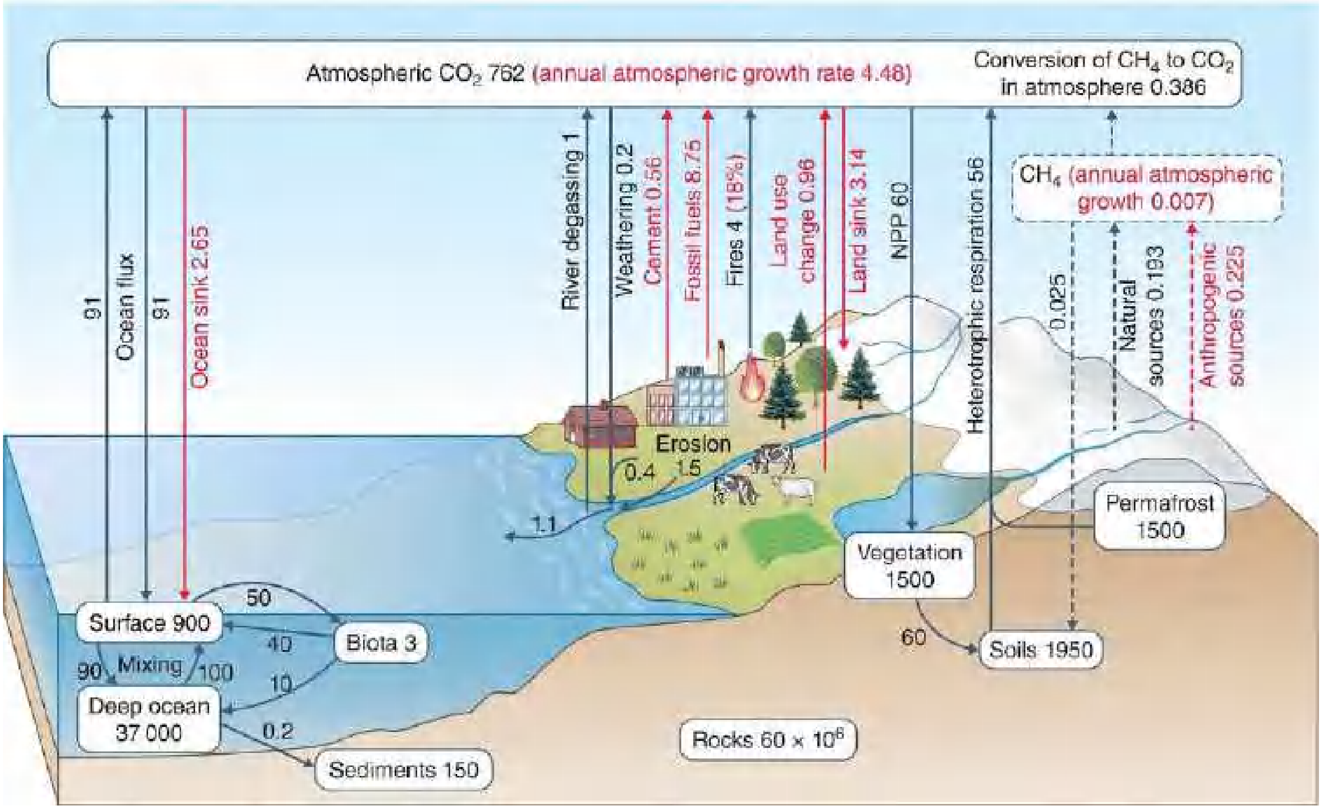


Figure 21.24 **The global carbon cycle.** Major pools are shown in boxes (Pg C) and fluxes as arrows (Pg C year^{-1}); 1 petagram = 10^{15} g. Anthropogenic contributions to fluxes are shown in brackets in red, or as red arrows where the human contribution is 100%. Dashed lines illustrate the much smaller pool (box) and fluxes of methane (CH_4), most of which is converted to CO_2 in the atmosphere.

Source: Based partly on Chapin *et al.* (2011), with updated estimates for anthropogenic sources and related sinks from Le Quéré *et al.* (2016), and Saunois *et al.* (2016). These publications collate estimates from a number of sources.

Terrestrial plants use atmospheric CO_2 as their carbon source for photosynthesis, whereas aquatic plants use dissolved carbonates (i.e. carbon from the hydrosphere). The two subcycles are linked by exchanges of CO_2 between the atmosphere and oceans as follows:

$$\text{atmospheric } CO_2 \rightleftharpoons \text{dissolved } CO_2$$

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \text{ (carbonic acid).}$$

In addition, carbon finds its way into inland waters and oceans as bicarbonate resulting from weathering (carbonation) of calcium-rich rocks such as limestone and chalk:

$$CO_2 + H_2O + CaCO_3 \rightleftharpoons CaH_2(CO_3)_2$$

Respiration by plants, animals and microorganisms releases the carbon locked in photosynthetic products back to the atmospheric and hydrospheric carbon compartments.

Historically, the lithosphere played only a minor role; fossil fuels lay as dormant reservoirs of carbon until human intervention in recent centuries. Anthropogenic emissions of CO_2 from the burning of fossil fuels, as well as cement production and land-use conversion (biomass burning and increased decomposition) represent additions to the atmosphere, counteracted only partially by extra dissolution of CO_2 in the oceans (the ocean sink) and absorption in enhanced NPP due to the fertilising effect of increased CO_2 (the land sink). A recent estimate of the overall net annual increase in atmospheric CO_2 is 4.48 ±0.11 Pg C year^{-1} (Le Quéré *et al.*, 2016), a build up with very significant consequences for global climate (see below).

Under anaerobic conditions, mineralization of organic compounds releases inorganic carbon to the atmosphere as methane (CH_4). Of the naturally produced CH_4, 72% comes from wetlands (0.125 Pg C year^{-1}) and much of the rest from fermentation in the guts of animals, including termites and ruminants. Anthropogenic sources of CH_4 are a little larger than natural sources (0.225 Pg C year^{-1}), deriving from combustion of coal, oil and gas (0.079 Pg C year^{-1}), biomass burning (both natural and anthropogenic, 0.025 Pg C year^{-1}), landfills and waste, rice agriculture and fermentation in the guts of cattle (0.141 Pg C year^{-1}). Note that CH_4 pools (atmospheric concentration by volume of CH_4 is 1.8 ppm, but 400 ppm for CO_2) and fluxes are very much smaller than those of CO_2 (Pg). Total CH_4 sources sum approximately to 0.421 Pg C year^{-1}, while sinks account for about 0.410 Pg C year^{-1}, the majority converted to CO_2 in the atmosphere and a small proportion removed by methanotrophic bacteria in the soil: thus, the annual build-up of CH_4 in the atmosphere is just over 0.007 Pg C year^{-1} (Saunois *et al.*, 2016).

The concentration of CO_2 in the atmosphere has increased from about 280 ppm in 1750 to more than 400 ppm today and it is still rising. The pattern of increase recorded at the Mauna Loa Observatory in Hawaii since 1958 is shown in Figure 21.25. (Note the cyclical decreases in CO_2 associated with higher rates of photosynthesis during summer in the northern hemisphere – reflecting the fact that most of the world’s landmass is north of the equator.)

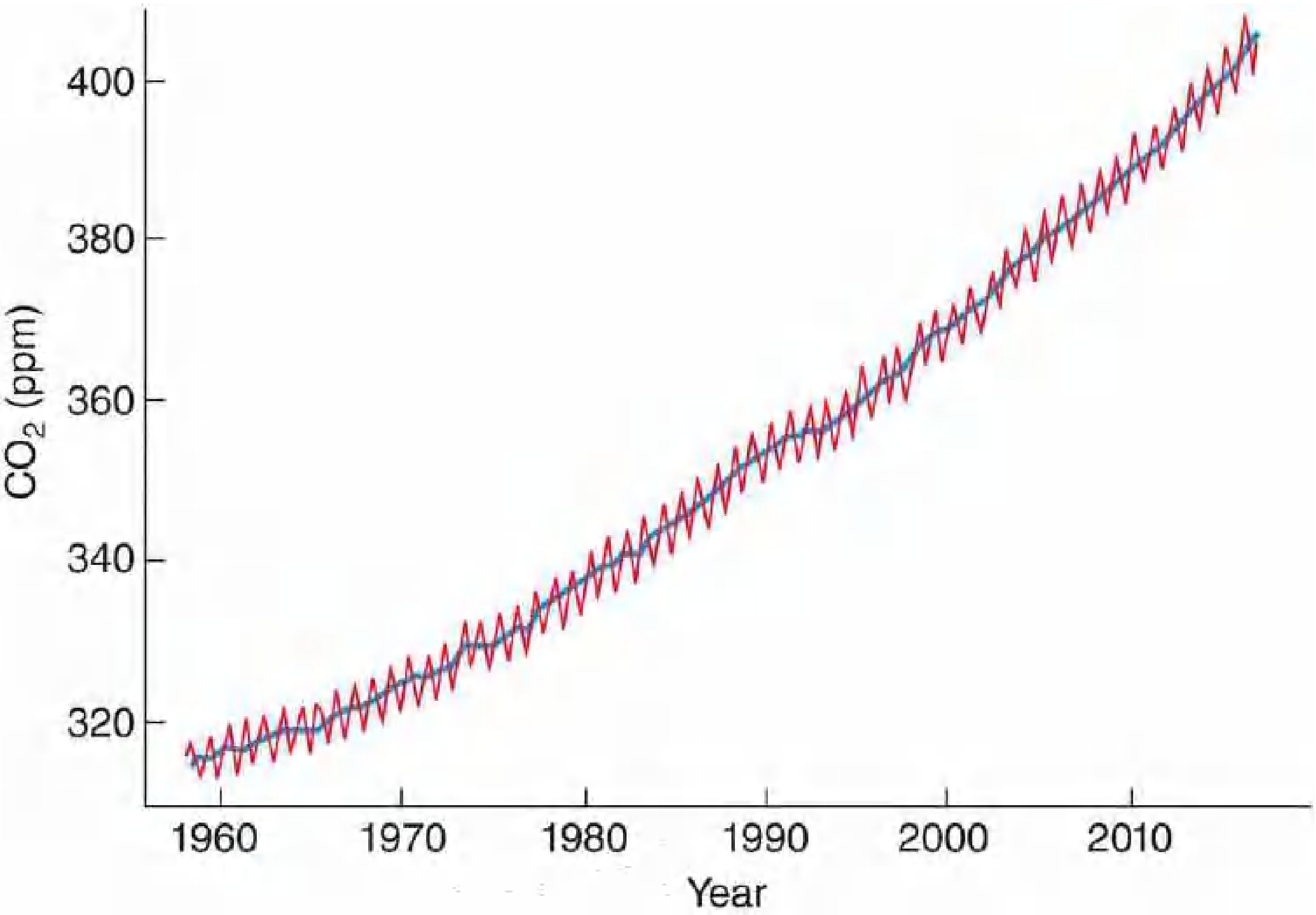


Figure 21.25 Concentration of atmospheric carbon dioxide at the Mauna Loa Observatory, Hawaii. Note the seasonal cycle (resulting from changes in photosynthetic rate) and the long-term increase that is due largely to the burning of fossil fuels.

Source: Courtesy of the Earth System Research Laboratory of the National Oceanic and Atmospheric Administration.

APPLICATION 21.10 Climate change and ocean acidification

Climate scientists are no longer in any doubt that CO_2 in the atmosphere is the major driver of global warming (see Section 2.9.2). Since about 1960, fossil fuel combustion has increasingly dominated the flux of CO_2 into the atmosphere (Figure 21.26). Not all this extra CO_2 accumulates in the atmosphere. A substantial proportion has been soaked up by terrestrial vegetation, fuelled in particular by the fertilising effect of the higher concentrations of atmospheric CO_2, a fundamental resource for photosynthesis. Moreover, a similar proportion has been absorbed by the oceans. Without these CO_2 sinks the concentration in the atmosphere would be rising much faster, with a concomitant acceleration in the rate of warming.

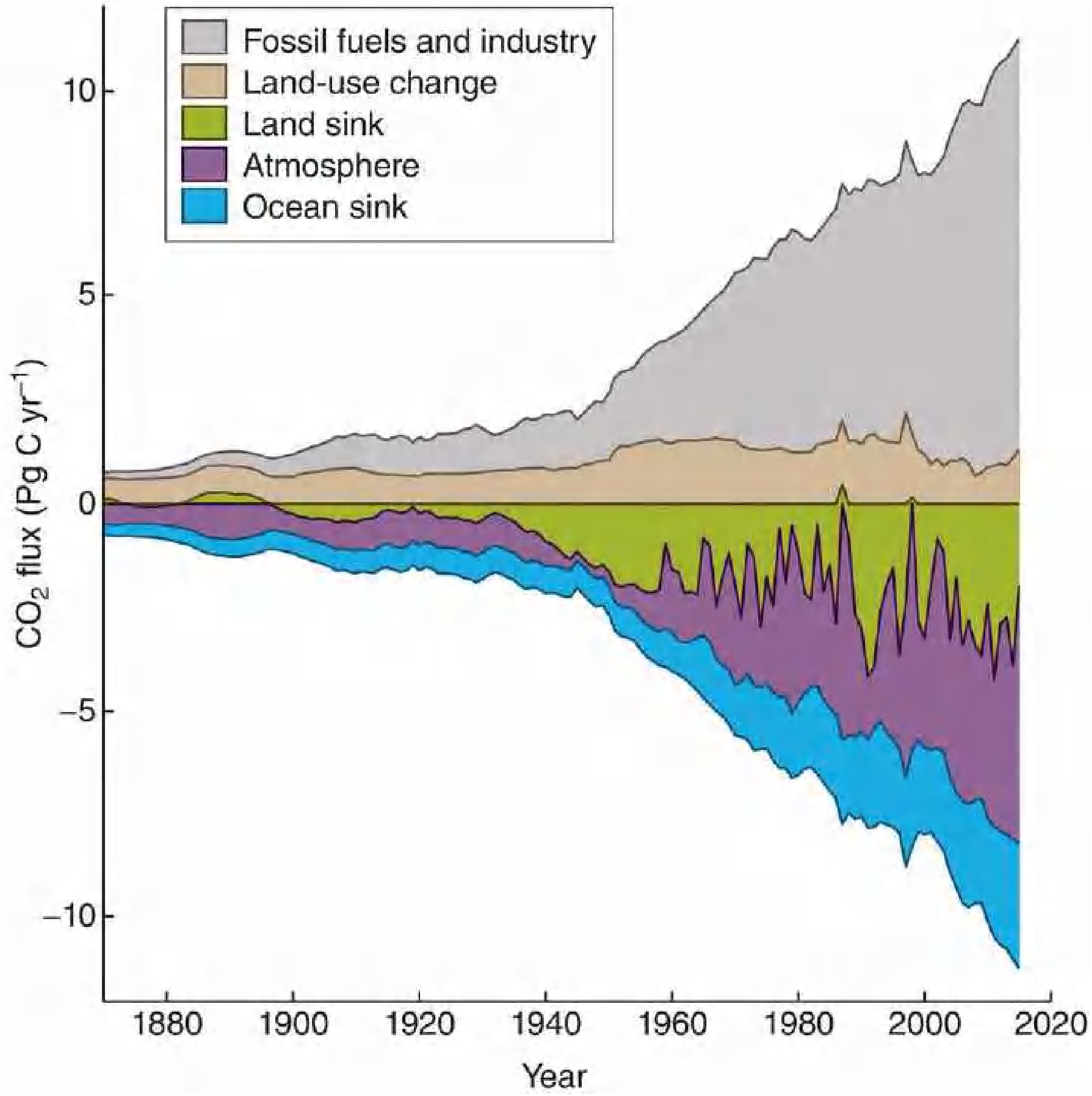


Figure 21.26 The global fluxes of carbon dioxide to the atmosphere (sources) and from the atmosphere (sinks) from 1870 to 2015.

Source: Redrawn from Le Quéré *et al.* (2016).

Other gases also contribute to global warming, of which CH_4 is of particular significance. This absorbs infrared radiation at wavelengths that CO_2 does not, and as a result, per molecule it is far more potent as a greenhouse gas. On the other hand, the annual accumulation of CH_4 in the atmosphere is much lower and it persists there for a much shorter period (about a decade) than CO_2 (thousands of years). Nevertheless, a unit of CH_4 has a global warming potential some 21 times greater than that of CO_2. Nitrous oxide is another highly potent greenhouse gas with a long residence time, and a global warming potential 265–298 times that of CO_2, but it is present at much smaller concentrations.

The deleterious effects of increasing CO_2 concentrations do not end with global warming. As ocean waters take up carbon dioxide, they become more acidic and adversely affect a diversity of marine organisms with shells or other structures composed of carbonate minerals. There is thus the potential to cause serious disruption of many marine ecosystems, including coral reefs.

We return to the many dimensions of the ecological challenges facing mankind in the final chapter.

Chapter 22

Ecology in a Changing World

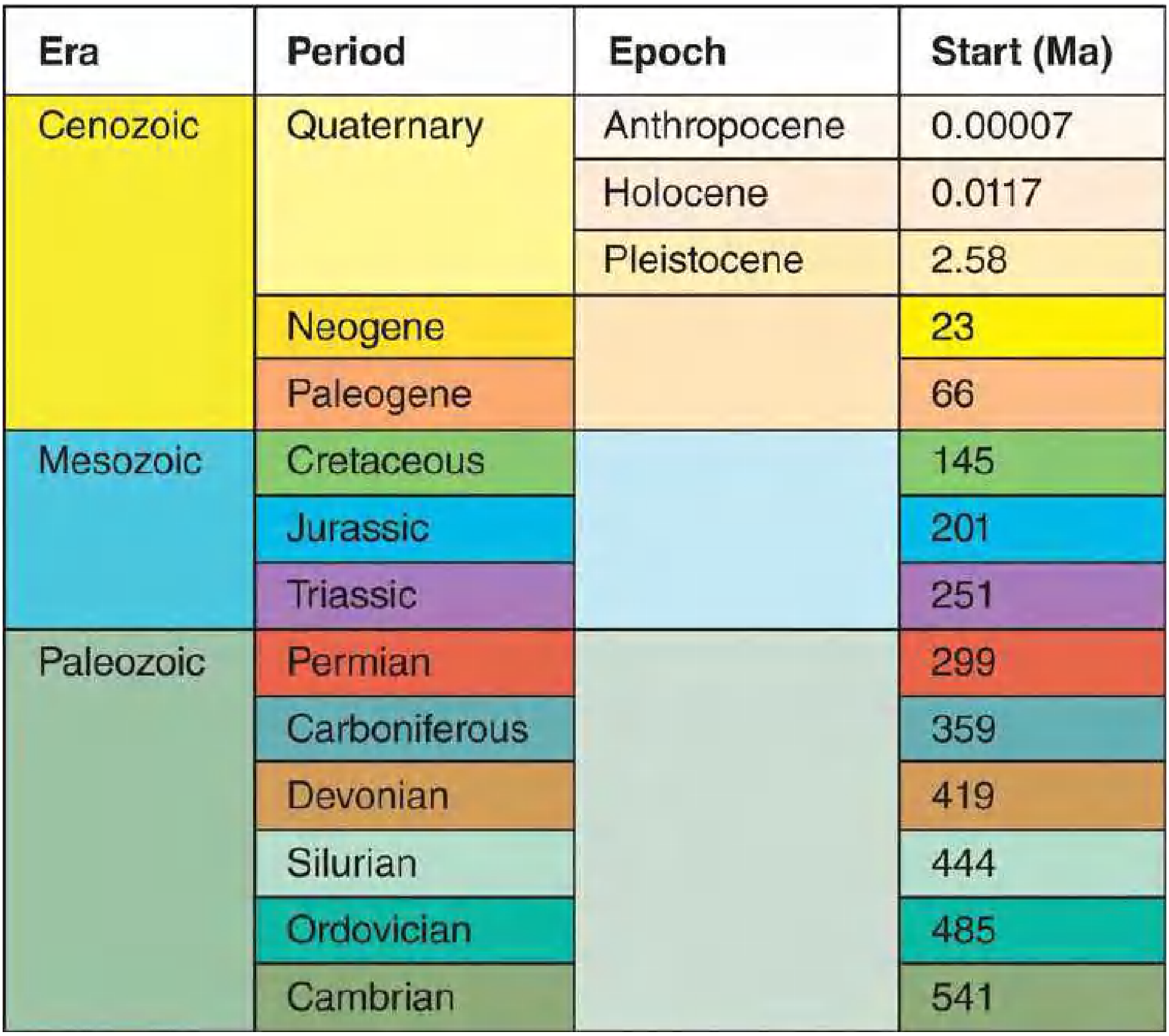
22.1 Introduction

Through the chapters of this book we have seen repeatedly that ecological systems are in a continual state of flux. At the small scale this may be driven by the dynamics of populations, with, for example, predator and prey affecting the densities of each other. Or externally driven disturbances may be responsible, such as the arrival of a new species (from local migration, or swept across seas or mountain tops by extreme weather events). Or ecosystems may change as a result of wildfire, drought, flood, hurricane or the impact of ungulate hooves in a patch of savanna. At a larger scale, ecosystems across the globe have been affected by gross climatic changes, such as the waxing and waning of ice ages, or periods of volcanic activity, or the impact of an asteroid. For most of the world's history, all these drivers of change were haphazard and natural.

But then *Homo sapiens* arrived on the scene, dispersed and multiplied, moving through phases of agricultural, industrial, nuclear and technological revolution. Since 1950 our population has increased more than 2.5-fold, and although the rate of increase is slowing, the world population is projected to increase to between 9.4 and 12.7 billion by 2100 (UN, 2019; and see Figure 5.20). While the haphazard factors are still operating in ecosystems, human actions have become the main driver of environmental change everywhere on the planet.

the human epoch – the Anthropocene

Indeed, since the turn of the millennium, support has been growing to recognise a new geological epoch that acknowledges our species' impacts. Once ratified by the International Commission on Stratigraphy (ICS), the 'Anthropocene' will represent the section of geological history during which people have become a major geological (and, undoubtedly, ecological) force. Geological eras, periods and epochs are determined by signals in the Earth's stratigraphy – evidence contained in layers of rocks, ocean sediments, ice cores, and so on, with particular reference to the appearance or disappearance of fossils. Thus, the 66 million years since the extinction of the dinosaurs is known as the Cenozoic era, while, within that, the Quaternary period occupies the last 2.58 million years when the planet cycled in and out of ice ages (Figure 22.1). The thin slice of the Quaternary since the end of the last cold snap (about 11 700 years) was designated in 2008 as the Holocene epoch, marked by a chemical signal of warming in a core of ice near the centre of Greenland, and mirrored by other signals in lake and marine sediments around the world.



Era	Period	Epoch	Start (Ma)
Cenozoic	Quaternary	Anthropocene	0.00007
		Holocene	0.0117
		Pleistocene	2.58
	Neogene		23
	Paleogene		66
Mesozoic	Cretaceous		145
	Jurassic		201
	Triassic		251
Paleozoic	Permian		299
	Carboniferous		359
	Devonian		419
	Silurian		444
	Ordovician		485
	Cambrian		541

Figure 22.1 **The last 541 million years of the Earth's geological history.** Once ratified by the International Commission on Stratigraphy, the slither of geological history during which people have become a major force will be named the Anthropocene epoch. For simplicity, the Quaternary is the only period for which epochs are shown. Ma, Millions of years ago.

Since 2009, the ICS 'Working Group on the Anthropocene' has been analysing evidence, reaching a majority opinion that this most recent slither of the Quaternary is stratigraphically real and recommending that the Anthropocene epoch be based on a mid-20th century starting point (around 1950). The stratigraphic proxies most likely to be used as a primary marker are radionucleotides associated with nuclear arms testing, with possible secondary markers of plastic, carbon isotope patterns and industrial ash (Zalasiewicz *et al.*, 2017). While the ecological impacts of human activities began much earlier (e.g. effects of agriculture 5000 years ago, or of the industrial revolution of the 19th century) these, unlike nuclear fallout, have not left unambiguous stratigraphic signals. Moreover, 1950 marks what might be called a great acceleration in human impacts, when the mass construction of dams, accelerating human consumption of water and production of fertilisers, plastics, aluminium, leaded petrol and concrete all began to leave signals in the sedimentary record (Monastersky, 2015).

a biodiversity crisis – the world's sixth mass extinction?

The vast majority of species that ever existed are now extinct. For most of history, species disappeared at a rate of a handful every 100 000 years. But during certain periods, extinction rates rose well above the background level and, in particular, five mass extinctions are recognised in the geological record. In the time spanning the late Ordovician–early Silurian periods (455–430 million years ago (Ma)) (Figure 22.1), nearly 85% of marine species (including many trilobites and brachiopods) went extinct when a severe ice age was associated with global cooling and a sea level fall of 100 m, eliminating or disrupting habitats. The Late Devonian mass extinction (375–360 Ma) involved the loss of up to 70% of invertebrate and vertebrate species over an extended period when severe sea level changes occurred and anoxic ocean conditions were experienced, possibly associated with extreme volcanic activity. During the particularly dramatic Permian–Triassic mass extinction (252 Ma), more than 95% of marine species disappeared, together with 70% of terrestrial vertebrates and most insect species. This period was again associated with volcanism, sea level change and ocean anoxia, with the possibility also of impact by an asteroid or comet fragment. The Triassic–Jurassic mass extinction (201 Ma) saw over 70% of species disappear, including certain groups of dinosaurs and large amphibians, another dramatic period of volcanism most likely playing a prominent role. And the fifth mass extinction, at the Cretaceous–Paleogene boundary (66 Ma), is linked to the impact of a large asteroid and associated with the demise of 75% of animal and plant species, including all non-avian dinosaurs. Some of these mass extinctions occurred over relatively short periods–less than 10 000 years in the case of the Triassic–Jurassic and Cretaceous–Paleogene events, and even perhaps just a few years in the latter case. By contrast, the Late Devonian mass extinction played out over tens of millions of years.

It is not fanciful to argue that in the Anthropocene we are already entering the world's sixth period of mass extinction. Recorded extinctions so far may fail to raise much concern among law makers and the general public. (For example, about 200 vertebrate species have gone extinct in the last century, which does not equate to mass extinctions of the past, though compared with ‘background' levels in the last two million years, the loss of 200 species would have been expected to take 10 000 years (Ceballos *et al.*, 2017).) But the pathway to a global extinction of a species involves declines in the sizes and number of local populations, and concomitant declines in geographic range, and when attention is switched from the finale (global extinction) to the prelude (population declines and local extinctions) the reality of the threat to biodiversity comes into stark focus.

The scale of the problem can be gauged from long-term assessments of population sizes of species that have not actually gone extinct. Balmford *et al.* (2003), for example, found evidence of general population declines in databases for British birds, and globally for amphibians, and other vertebrates in terrestrial, freshwater and marine environments (Figure 22.2).

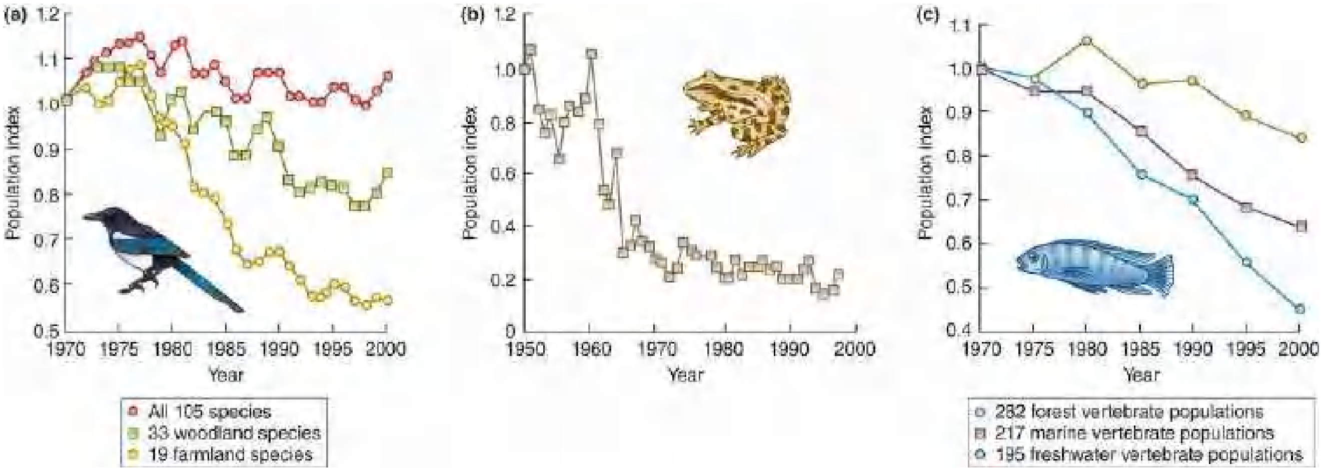


Figure 22.2 Indexes of change in population size (standardised at 1.0 for the first year of each dataset) for various animal groups for which long-term data are available. (a) British bird species from 1970 to 2000. (b) Amphibian populations worldwide from 1950 to 1997 (936 populations of 157 species). (c) Vertebrate populations worldwide from 1970 to 2000.

Source: After Balmford *et al.* (2003), where original references can be found.

In a more general analysis, Ceballos *et al.* (2017) determined the number of mammal, bird, reptile and amphibian species that, according to the International Union for the Conservation of Nature (IUCN), are decreasing (either in terms of range contraction or reduction in numbers of extant populations) and then expressed these as a percentage of total species richness (27 600 vertebrate species in all) in 22 000 grid squares (each 10 000 km^2) across the continents (Figure 22.3). Many parts of the world have high proportions of decreasing species across most vertebrate groups, but some groups show high percentage declines in particular regions. For example, in mammals this tends to be in tropical regions, while in reptiles, percentage declines are greatest across Eurasia and in Madagascar and parts of Australia. Declines in amphibians and birds are concentrated in regions of all the continents. Not surprisingly, most of these declining species have also been classified by the IUCN as ‘endangered' (including ‘critically endangered', ‘endangered', ‘vulnerable' and ‘near-threatened' classifications). But also disturbing is the finding that even among species that are currently classed as of ‘low concern' (including ‘low concern' and ‘data-deficient' classifications), at least 20–30% are declining, and in the case of birds more than 50%. Notwithstanding the fact that other species in the database were stable or increasing, these statistics highlight the dramatic nature of population declines worldwide. Moreover, in the detailed analyses of geographic ranges that were possible for 177 species of terrestrial mammals, Ceballos *et al.* (2017) found that between about 1900 and 2015 most species lost more than 40% of their historic range and almost half lost more than 80%. Finally, a landmark report from the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem services (IPBES 2019), based on a systematic review of 15 000 scientific and government documents, concludes that around one million species of animals and plants are currently threatened with extinction, many within decades.

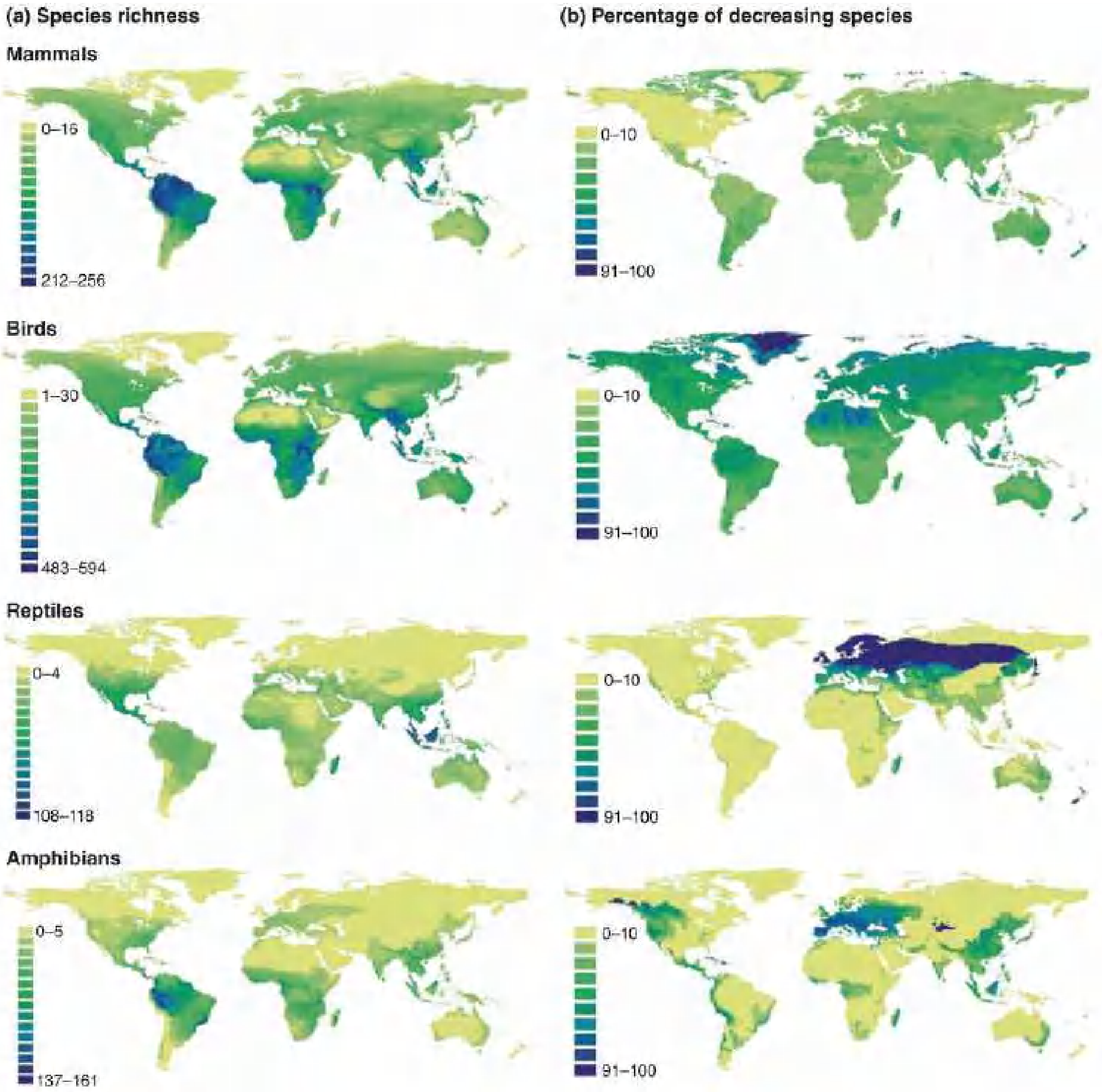


Figure 22.3 Global patterns of terrestrial vertebrate species richness and loss. (a) Global distribution of species richness per 10 000 km^2 grid squares of terrestrial vertebrates according to the IUCN (2015). (b) Percentage of total species assessed by the IUCN as experiencing population losses.

Source: After Caballos *et al.* (2017).

the anthropogenic drivers of biodiversity loss

The anthropogenic factors responsible for local and global extinctions include terrestrial and ocean acidification (Section 22.3), destruction of natural habitats and land-system change (Section 22.4), the pumping of chemical and physical pollutants into the biosphere (Section 22.5), overexploitation of harvested species (Section 22.6) and the spread of invasive species (Section 22.7) (Townsend, 2008). Probably the biggest driver of ecological change in coming decades will be climate change, resulting from chemical pollution of the atmosphere (with carbon dioxide from the burning of fossil fuels the major culprit). For this reason, and because climate change often interacts with the other drivers of change, we deal with it first (Section 22.2). A final section evaluates whether ‘planetary boundaries' for the various drivers (beyond which the whole ‘earth system' may be destabilised) have yet been reached or when they may be reached in future (Section 22.8). Overall, this final chapter is about projected trends in chemical, physical and biological factors for which our species is responsible, the predicted ecological consequences, and some potential management actions to prevent or mitigate adverse effects on biodiversity, ecosystem functioning and the condition and geographic distribution of biomes.

22.2 Climate change

IPCC

Under the auspices of the United Nations, the Intergovernmental Panel on Climate Change (IPCC) has the task of providing the world's governments with an objective, scientific view of climate change and its impacts. Its assessments are based on published literature, with thousands of scientists contributing (on a voluntary basis) to writing and reviewing IPCC reports. The IPCC's Fifth Assessment Report (AR5; IPCC, 2014), concluded that warming of the climate system since the 1950s is unequivocal and that human influence on the climate system is clear, with an estimated 95–100% probability that human influence is the dominant cause.

representative concentration pathways

We dealt with the *history* of industrial gases and the greenhouse effect in Section 2.9. Here we focus on the *future*. IPCC bases its projections on representative concentration pathways (RCPs) that are consistent with a range of possible changes in anthropogenic greenhouse gas emissions. These scenarios, which model different plausible futures, explore the extent to which humans could contribute to future climate change given large uncertainties in factors such as population growth, economic and technological development, social and political change, and their consequences for emissions. Emissions projected under different socioeconomic scenarios are the input to global models that calculate ‘radiative forcing', a measure of the additional energy absorbed by the Earth system due to increases in climate change pollution: the RCPs are named according to their associated radiative forcing in 2100. Each scenario is associated with particular patterns of annual greenhouse gas emissions and other air pollutants – including atmospheric aerosols such as dust, smoke and sulphate pollution that have a counteracting cooling effect.

global and regional climate models

The patterns of radiative forcing are then used as the input for future-climate model simulations – ‘global climate models' (GCMs) – that project global and regional changes in atmospheric and ocean circulation and in temperature, precipitation, general aspects of storm tracks and ocean heat content. So-called dynamical downscaling models, or regional climate models (RCMs), simulate the same processes at a higher spatial resolution. The RCPs also provide climate modellers with gridded trajectories of land use and land cover.

CO_2 and other greenhouse gas emissions and radiative forcing

Thus, as an illustration, RCP8.5 (*high emission*) will reach an overall radiative forcing of 8.5 W m^{-2} (watts per square metre) and is consistent with a complete lack of policy change to reduce emissions (heavy reliance on fossil fuels, increased use of cropland and grassland driven by population increase to 12 billion, low rate of technological development). At the other extreme, RCP2.6 (*low emission*) has radiative forcing reaching 3.1 W m^{-2} before returning to 2.6 W m^{-2} by 2100, and is consistent with stringent climate policies (declining use of fossil fuels, high rate of technological development, 2100 world population of 9 billion, greater use of croplands for bioenergy production, declining use of grassland, extensive reforestation). In this most optimistic scenario, carbon dioxide (CO_2) emissions were projected to remain at current levels until 2020 then decline by 2100 to a level lower; the resulting pattern for CO_2 concentration in the atmosphere has a peak around 2050, followed by a modest decline to 400 ppm by 2100. Between the two extremes are the *intermediate emission* scenarios RCP4.5 and RCP6.0 (Figure 22.4a, b). Within these RCP classes, the results of multiple models developed by different research teams have been synthesised by IPCC to provide average projections to 2100 together with measures of uncertainty.

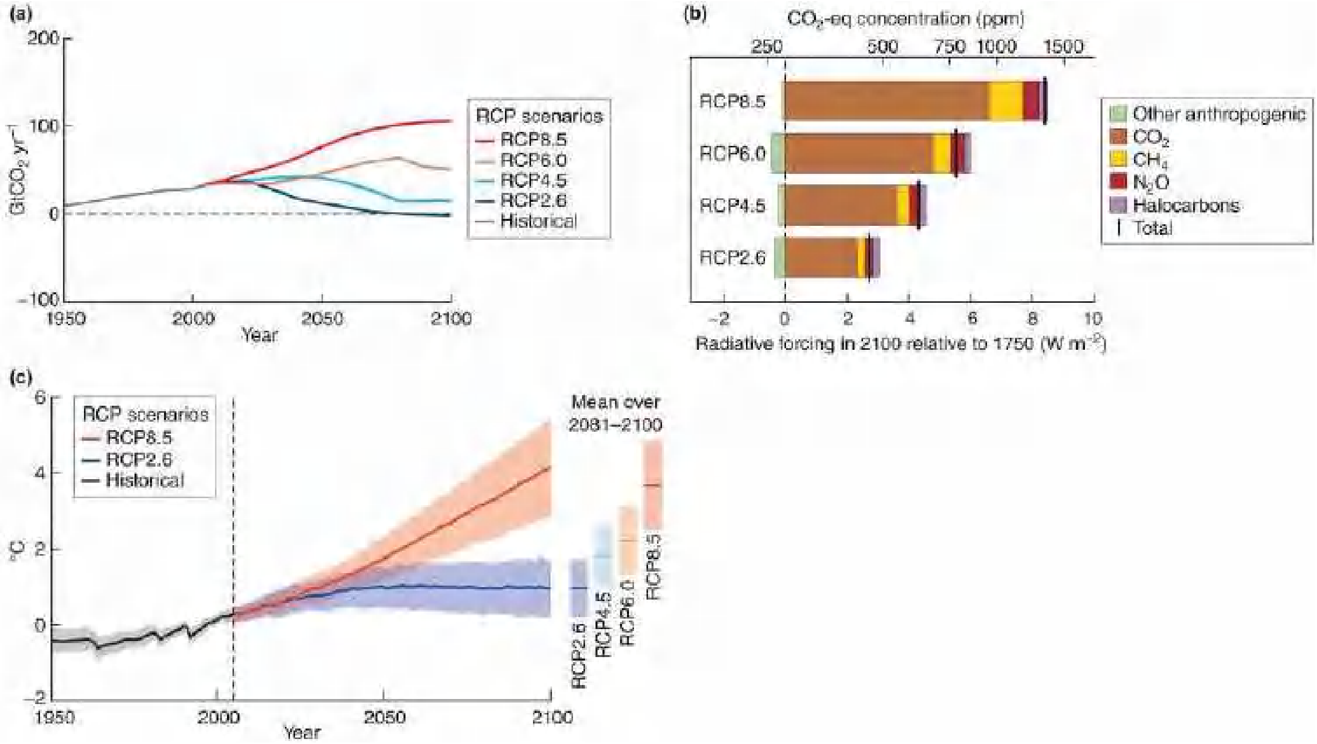


Figure 22.4 Projected carbon dioxide emissions and surface temperature for four representative concentration pathways. (a) Historical and projected CO_2 emissions for four representative concentration pathways (RCPs) until 2100. (b) The contribution of projected CO_2 emissions (and other greenhouse gas emissions) by 2100 to atmospheric concentrations (in CO_2 equivalents) and radiative forcing relative to 1750. The green category ('Other anthropogenic') includes the cooling effect of atmospheric aerosols such as dust, smoke and sulphate pollution. (c) Projected changes in global annual mean surface temperature until 2100, relative to 1986–2005, for RCP8.5 (high emission scenario – red line and shading) and RCP2.6 (low emission scenario – blue line and shading). The lines show means and the shading indicates the range of results from multiple models (39 models for RCP8.5, 32 models for RCP2.6). The black line with grey shading shows the modelled change until 2010 using historical reconstructed forcings (42 models). The side bars contrast the means and ranges averaged between 2081 and 2100 for all four scenarios, including intermediate emission scenarios RCP4.5 and RCP6.0.

Source: IPCC (2014).

temperature

The results projected for global average surface temperature change relative to 1986–2005 are shown in Figure 22.4c. The likely temperature increases until the period 2081–2100 are 0.3–1.7°C, 1.1–2.6°C, 1.4–3.1°C and 2.6–4.8°C for RCP2.6, RCP4.5, RCP6.0 and RCP8.5, respectively. This projected warming will, in turn, cause further profound effects certain to have ecological consequences, for example in rising sea levels (projected to increase by 2081–2100 by 0.26–0.82 m across the range of RCPs), and reductions by 2100 in the extent of glaciers (reductions of 15–85%), northern hemisphere spring snow cover (7–25%), near-surface permafrost (37–81%), and arctic sea-ice (43–94% in September, 8–34% in February) (IPCC, 2014).

geographic patterns of warming, and precipitation, are not uniform

The models used in IPCC assessments divide the world into 518 400 cells measuring half a degree of latitude and longitude. Switching from averages through time to a spatial representation, it becomes immediately obvious that projected patterns of warming are not uniform (Figure 22.5a). Thus, for example, the Arctic region will continue to warm more rapidly than the global mean under both RCP2.6 and RCP8.5 scenarios and the strongest ocean warming is projected for surface waters in tropical and northern hemisphere subtropical regions under the RCP8.5 scenario. The effects of warming on other critical aspects of climate, including winds and currents, but particularly precipitation, are also expected to be non-uniform (Figure 22.5b). For example, under the RCP8.5 scenario, high latitudes and the equatorial Pacific are expected to experience an increase in annual mean precipitation, mean precipitation will be likely to decrease in many mid-latitude and subtropical dry regions, while mean precipitation is expected to increase in many mid-latitude wet regions.

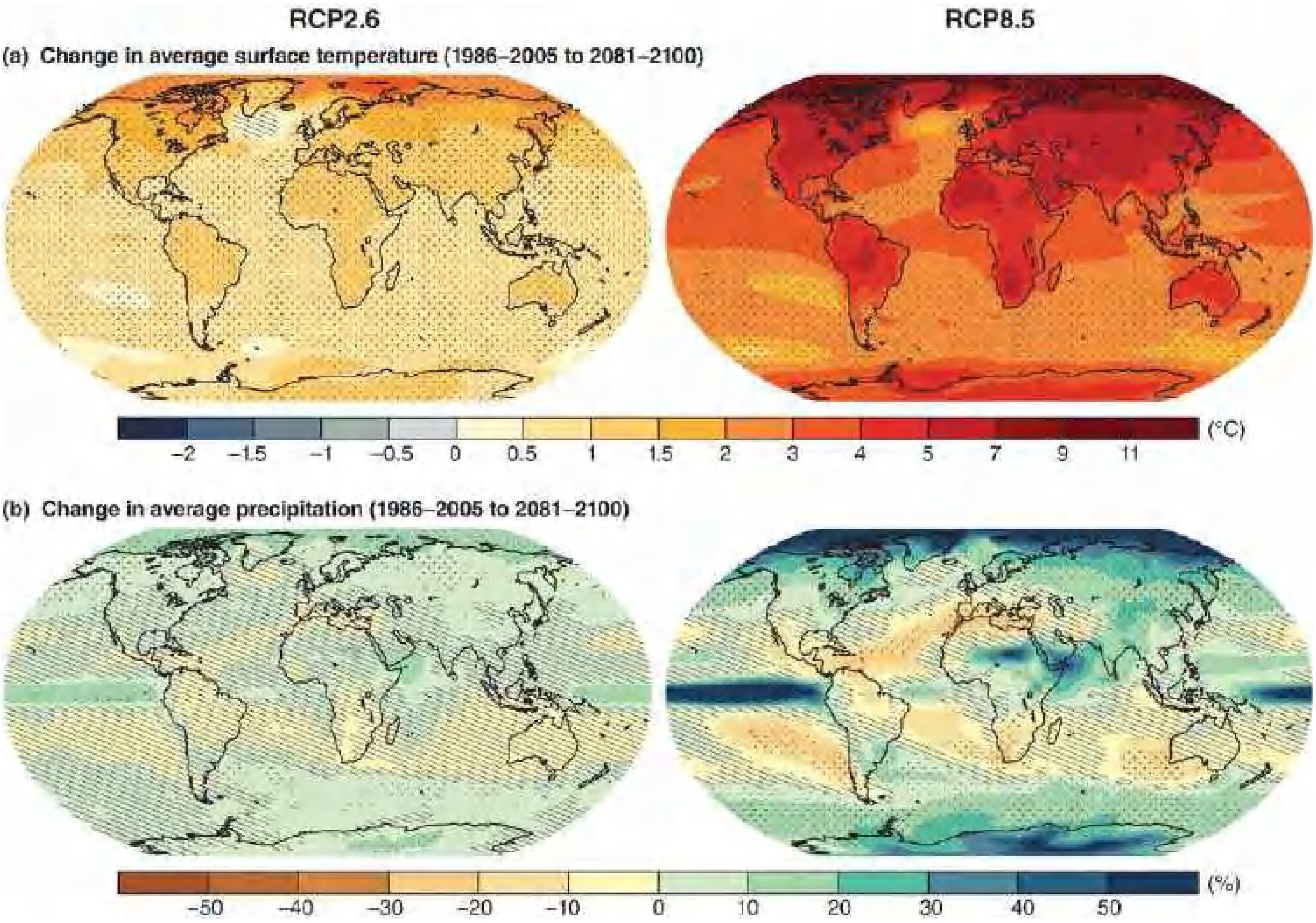


Figure 22.5 Annual mean change in (a) surface temperature and (b) precipitation. Averaged for multimodel results (32 for RCP2.6 and 39 for RCP8.5) comparing the period 2081–2100 with 1986–2005. Hatching shows regions where the multimodel mean is small compared with natural internal variability (i.e. less than one standard deviation of natural internal variability in 20-year means). Stippling indicates regions where the multimodel mean is large compared with natural internal variability (i.e. greater than two standard deviations of natural internal variability in 20-year means) and where at least 90% of models agree on the sign of change. Projections for the stippled regions are thus more reliable.

Source: IPCC (2014).

22.2.1 Ecological risks

The global climate changes that occurred naturally over geological eras, at rates much lower than those in the current epoch (Glikson, 2016), caused significant extinctions, biodiversity changes and ecosystem shifts. Hence, it is to be expected that many plant and animal species will be unable to adapt locally to changing climate or, depending on their migratory ability, to shift to track suitable climates in future. In this section, we present examples of likely ecological consequences of climate change ranging from the population level, through community interactions to the distribution of whole biomes.

shifting distributions of forest trees – unlikely to keep up with climate change

Coupled with the wholesale fragmentation and conversion of habitats, we can expect not only shifting distributions of species in response to climate change but also migration failures. Using climate projections from the high emission climate scenario RCP8.5 at 2055, Rogers *et al.* (2017) modelled changes in the distribution of temperate tree species in the eastern USA (east of

100°W). Their models start with a measure of changing habitat suitability in a grid of 800 × 800m pixels. This ‘potential impact' (PI), on a scale from −1 to +1, reflects the different species' niches, primarily based on their physiology and geographic climatic and soil patterns, captured in species distribution models (SDMs: see Ecological Niche Models in Section 2.2). But in addition, they modelled what they called ‘adaptive capacity' (AC; range 0–100) – capturing migration potential (from surrounding grid cells), the density of seed sources and forest fragmentation (a barrier to dispersal). Finally, for each species and every pixel they assessed ‘vulnerability', V, which combines potential impact and adaptive capacity such that

$$\begin{aligned}&V = \mathrm{PI} \times (1 - \mathrm{AC}/100) \text{ when PI is positive}\\&(\text{worsening future conditions})\end{aligned}$$

$$\begin{aligned}&V = \mathrm{PI} \times (\mathrm{AC}/100) \text{ when PI is negative}\\&(\text{improving future conditions})\end{aligned}$$

The average vulnerability scores for 40 tree species and 10 forest types (particular combinations of trees) are shown in Figure 22.6a. The most vulnerable were generally northern and mountainous species that require lower annual mean annual temperatures, including members of the spruce-fir, northern hardwoods and maple–beech–birch forest types. Vulnerability was particularly related to the high velocity of climate warming in the northern lowlands and limited upslope area for range expansion in the mountains. In contrast, members of the elm–ash–cottonwood, southern oak–hickory, oak–hickory, longleaf–slash pine, and loblolly–shortleaf forest types were only minimally vulnerable (in fact, loblolly–shortleaf forest was on average negatively vulnerable). Topographic complexity and climatic heterogeneity in the central Appalachian Mountains allowed for a low climate velocity, and the relatively unfragmented landscape there allowed for high ACs.

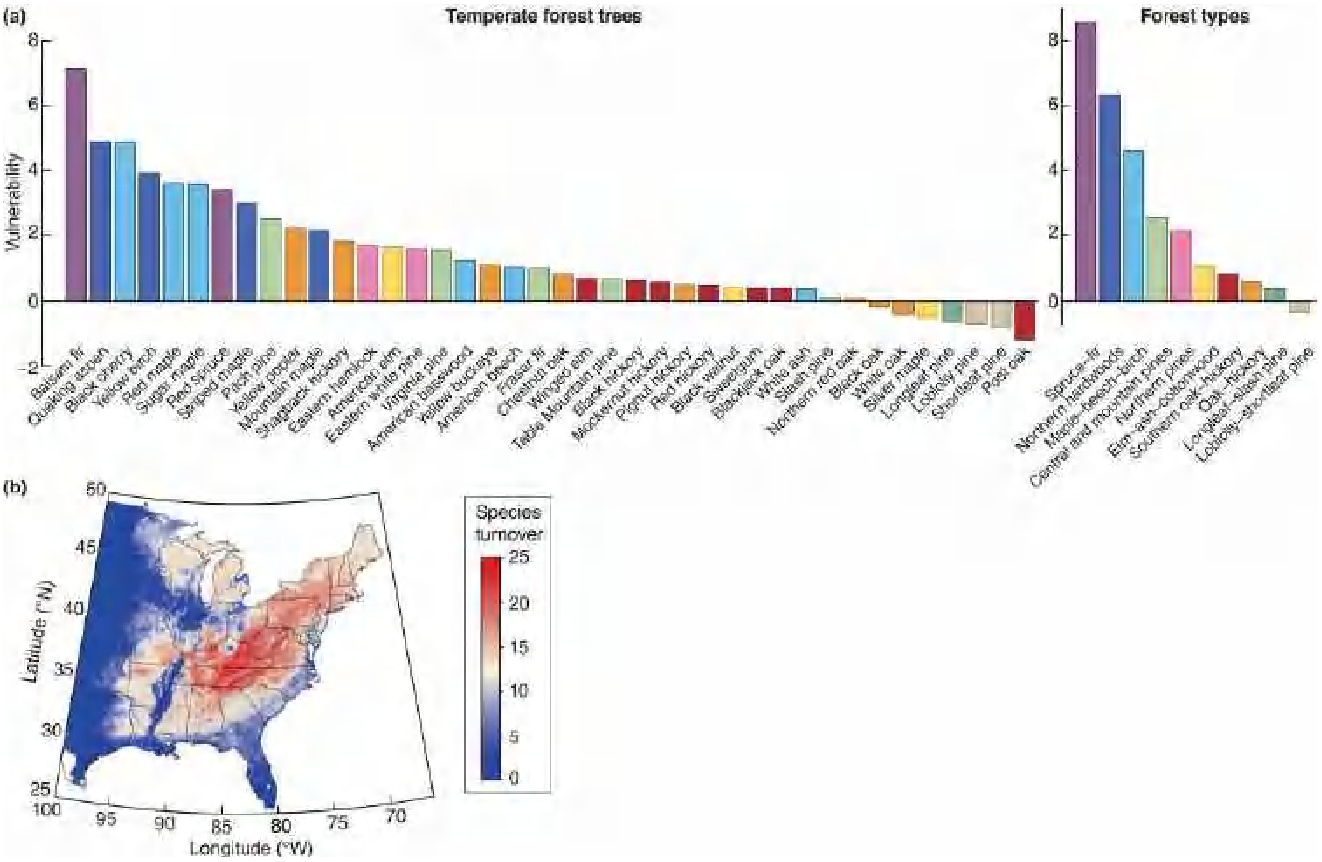


Figure 22.6 **Tree species vulnerability to projected climate warming.** (a) Ranked vulnerability scores (which take into account both the potential impact of changing habitat suitability and each species' adaptive capacity) for temperate forest trees and forest types in the eastern USA, in response to projected climate warming based on the high emission climate scenario RCP8.5 by the year 2055. (b) Species turnover by 2055 in each grid square, represented by the number of species with extreme vulnerabilities, less than −1 (gaining) or greater than +1 (losing).

Source: From Rogers *et al*. (2017).

Note also that as a result of species-specific differences, both in the potential impact of climate change on habitable areas and in migration potential, some species showed gains while others showed losses: as a consequence, results aggregated at the landscape level suggest high species turnover over much of the mountainous area (Figure 22.6b).

The authors conclude that their vulnerability patterns are consistent with other studies (e.g. Coops *et al*., 2016) that suggest migration and range shifts will be largely unable to keep pace with the velocity of 21st century climate change without human intervention. The extent of the risk is obviously related to the severity of climate warming, and the authors note that the average of vulnerability scores by 2055 was 4.1 when based on the high emission RCP8.5 scenario, but 3.5 under the intermediate RCP4.5 scenario.

APPLICATION 22.1 Combating the increased risk of forest fires in the boreal region

The world's conifer-dominated boreal forests are experiencing increased disturbance as a result of climate warming, particularly in terms of the prevalence of forest wildfires, which themselves add to greenhouse gas emissions. These forests, both managed (producing more than half of the world's harvested timber) and unmanaged (Figure 22.7a), store a huge quantity of carbon (367.3–1715.8 Pg C) and provide an annual sink (net uptake of C) of about 0.5 Pg C year^{-1}. However, during the period 1997–2014 nearly 6 Mha year^{-1} have succumbed to wildfires (Figure 22.7b), releasing CO_2 at a rate equivalent to 79% of the biome's total net uptake. Astrup *et al.* (2018) propose a strategy of increasing the proportion of broad-leaf species in the managed areas, with the joint aims of contributing to the mitigation of climate warming and adapting to its effects. Having a higher moisture content and lower flammability, broad-leaved species are considerably less likely to burn than conifer species (a pure broad-leaved stand is about 24 times less likely to burn than one of conifers). Hence, increased representation of broad-leaved trees would have the effect of reducing greenhouse emissions via fewer wildfires, but in addition, the higher year-round surface albedo of broad-leaved species is predicted to equate to less total energy absorption (more light reflected), adding to the mitigation potential of the proposal. Given the enormous cost of forest fires in lost productivity, damage to infrastructure and impacts on human health and safety, increasing the broad-leaved tree composition can also be viewed as a socioeconomic adaptive response to climate change. Bringing all this together, the authors argue that converting just 0.1–0.2% of forested area in southern Canada per year from 2020 would be sufficient to mitigate the projected increase in fires due to climate warming.

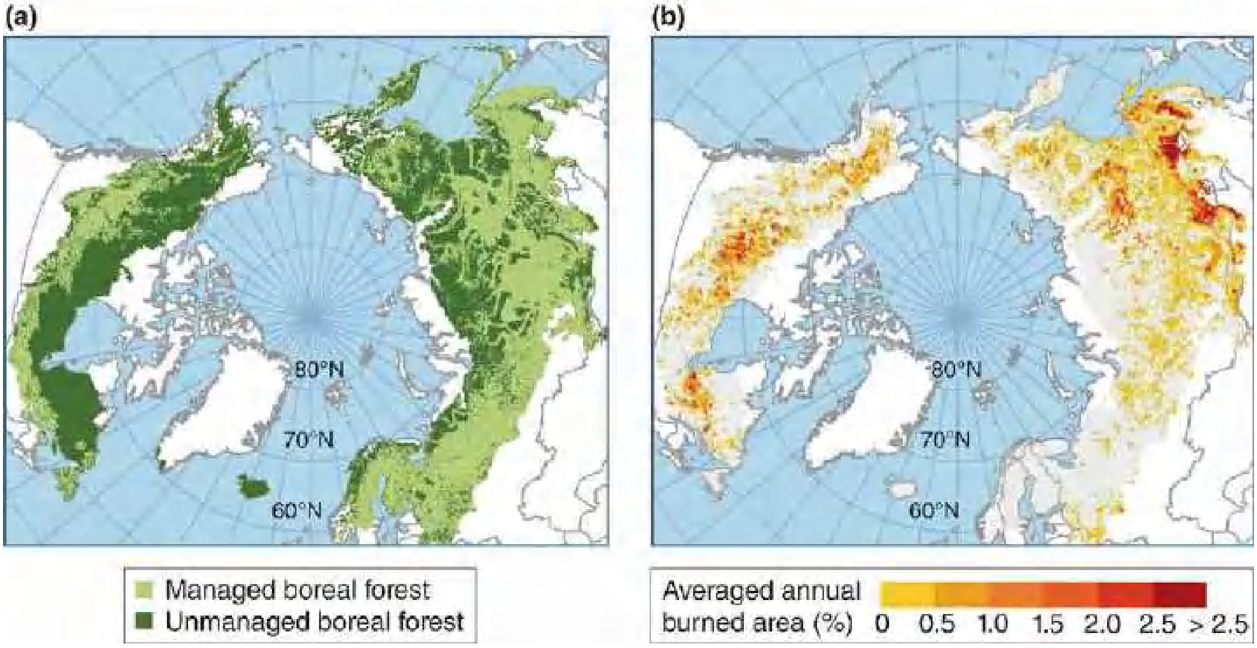


Figure 22.7 Loss of circumboreal forest to wildfire. (a) Distribution of the managed and unmanaged portions of the circumboreal forest. (b) Estimates of percent annual area burned 1997–2014.

Source: From Astrup *et al.* (2018).

upward expansion of alpine plants – likely to keep up but will some run out of space?

The annual herbs, perennial herbs and woody plants of alpine zones can be expected to show an upward altitudinal expansion with climate change and, indeed, there is strong evidence that this has already been happening. Dainese *et al.* (2017) analysed more than 124 380 GPS records of native alpine plants over a 20-year period (1989–2009) in north-east Italy, allowing them to estimate mean vertical spread rate for 1208 species. The overall rate of upward spread was 16.6 m year^{-1}. (About 20% of species showed negative vertical spread, probably explained by stochastic population fluctuations.) Rates of upward spread were greater for species that occurred near roads, indicating a role for unintentional human-assisted dispersal in the ability of some species to spread upslope at faster rates. Based on a future projection of climate change specific to the European Alps (about 0.25°C warming per decade until mid-century), a vertical spread rate of 3.85–5.54 m year^{-1} would be necessary to keep track of suitable temperatures and it seems that many alpine plants will indeed be able to keep up. However, species at higher elevations have limited space to move upwards and immigration of downhill species may contribute to the decline or extinction of higher altitude residents. Of more concern is the finding that non-native species tend to spread upwards at an even faster rate than natives (mean of 27.4 m year^{-1}), with implications for native extinctions and effects on native species composition.

climate warming, migratory birds and mismatches in spring arrival

Climate change can be expected to be particularly problematic for migratory species, because it can lead to a mismatch between individuals' arrival time at their destination and the optimum time for their arrival. Billions of birds migrate from subtropical/tropical wintering grounds to higher latitude breeding areas, each journey consisting of a series of fuelling stop-overs and migratory flights to the next destination. The rate of accumulation of energy at stop-overs is less than the rate of expenditure in flight. Hence, total migration time depends mainly on stop-over time and less on flight speed (Figure 22.8a). For example, Schmaljohann and Both (2017), based on the results of 49 tracking studies (using miniaturised tracking devices) of 320 individual birds of 46 species, estimated that a 20% reduction in total stop-over time would result in an average two-day reduction in migration time for a migration distance of 5000 km and a seven-day reduction for a 10 000 km migration (Figure 22.8b). It seems, indeed, that migrant birds have responded to changing climate by arriving somewhat earlier on their breeding grounds, probably because of climate-warming enhancement of food availability in their wintering grounds or stop-over sites. However, in temperate European forests, the timing of peak numbers of caterpillars (primary prey for nestlings of many songbird species) has already advanced as a result of climate warming by between nine and 20 days. Thus, the speedier migrations seem not to be fast enough to compensate for changes in food availability at the destination sites, and these mismatches between arrival time and nestling food have most likely contributed to widespread population declines. With further climate warming, therefore, it seems that evolutionary change would need to be the prime process for reducing mismatches, on a timescale that is unknown.

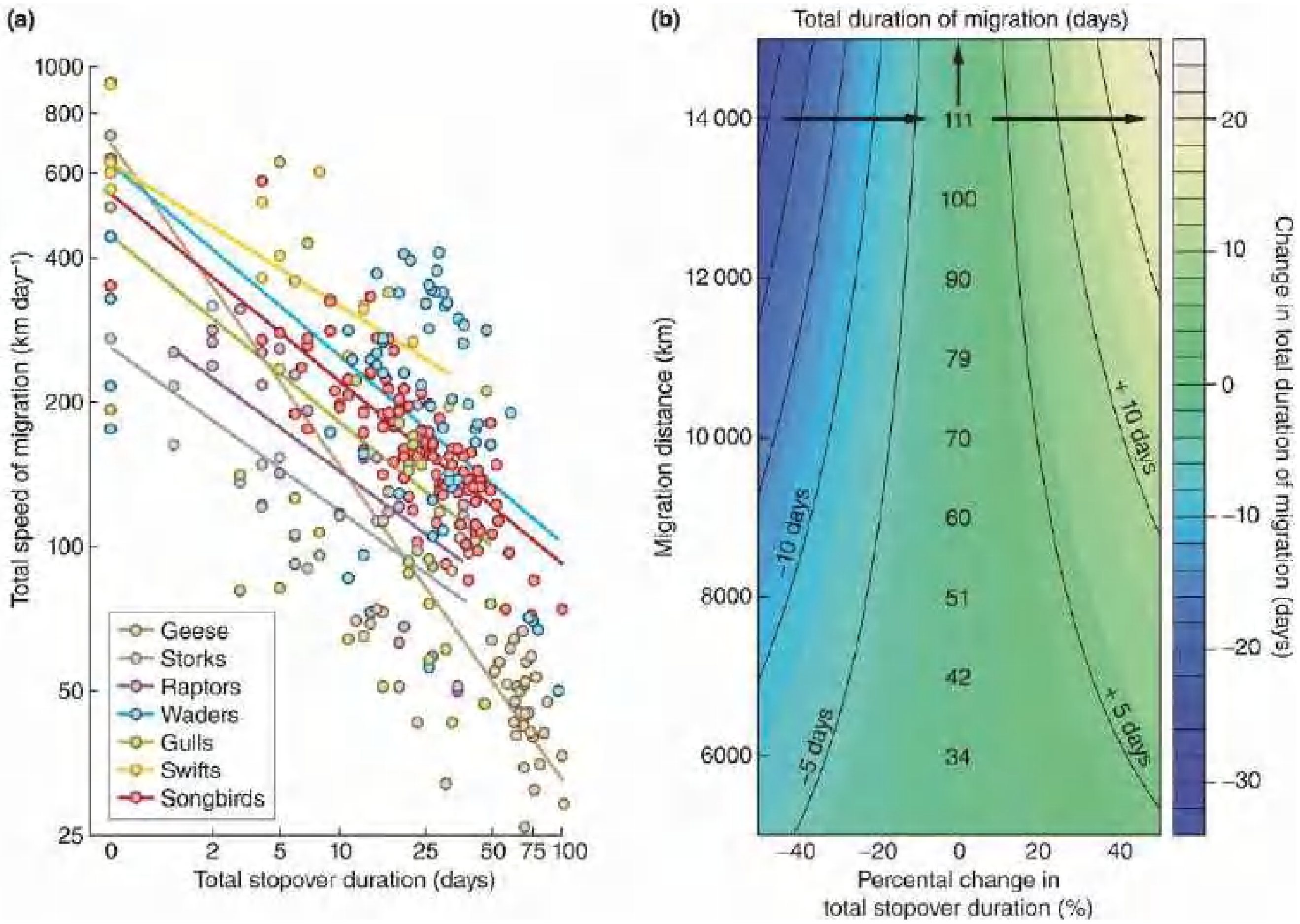


Figure 22.8 Projected mismatches between migratory birds' arrival time and the optimum time for arrival at their destination. (a) Individual total speed of migration (km day^{-1}) plotted against individual total stop-over duration for seven bird groups. Total stop-over duration and bird group together explained 66% of variation in total migration speed. (b) Change in predicted total duration of migration as a function of a change in total stop-over duration in migratory songbirds. Numbers in the middle of the figure show total duration of migration in days for a given migration distance, e.g. the arrows indicate total duration of migration for a distance of 14 000 km, where a 40% change in stop-over duration would reduce the migration time by about 20 days (right-hand scale).

Source: From Schmaljohann & Both (2017).

king penguins – island colonies and shifting foraging grounds

While the primary focus for ecological consequences of climate change are shifts in species ranges that reflect their niche requirements, distributional shifts to newly appropriate climatic zones are often hampered by geography. Colonies of king penguins (*Aptenodytes patagonicus*) exist only on year-round ice-free (and predator-free) areas of scattered islands in the Southern Ocean. The islands cannot shift, but penguin foraging grounds do move in response to the distribution of their principal prey (myctophid fish) that thrive around the Antarctic polar front (APF), where the cold waters from the Antarctic region meet and sink beneath warm waters from mid-latitudes. Penguins can potentially forage in the ocean to the north as far as the APF extends and to a southern limit set by sea ice in September. However, successful breeding is not expected if the penguins have to move further than about 700 km from their breeding grounds for food. Warming has led to a poleward displacement of the APF and penguin foraging grounds since the last glacial maximum (21 000–19 000 years ago, through the mid-Holocene (6000 years ago) to the current period (1981–2005) (Figure 22.9a–c). According to models based on the high emission scenario RCP8.5, this is set to continue with further climate warming as we move further into the Anthropocene (Figure 22.9d). During the last glacial maximum, the extent of sea ice cover limited penguin distribution to a fraction (four islands) of its current extent (eight islands). Now, though, warming to the end of the century is set to see foraging distance increasing steadily in the world's largest colonies located north of the APF (outside the dotted red line in Figure 22.9d), and as a result, these populations are expected to decline or disappear. Taking Crozet Island (location 8) as an example, Figure 22.9e shows predictions based on three of the greenhouse gas scenarios, RCP2.6, RCP4.5 and RCP8.5, and even in the most optimistic scenario the risk is high of exceeding the critical 700 km foraging distance maximum. On the other hand, conditions should become more favourable on the colder South Sandwich and Bouvet Islands south of the APF (inside the dotted red line in Figure 22.9d), with shorter foraging distances and decreased sea ice.

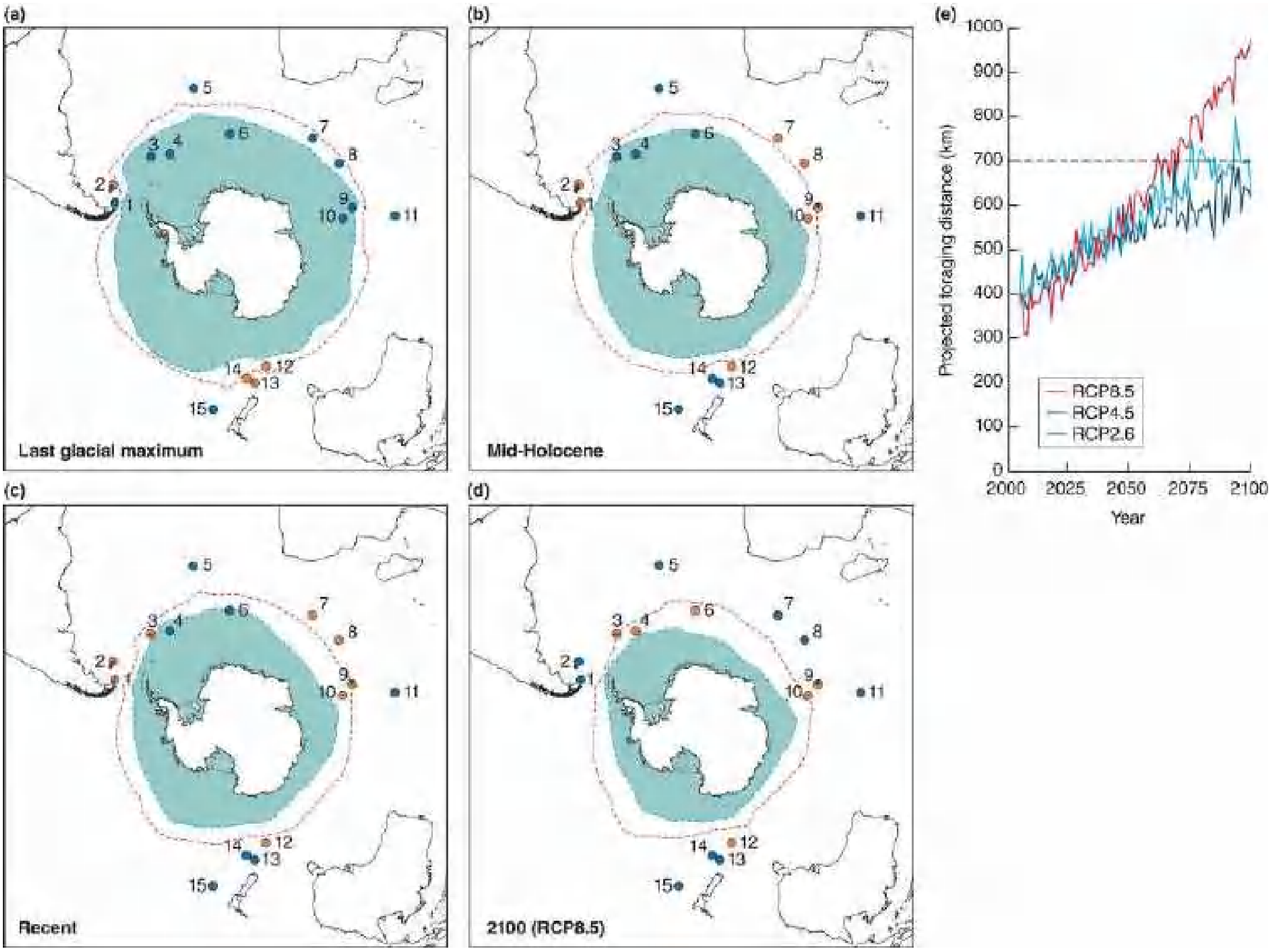


Figure 22.9 Distributional shifts to newly appropriate climatic zones may be hampered by geography. (a–d) Inferred position of the Antarctic polar front in February (sea surface temperature 5°C; dotted red line), which defines the most important foraging grounds for king penguins, and the extent of sea ice in September (light turqoise shading) at the time of the last glacial maximum, mid-Holocene, recent (1981–2005) and projected for 2100 according to RCP8.5. Islands with penguin colonies are shaded orange; islands without, shaded blue. Island 1: Tierra del Fuego; 2: Falklands; 3: South Georgia; 4: South Sandwich; 5: Gough; 6: Bouvet; 7: Marion and Prince Edward; 8: Crozet; 9: Kerguelen; 10: Heard and McDonald; 11: Amsterdam; 12: Macquarie; 13: Auckland; 14: Campbell. 15: Chatham. (e) Projected foraging distances under three greenhouse gas concentration scenarios for Crozet Island: RCP2.6, RCP4.5 and RCP8.5. The dashed black line shows the 700 km limit, beyond which no successful breeding is expected.

Source: From Cristofari *et al.* (2018).

Once again, we see how the ecological consequences of climate change are likely to be moderated by the ability of species to move to newly suitable climatic areas, in this case due to geographic barriers.

Atlantic cod – shifts in location of fisheries and size of fish

Ocean warming is expected to lead to smaller body sizes in cold-water marine ectotherms because metabolic rates increase exponentially with temperature while the cardiorespiratory capacity to match enhanced oxygen demands is limited – an argument similar to that discussed in Section 3.9 on the metabolic theory of ecology. This has significant implications for the demography of Atlantic cod (*Gadus morhua*), the basis of some of the world's important cod fisheries, and in particular for their maximum body size, a key aspect of their long-term economic viability. The currently observed range of average temperatures for cod breeding is 0–10°C. Butzin and Pörtner (2016) used high emission climate scenario RCP8.5 as the basis for modelling geographic patterns of waters with these temperatures during the cod spawning season (typically February to May), comparing 1985–2004 with 2081–2100 (Figure 22.10a, b). The model projections showed spawning season temperatures rising by 2081–2100, leading to potential loss of cod habitat in the southern north-east Atlantic and southern North Sea, but an expansion of cod habitat in some Arctic Ocean margins. The researchers were then able to estimate geographic patterns in the potential maximum size of fish (Figure 22.10c, d) by combining the temperature estimates with physiological modelling based on laboratory-determined growth rates (Figure 22.10e). The shift to heavier fish reflects substantial weight increases at high latitudes associated with enhanced warming, approaching optimum conditions for cod growth in the Barents Sea and on the Canadian Shelf.

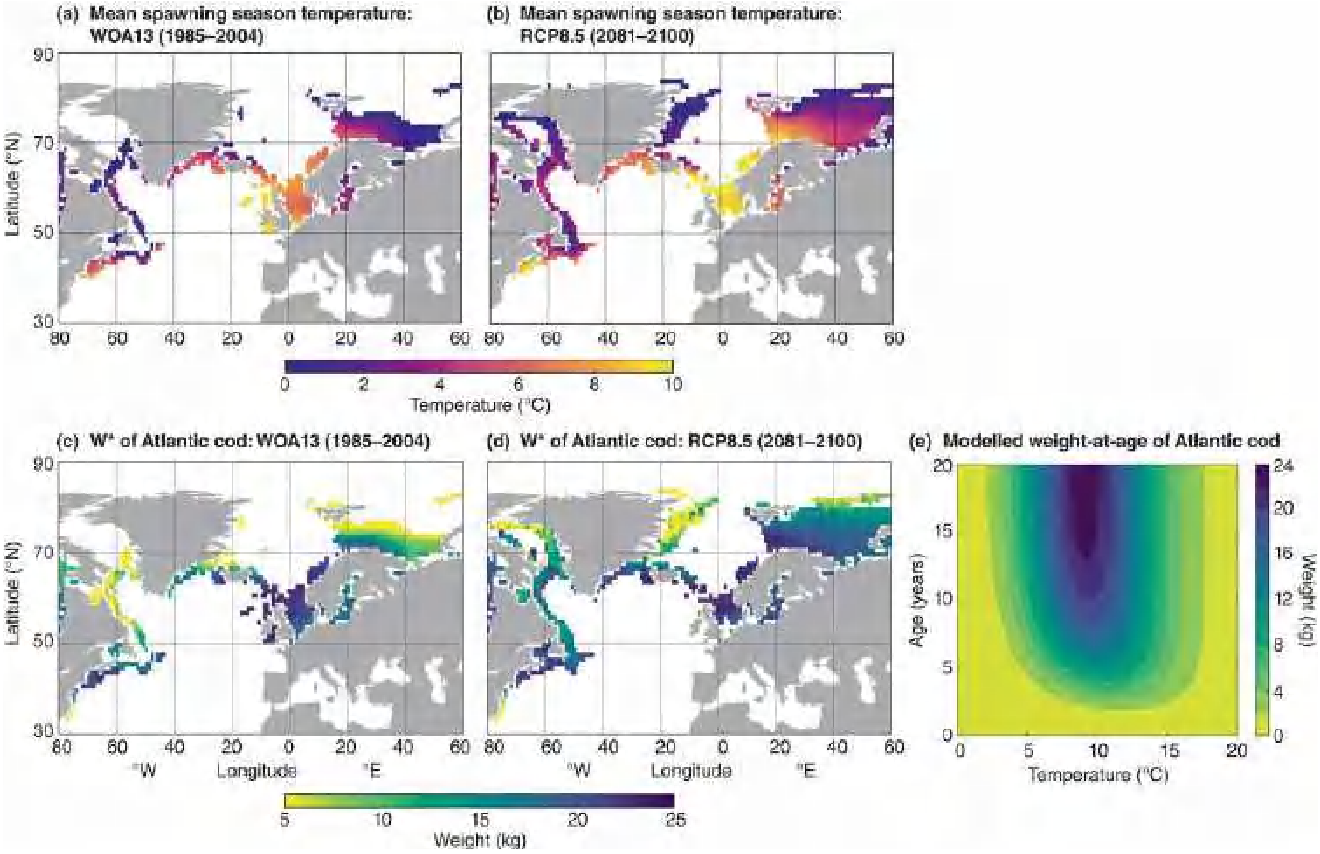


Figure 22.10 **Ocean warming, body size and viability of an important fishery.** (a) Water temperature during the cod breeding season averaged from February to May at 45–150 m depth (25–35 m in the shallower southern North Sea) for 1985–2004 (based on temperatures provided by the World Ocean Atlas 2013 (NOAA, 2013) and (b) for 2081–2100 (projections based on climate scenario RCP8.5). (c, d) Projected maximum cod sizes achieved (body weight at age 20; W*) in the scenarios in (a) and (b). (e) Modelled growth (weight-at-age) of Atlantic cod based on laboratory temperature experiments.

Source: From Butzin & Pörtner (2016).

Note, though, that these cod models take no account of potential future climate-related mismatches in the distributions of predatory fish such as cod and their principal prey (herring, *Clupea harengus*, and mackerel, *Scomber scombrus*). In fact, the geographic models of Selden *et al.* (2017) suggest that future warming may lead to a decline in the proportion of the prey species' ranges occupied by cod, with potential consequences for the cod populations not considered in Figure 22.10. Other top predators, such as spiny dogfish (*Squalus acanthias*), on the other hand, are projected to show enhanced overlap with prey distributions, probably increasing their importance as predators in the ecosystem.

shifting biomes ...

Even the simplest climate-related ecological models are enormously complex, attempting to capture the future effects of greenhouse gas emissions when political, social and technological developments can only be guessed at, and translating these into climatological consequences (temperature, precipitation, ice cover) across the face of the earth. In essence, the examples just discussed seek to superimpose abiotic species niche requirements on the projected geographic climate patterns. But this of course misses the biotic aspects of species niches, related to food availability and interactions with other species. We have countered this to some extent by noting food-web mismatches that may be expected in the cases of migrating songbirds, king penguin feeding grounds, and Atlantic cod and their prey. Undoubtedly, competitive interactions will be influential too, for example among the species of forest trees and alpine plants. Attempting to create a comprehensive and plausible ecological model for even a single species is an enormous challenge, given all the uncertainties. Some might even argue that it is a somewhat vain quest, given all the uncertainties relating to the details of greenhouse gas emissions, climatic consequences, as well as population dynamics and food-web relationships.

... and anthromes

So, moving up the ecological scale needs to be approached with caution. However, an ecosystem level of enquiry is possible in terms of changing distributions of the earth's biomes. Boit *et al.* (2016) have projected the distribution of biomes in Latin America, a high diversity ecoregion, based on the high emission climate scenario RCP8.5 and the low emission RCP2.6, comparing land cover in 2005 with what is predicted for 2099. They focused especially on plant physiological responses to water stress from regionally increasing temperatures and declining precipitation, and they considered not only classical biomes, such as desert, savanna and rainforest, but also three biomes unique to the Anthropocene, termed anthromes by Ellis *et al.* (2010): urban, pastoral and cropland (Figure 22.11).

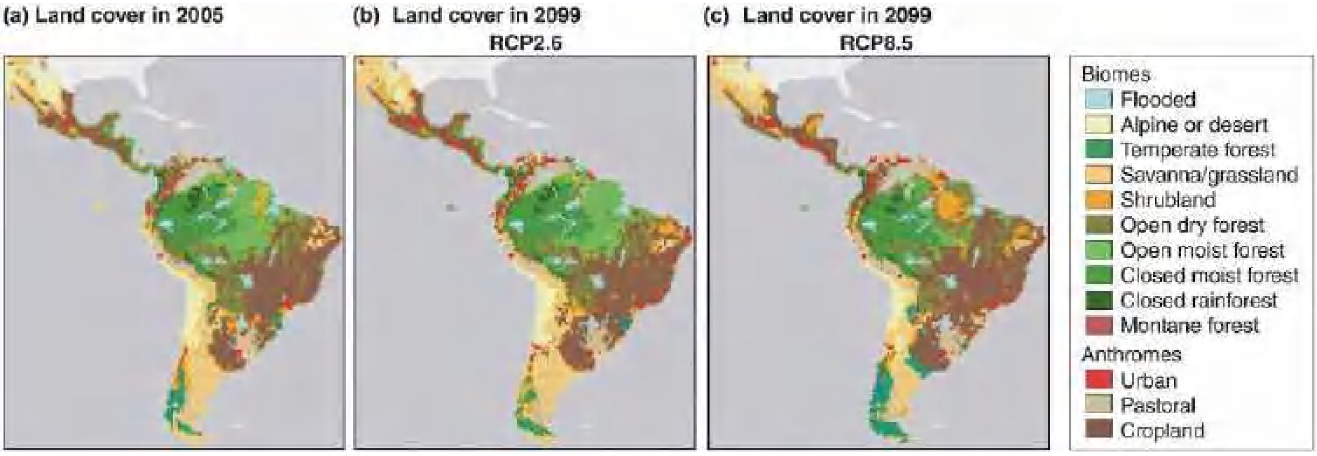


Figure 22.11 Current and projected distributions of 10 biomes and three anthromes. (a) In 2005, (b) in 2099 using climate change patterns projected for low emission climate scenario RCP2.6 and (c) in 2099 on the basis of high emission scenario RCP8.5.

Source: From Boit *et al.* (2016).

The modelled plant responses see some closed forests turn into open forests and some open forest into shrubland, particularly in the northern part of the Amazon basin and Mexico's Yucatan

Peninsula, while dry forest in the southern Amazon basin becomes increasingly fragmented. Just as significant are projected transformations of forest to cropland or pastureland in the two scenarios. Overall, the total area undergoing biome shifts between 2005 and 2099 is projected to be 8% for RCP2.6 and almost double that (15%) for RCP8.5. Because all the models assume CO_2-fertilisation effects on plant productivity (see Section 3.3.2), even the most severe scenario does not lead to tropical forest loss in these projections.

APPLICATION 22.2 A zero-deforestation policy?

Boit *et al*. (2016) note that if an early zero-deforestation policy were to be enacted across Latin America, this would not only prevent further anthrome formation, preserving natural biomes, but also would contribute to climate-change mitigation by locking carbon in forest biomass and preventing increased greenhouse gas emissions arising from the conversion of biomes to anthromes. The authors argue that the time has come to implement less 'land-hungry' policies, in concert with plans to protect the remaining natural vegetation and secure its potential to reduce greenhouse emissions.

22.3 Acidification

acid rain and ocean acidification – different causes but both with profound effects

Two distinct categories of acidification have profoundly affected ecosystems during the Anthropocene. Given the significance of pH as an ecological condition (Section 2.5), it is to be expected that human activities that alter the pH of precipitation, freshwaters and oceans will affect individual species, community composition and ecosystem functioning, and such effects have already been witnessed in both terrestrial and aquatic habitats. In terrestrial and freshwater systems, problematic acidification has occurred largely as a result of atmospheric pollution with sulphur dioxide (SO_2) and oxides of nitrogen (NO_x) produced by the burning of fossil fuels: 'acid rain' (atmospheric deposition of sulphuric and nitric acids) then falls in the vicinity and downwind of centres of human industrial, domestic and transport activity, damaging forests, soil, river and lake ecosystems, often across national boundaries (Sections 2.8, 21.4.3 and 21.4.4). Ocean acidification, on the other hand, with the same root cause of the burning of fossil fuels, results when emitted CO_2 is dissolved in seawater and some of this is converted into carbonic acid.

acid rain – differential patterns in acid deposition and recovery around the globe

Acid rain and its ecological consequences (Section 2.8) were first described in Europe and North America in the early 1970s, and a decade later in Asia. With emissions in North America and Europe declining, East Asia has become the global hotspot of nitrogen and sulphur deposition (Figure 22.12a). However, China, the major emitter in the region, has seen emission reductions of SO_2 since 2006 and of NO_x since 2012, as a result of widespread application of flue-gas desulphurisation and catalytic reduction in coal-fired plants (Duan *et al*., 2016). The severity of acid rain has declined as a consequence (Figure 22.12b). Across Asia it seems surface waters are not as sensitive to acid rain as in Europe and North America, because of high buffering capacity of the soils, high alkalinity of inland waters, and high base cation deposition, in particular Ca^{2+} from soil dust. Soil acidification has however been widely observed in China, Japan and Korea, with abnormal forest defoliation and increased tree mortality. Recovery of soil systems from acidification is beginning to be seen, but large stores of adsorbed sulphur are expected to be desorbed, delaying recovery, and the risk of regional soil acidification still exists. Moreover, Duan *et al*. (2016) warn that if attempts are made to control erosion and the production of dust rich in Ca^{2+} in the future, more effort still will be needed to combat soil acidification.

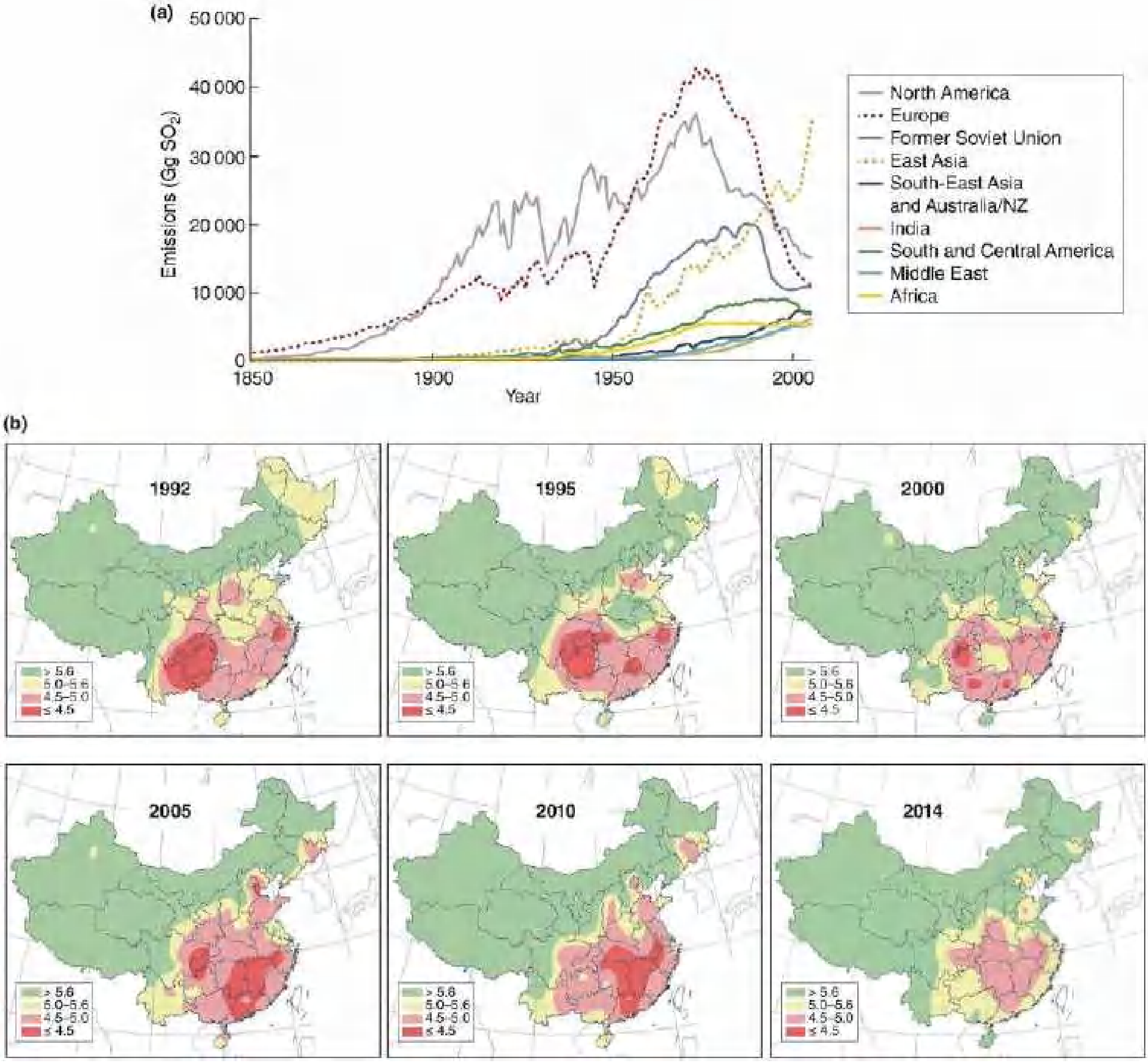


Figure 22.12 Acid rain declining in Europe and North America but increasing in other parts of the globe. (a) Global sulphur dioxide emissions by region 1850–2005. East Asia includes China, Japan and South Korea. (b) Distribution of acid rain in China 1992–2014 (four pH ranges are shown).

Source: (a) From Smith *et al*. (2011). (b) From Duan *et al*. (2016).

ocean acidification – a growing problem

While acid rain is a regional problem and there are signs of improvement due to reductions in SO_2 and NO_x emissions, ocean acidification is a global problem because of thorough mixing and longevity of CO_2 in the atmosphere, and it is a worsening problem because of projected increases in CO_2 concentrations.

Since the beginning of the industrial era, oceanic uptake of CO_2 has caused a decrease in average surface water pH of 0.1 units, representing a 26% increase in acidity (IPCC, 2014). Projections for the future are a further decrease by 2100 of between 0.06 and 0.32 pH units (15–109% increase in acidity) across the range of scenarios from RCP2.6 to RCP8.5. Reductions in pH influence species because of their specific tolerance ranges for this factor. But the consequent pH-related undersaturation of calcium minerals in the oceans is having further major consequences for calcifying species such as corals, molluscs, echinoderms and many microscopic plankton that use calcium carbonate to construct their skeletons or shells (see Sections 2.9.1 and 21.4.5). Paradoxically, most primary producers, and therefore potentially their consumers, can benefit from the positive effect of elevated CO_2 on photosynthesis and nitrogen fixation (see Section 3.3.2).

The distribution of surface acidification across the oceans will be far from uniform, depending as it does on ocean currents and patterns of upwelling and downwelling, which themselves can be expected to change under climate warming. Figure 22.13a shows the projected distribution of ocean pH by 2100 under RCP8.5, and superimposed are the current distributions of some of the most vulnerable animal taxa that are also of socioeconomic importance (related to fisheries and coastal protection). Based on multiple laboratory studies, Figure 22.13b shows how species in four important taxa can be expected to respond to different categories of elevated CO_2 by 2100 for RCP4.5 (500–650 μatm partial pressure CO_2, a measure approximately equal to ppm CO_2), RCP6.0 (651–850 μatm) and RCP8.5 (851–1370 μatm). RCP2.6 is unlikely to have effects different from controls (non-elevated CO_2) and is not included. A final category of CO_2 elevation (1371–2900 μatm) corresponds to RCP8.5 by 2150, highlighting in particular the vulnerability of molluscs and corals to extreme acidification.

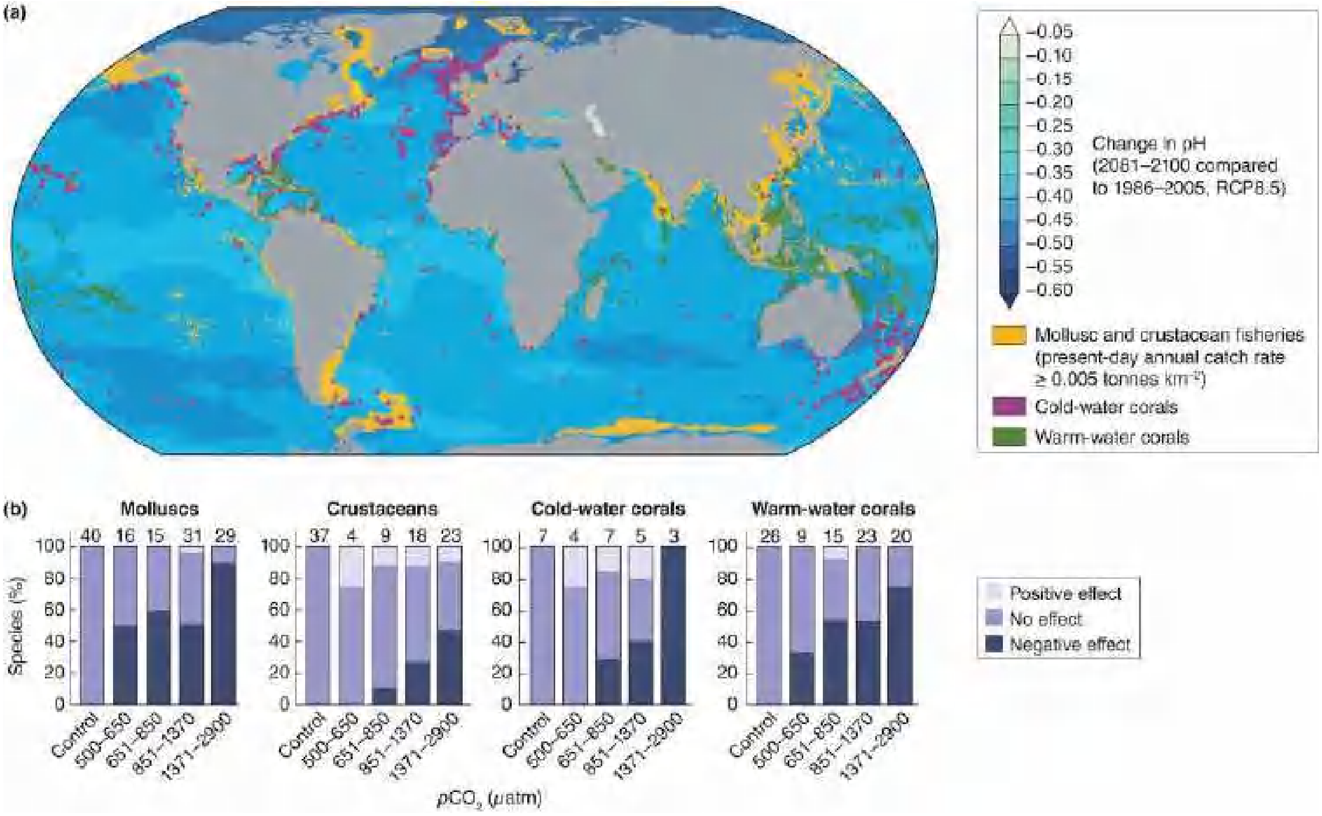


Figure 22.13 Projected global pattern of ocean acidification in relation to important fisheries. (a) Projected distribution of surface ocean acidification by 2100 under RCP8.5. Superimposed on the map are known locations of cold-water and warm-water corals and significant marine mollusc and crustacean fisheries (≥0.005 tonnes km^{-2} year^{-1}). Those in regions of more extreme acidification are at most risk (IPCC, 2014). (b) The percentage of species in four animal groups sensitive to increasing categories of ocean acidification in comparison with non-elevated controls (380 µatm), derived from multiple laboratory and field studies (number of studies shown above each histogram).

Source: From Wittmann & Pörtner (2013).

22.3.1 Interactions among drivers

More than one of the suite of anthropogenic drivers of ecological change are generally experienced together and may interact to produce ecological surprises. We discuss two examples next.

acid rain recovery and thermally sensitive lake fish

First, many lakes are sufficiently deep for stratification to occur, producing a zone of warmer water (epilimnion) at the surface separated by a thermocline from cooler water below (hypolimnion). However, with climate warming, extreme summer temperature events, predicted to become more common, can stress and kill cold-water fish especially in lakes where thermal stratification is absent or limited. Cold-water species such as brook trout (*Salvelinus fontinalis*) in lakes in eastern North America, for example, become stressed at higher temperatures and use the hypolimnion as a refuge during hot summers. The depth of the hypolimnion depends to an important degree on water clarity, with lakes that transmit less light having thermoclines that develop earlier and at shallower depths, leaving a larger hypolimnion refuge. With the implementation of more stringent air quality regulations in North America and Europe, surface water pH has been slowly increasing. This has led Warren *et al.* (2017) to point out that recovery from lake acidification can be expected to lead to reduced water clarity, as a result of an increased supply and concentration of dissolved organic carbon (DOC) in the water, providing a greater cool-water refuge for the fish in the face of warming temperatures.

acidification, climate warming and marine food webs

Second, marine plants can be expected to benefit from future climate change because both CO_2 enrichment and warmer temperatures can enhance primary production, with a potential strengthening of bottom-up control of food webs (see Section 17.1.2). On the other hand, a temperature-related increase in food demand of consumers might intensify trophic cascades and top-down control of prey. In their laboratory mesocosm experiment, Goldenberg *et al*. (2017) studied a sediment-associated, three-level food web of predatory fish (seven juveniles of *Favonigobius lateralis*), which take bites of sand to catch small herbivores present in the sand collected for the experiments (copepods, annelids and tanaid shrimps), which feed on the microalgae there. The fish were first habituated to captivity for one month, before the different microcosm treatments were introduced and followed for 3.5 months.

As expected, both elevated CO_2 (pCO_2 of 900 ppm versus ambient) and elevated temperature (+2.8°C versus ambient), or a combination of both, boosted primary production (assessed as chlorophyll a) (Figure 22.14a). When just CO_2 was elevated, both secondary and tertiary production (herbivores and fish) were also elevated, indicating bottom-up control of production through the trophic levels (Figure 22.14b, c). But when temperature was elevated, alone or in combination with CO_2, neither secondary nor tertiary production were different from ambient. It seems that elevated temperature negated the benefits of elevated CO_2 because predators consumed their prey more rapidly in the face of higher metabolic demands, strengthening top-down control and preventing an increase in secondary production.

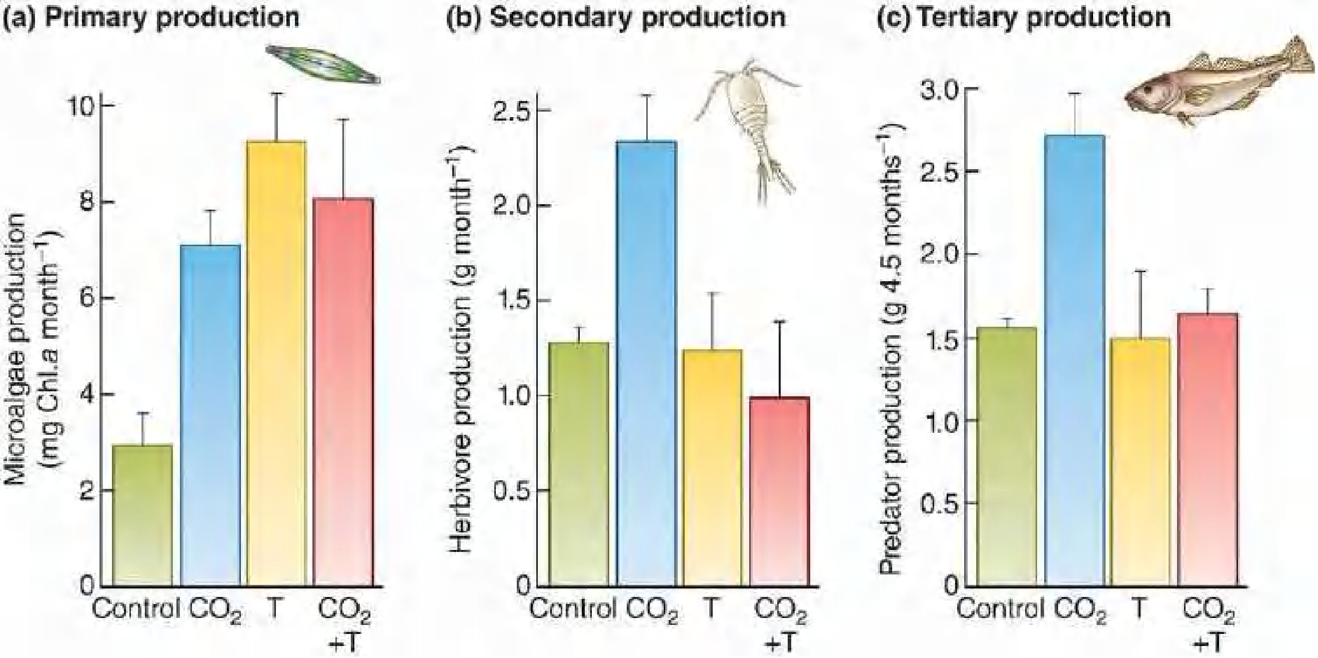


Figure 22.14 Effects of ocean acidification on different trophic levels. Effects of elevated CO_2 (ocean acidification) and temperature (T) on production (mean + SE) in a three-trophic level mesocosm experiment (three replicates per treatment) compared with ambient conditions (control). There was a statistically significant interaction between CO_2 and T at all trophic levels.

Source: From Goldenberg *et al*. (2017).

APPLICATION 22.3 Assisted evolution as a management response to climate warming and ocean acidification

Given the strong selective forces exerted by climatic factors on fitness-related traits, we must expect evolutionary responses to a changing climate, as there have been throughout the world's history. Indeed, evidence for contemporary evolution in response to climate change to date has been described for a wide range of species, including vertebrates (Grant, 2018), invertebrates (Kingsolver & Buckley, 2017) and plants (Lustenhower *et al.*, 2017). This capacity of species to evolve challenges our ability to predict potential changes to their ranges, because there is little theory to guide expectations of how climatic change will affect evolution. Notwithstanding these challenges, however, it is possible to think in terms of assisted evolution as a management tool. Coral reefs are among the most threatened of ecosystems, but management to date (marine protected areas, land catchment management to improve marine water quality, restoration) has been insufficient to maintain their health and abundance. Van Oppen *et al.* (2017) are now evaluating 'assisted evolution', which they define as the acceleration of natural evolutionary processes to enhance certain traits. These include the exposure of adult coral colonies to environmental stresses with the aim of inducing heritable increased stress tolerance, and inducing evolution in the coral's algal symbiont *Symbiodinium*, under elevated temperatures and CO_2 concentrations in the laboratory, for later inoculation into the coral hosts.

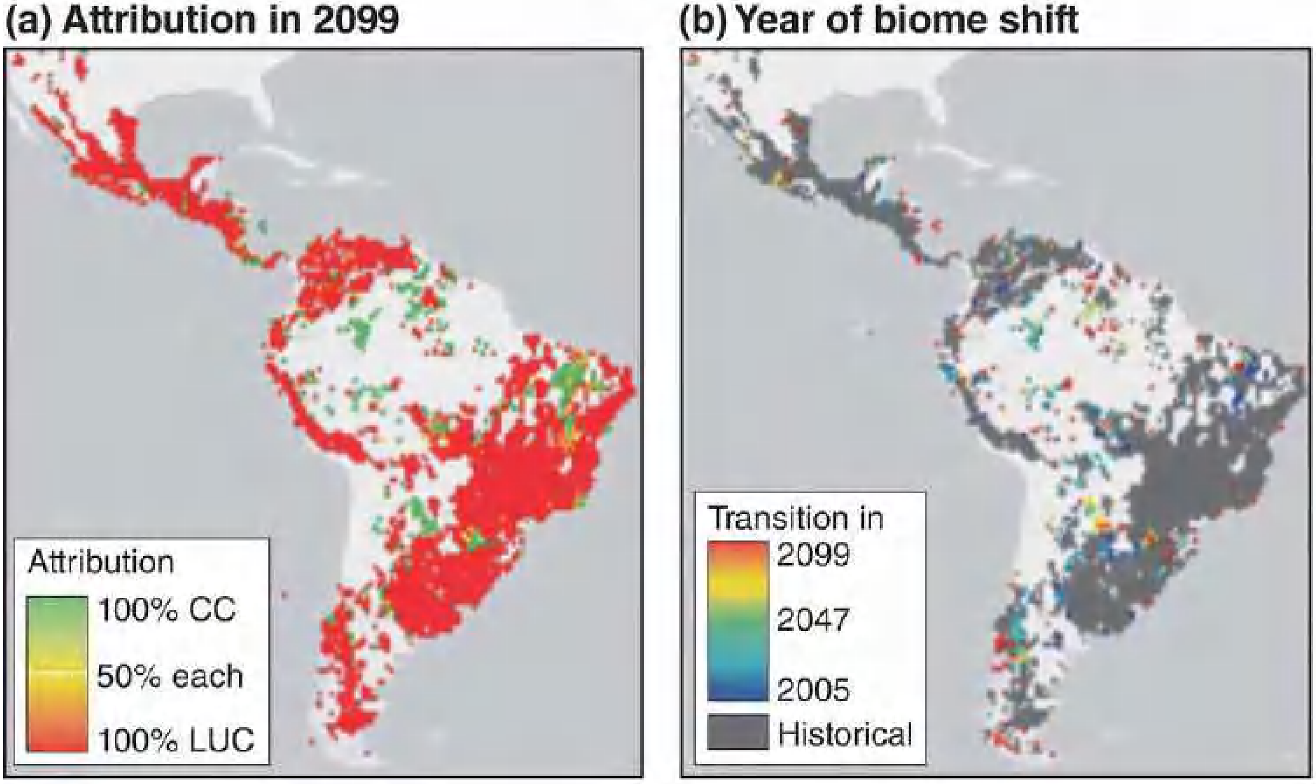
22.4 Land-system change

Humans began to have widespread impacts on ecosystems well before the beginning of the Anthropocene. This started more than 10 000 years ago with the Pleistocene extinction of giant herbivores, a megafauna including mammoths, giant sloths and giant kangaroos, whose demise was at least partly due to overexploitation by hunters. The use of anthropogenic fire to clear vegetation, the development of agriculture, beginning 5000 years ago, and more recent urbanisation and industrialisation, have all had even more profound ecological consequences. So far, we have discussed the global effects of emissions of greenhouse gases associated with power generation. Here we turn to human use and abuse of habitats. Agriculture has been particularly influential in this respect, both directly, by destroying natural ecosystems to make way for agroecosystems, and indirectly by disrupting natural nutrient cycles, with consequences that extend well beyond the boundaries of agricultural land.

22.4.1 Expansion of the anthromes

Habitat loss has undoubtedly been the most far-reaching of human influences on ecosystems during the Anthropocene and is set to remain a potent driver into the future. We saw an example in Figure 22.11 of how boundaries of biomes and anthromes (urban, pastoral and cropland areas) are projected to shift in the future. In their analysis of the situation in Latin America, Boit *et al.* (2016) estimated the proportions of changes in land cover, both historical (before 2005) and projected until 2099, that can be attributed to human land-use change as opposed to climate change (Figure 22.15a–d; and compare with Figure 22.11). Historically, it is clear that most biome shifts have been due to land-use change (compare red areas with dark grey areas). In future, climate change looks set to play a more prominent role, particularly in the case of high emission scenario RCP8.5 (green areas in Figure 22.15c) as we move towards the end of the century (yellow and red areas in Figure 22.15d).

RCP2.6


(a) Attribution in 2099
(b) Year of biome shift
Attribution
100% CC
50% each
100% LUC
Transition in
2099
2047
2005
Historical

RCP8.5

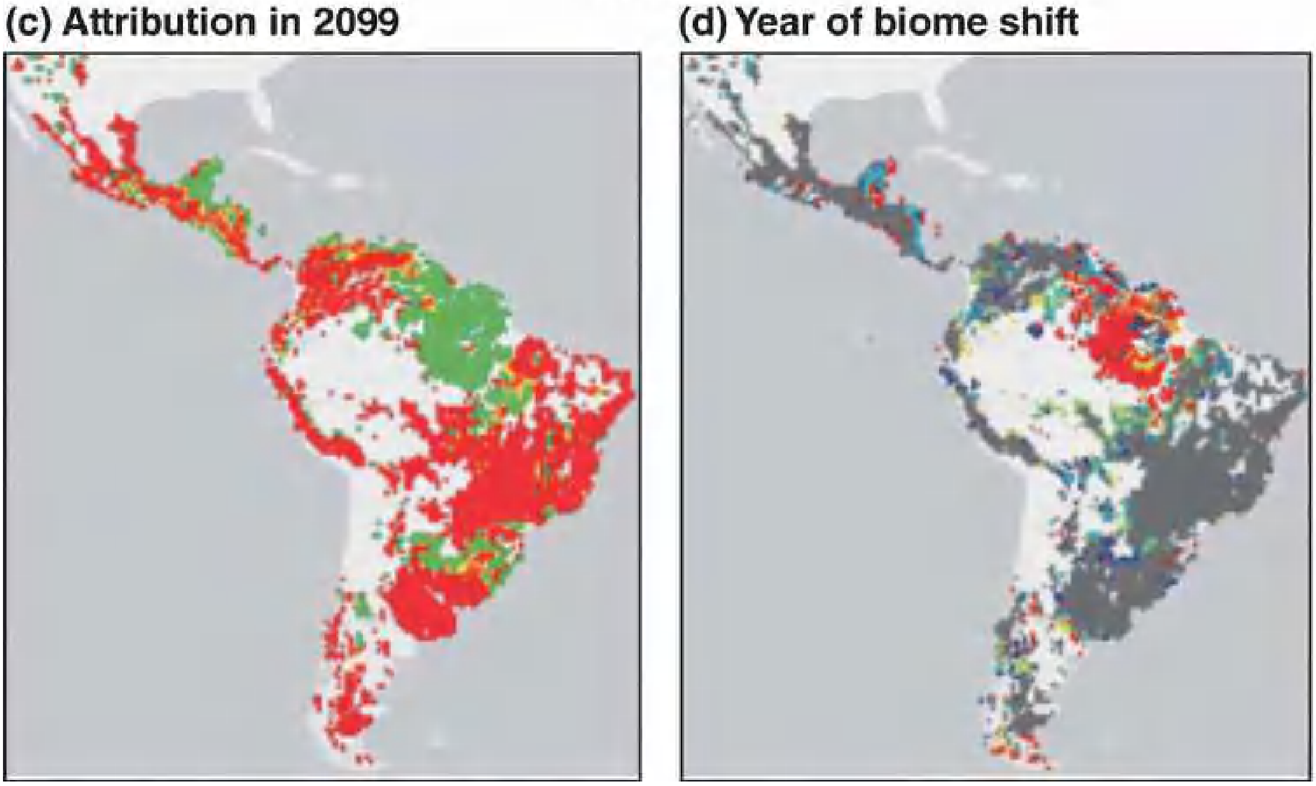

(c) Attribution in 2099
(d) Year of biome shift

Figure 22.15 Individual and combined effects of land-use change and climate change on biome distributions. Attribution of biome shifts (illustrated in Figure 22.11) to land-use change (LUC), climate change (CC) or both by 2099 under (a) the low emission scenario RCP2.6 and (c) the high emission scenario RCP8.5. The predicted years of biome shifts are shown in (b) and (d), respectively. Note that most change is historical (prior to 2005). *Source*: From Boit *et al.* (2016).

APPLICATION 22.4 Establishing and managing protected areas

A key response to global and national biodiversity declines has been to safeguard natural habitats through the establishment of protected areas (PAs) (see Application 19.1). These now cover 15.4% of the land surface globally and their importance is set to increase with the latest Convention on Biological Diversity target for 2020 being 17%. Given the objective of maintaining biodiversity within their boundaries, it is critical to quantify how well PAs are preserving wildlife and to identify factors that contribute to their effectiveness. Barnes *et al.* (2016) compiled bird and mammal long-term datasets (minimum of five years) between 1970 and 2010, for 1902 populations of 556 species from 447 terrestrial PAs worldwide (Figure 22.16). They found that on average the abundance of monitored populations in PAs was being maintained (annual population change +0.52%), with birds faring somewhat better than mammals (+1.72% and –1.0%, respectively). Populations in European PAs did better than in African PAs (+2.05% and –1.79%, respectively) and, more generally, wildlife population trends were more positive in PAs in countries with higher development scores (related to higher wealth and lower corruption). This finding has promise, because it emphasises that economic development and decreased human dependence on wildlife resources in developing countries should ameliorate wildlife declines in their PAs. A further finding is that larger species had particularly positive population trends, probably because management effort is commonly prioritised for large-bodied charismatic species. (Note, however, that the datasets end before the poaching crisis that has significantly affected African elephant and rhinoceros populations.) It seems that small-bodied species also did relatively well (probably because of their high reproductive rates) while medium-sized species, particularly in Africa, were more likely to show negative population growth rates because they tend to lack active management and have lower reproductive rates. The optimal management of PAs clearly has both ecological and sociopolitical dimensions.

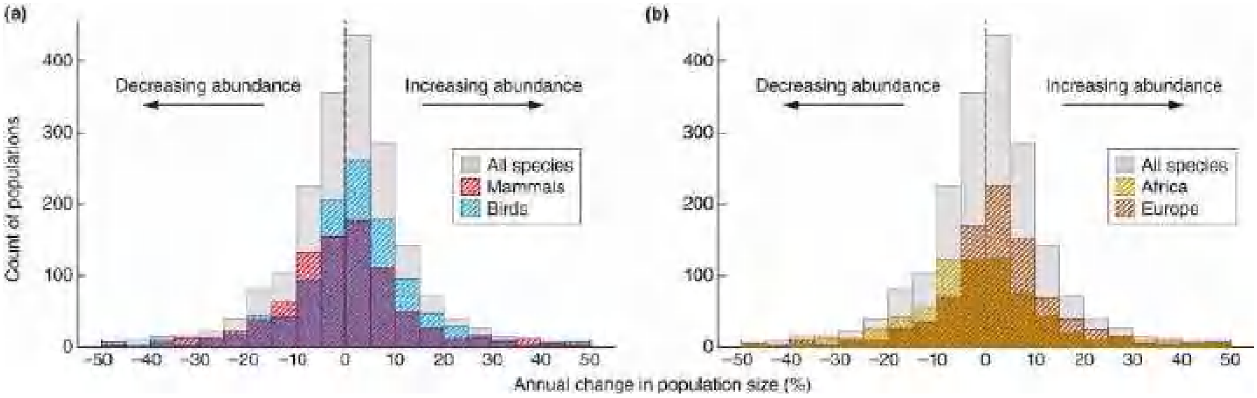


Figure 22.16 Frequency distributions of wildlife abundance changes between 1970 and 2010 in protected areas globally. (a) Abundance changes of mammals, birds and all species. (b) Abundance changes by location – Africa, Europe and all locations.

Source: After Barnes *et al.* (2016).

22.4.2 Perturbation of nitrogen and phosphorus cycles

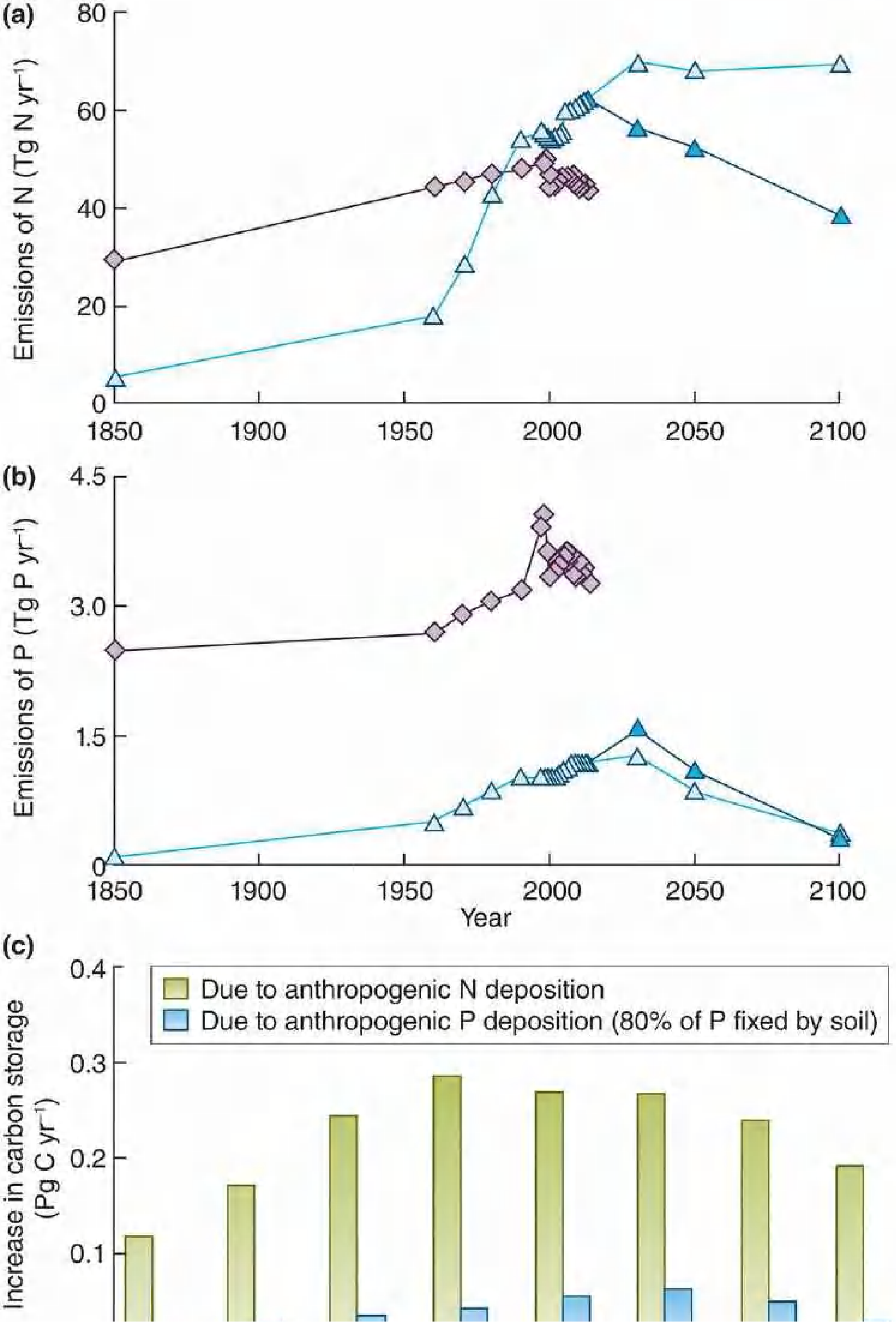
agroecosystems

Agricultural inputs of nitrogen (N) and phosphorus (P), the two key limiting plant nutrients, are now so large that they perturb the global cycles of these elements (Section 21.4). On the positive side, inputs of fertilisers have augmented food production for a growing human population. On the negative side, their inefficient use in agriculture is having profound downstream environmental effects. Inputs of both N and P are expected to increase in future.

A key event for the global N cycle was the invention in 1910 of the Haber Bosch process to artificially fix atmospheric N into ammonia, leading to dramatic increases in N fertiliser use in developed countries from the mid-1930s. Fertiliser manufacture, together with the cultivation of leguminous crops (whose bacterial symbionts fix atmospheric N), now converts about 120 million tonnes of N_2 gas from the atmosphere into reactive forms of N that can be used to enhance primary productivity of crops, and thus secondary productivity of livestock. This amount is more than the combined contributions to N fixation from all the earth's natural terrestrial processes. In contrast to N, P is a fossil mineral; about 20 million tonnes are mined per year, equivalent to approximately eight times the natural rate of influx globally (Rockström *et al.*, 2009).

forest ecosystems

It is not only agricultural systems that are transformed by anthropogenic inputs of N and P. Productivity (and thus carbon capture) of forest ecosystems is expected to be strongly affected in the 21st century by the availability of N and P, and especially by future patterns of deposition of N and P from the atmosphere. Wang *et al.* (2017) modelled emissions of P (on sea salt and dust and biogenic aerosol particles) and N (oceanic emissions of ammonia (NH_3), vegetation emissions of nitric oxide, fertiliser application in agriculture), together with fuel combustion emissions of both N and P, through to 2100, comparing emissions scenarios RCP4.5 and RCP8.5 (Figure 22.17a, b).


(a)
Emissions of N (Tg N yr^{-1})
80
60
40
20
0
1850
1900
1950
2000
2050
2100
(b)
Emissions of P (Tg P yr^{-1})
4.5
3.0
1.5
0
1850
1900
1950
2000
2050
2100
Year
(c)
Increase in carbon storage (Pg C yr^{-1})
0.4
0.3
0.2
0.1
Due to anthropogenic N deposition
Due to anthropogenic P deposition (80% of P fixed by soil)

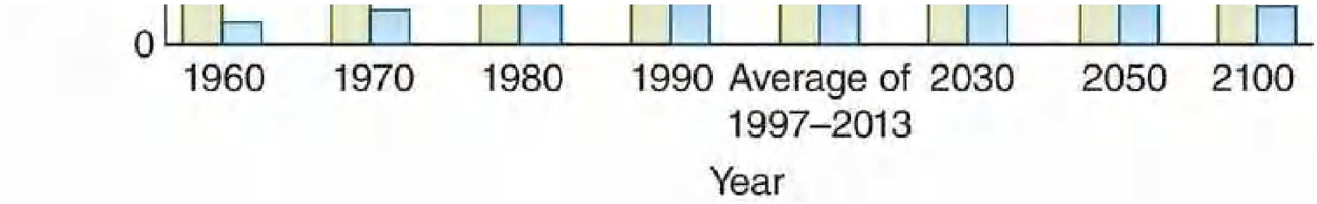


Figure 22.17 Influence of projected global nitrogen and phosphorus emissions and deposition on forest carbon storage. (a) Projected global atmospheric nitrogen (N) emissions. Blue triangles, from fossil fuels plus biofuel plus agricultural activities (pale blue, RCP8.5; dark blue, RCP4.5); purple diamonds, from biomass burning, oxides of nitrogen and ammonia (NH_3) soil emissions plus NH_3 ocean emissions. (b) Global atmospheric phosphorus (P) emissions. Blue triangles, from fossil fuels plus biofuel (open triangles, RCP8.5; closed triangles, RCP4.5); purple diamonds, from biomass burning, dust, sea salt, volcano particles plus primary biogenic aerosol particles. (c) Increase in carbon storage in the world's forests due to fertilisation by deposited N and P under RCP4.5. In the case of P, it is assumed that a large fraction (80%) is fixed by the soil as inorganic P and is unavailable for uptake by trees.

Source: From Wang *et al.* (2017).

The projected future emissions of N differ considerably under the two scenarios, with, for example, NH_3 emissions 54% higher in 2100 under RCP8.5 due to the greater agricultural activity needed to meet food demands of a larger human population. In contrast, projected P emissions differ much less, because 70% of P emissions from fossil and biofuels are expected to be removed by cleaner technology in industrial and domestic sectors under both scenarios. The consequences of these projected changes for carbon storage by the world's forests (i.e. their net productivity) for anthropogenic N alone and anthropogenic P alone are shown in Figure 22.17c. For 1997–2013, Wang *et al.* (2017) estimated that anthropogenic deposition of N and P contributed carbon storage increases of 0.27 ± 0.13 and 0.054 ± 0.10 Pg C year^{-1}, respectively (9% and 2% of the total terrestrial carbon sink). Their contribution during the rest of the century while remaining high is projected to decline somewhat.

22.4.3 Downstream effects of nutrient cycle perturbations

Significant fractions of anthropogenically mobilised N and P from agricultural, industrial and domestic sources (including sewage and wastewater) flow from the land into groundwater and surface water, affecting the ecology of streams, rivers, ponds, lakes and oceans. By modelling flows from land via individual catchment areas (i.e. land areas drained by individual rivers and their tributaries) into coastal waters, Seitzinger *et al.* (2010) have provided projections of future trends in river nutrient export at global, continental and regional scales.

socioeconomic drivers

They pay particular attention to the way that socioeconomic factors may come into play by comparing two scenarios developed for the Millennium Ecosystem Assessment (MEA, 2005a). The ‘global orchestration' scenario assumes that the whole world is focused, together, on trade, economic liberalisation and economic growth, reacting to ecosystem problems rather than being proactive. The ‘adapting mosaic' scenario, on the other hand, sees political and economic activity focused at the river catchment scale and assumes a strongly proactive and learning approach to local ecosystem management.

Projected changes in river export of N and P at the global scale until 2030 differ for the two scenarios and for the three forms of each element, dissolved inorganic (DI), dissolved organic (DO) and particulate (P) (Figure 22.18a). An increase in DIN export is projected for the global orchestration scenario (with manure the most important contributor because of assumed high per capita meat consumption) but a decrease for the adapting mosaic scenario (where manure production also increases but is more than counteracted by a decline in fertiliser use). DON, on the other hand, shows an increase from 2000 to 2030 under both scenarios, but less so for the adapting mosaic case because of better catchment management, including restoration of wetlands. DIP and DOP show increases under both scenarios, but to a greater extent in the global orchestration scenario because of bigger increases in sewage, fertiliser use, P-based detergents and manure production. In contrast, particulate forms of the two elements (PN and PP) decrease from 2000 to 2030 under both scenarios, largely because of increased trapping of particulates due to the construction of more dams and reservoirs. When results for catchment areas are aggregated, substantial differences in magnitude and direction of change are expected for different regions and continents, as illustrated for DIN and DIP under the global orchestration scenario (Figure 22.18b). Largest increases in DIN export are projected for South Asia, which accounts for just over 50% of the projected increase in the world's river export until 2030 under both scenarios, with South America responsible for another 21%. European and North American countries, on the other hand, are expected to see marked decreases. The pattern for DIP is similar.

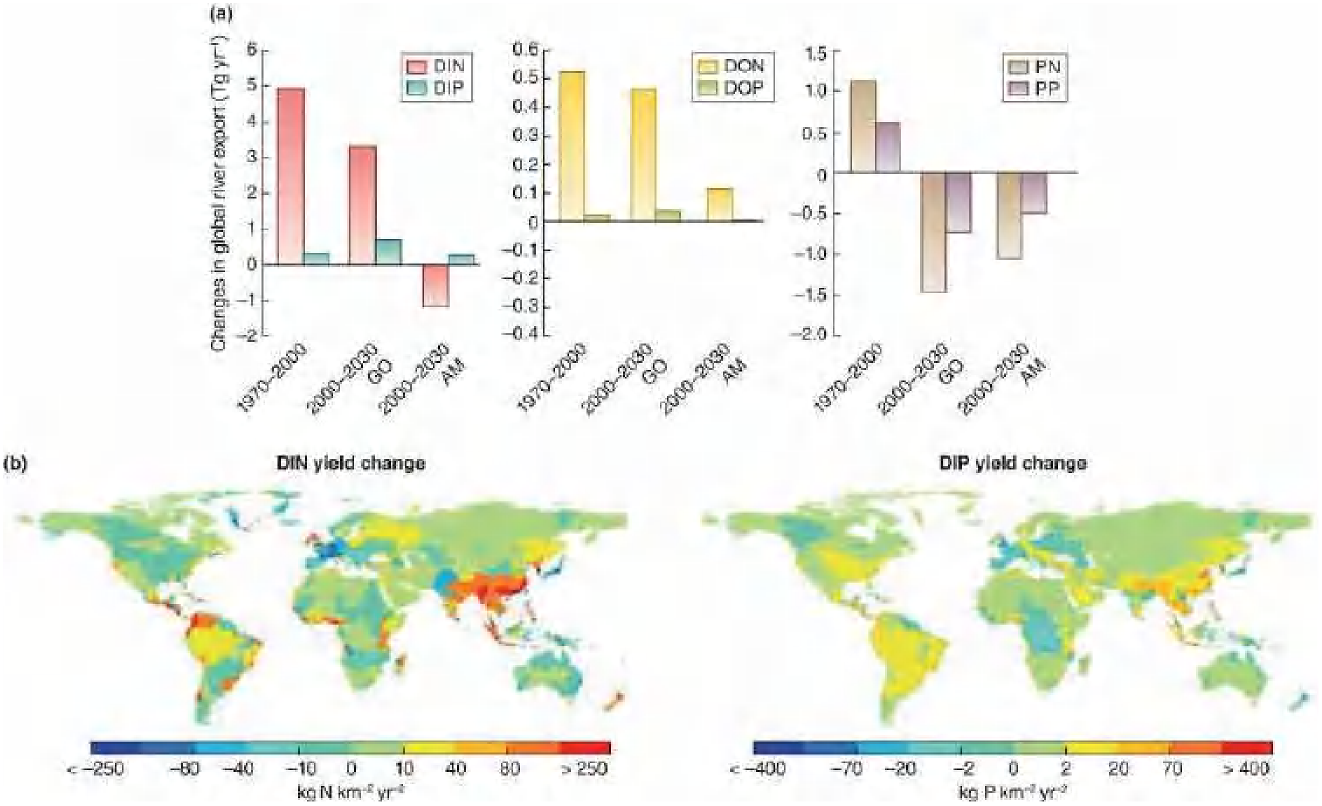


Figure 22.18 Projected global and regional patterns of change in export of nitrogen and phosphorus from rivers. (a) Changes in river export to coastal waters of three forms of nitrogen and phosphorus (dissolved inorganic, DIN and DIP; dissolved organic, DON and DOP; and particulate, PN and PP) estimated for the period 1970–2000, and projected for 2000–30 under two socioeconomic scenarios – 'global orchestration' (GO) and 'adapting mosaic' (AM). (b) Changes in regional yields of DIN and DIP for 2000–30 under the 'global orchestration' scenario.

Source: From Seitzinger *et al.* (2010).

hypoxia and anoxia in aquatic ecosystems

We saw in Section 21.1.3 how lakes, estuaries and coastal seas are strongly influenced by human activities in their terrestrial catchment areas that augment the supply of plant nutrients, fuelling algal primary productivity and increasing the rate of delivery of organic matter to bottom waters where microbial decomposition consumes oxygen. Hypoxic conditions refer to the part of the dissolved oxygen spectrum where only species highly tolerant of low oxygen levels can thrive, while anoxic conditions (a complete absence of oxygen) rule out all but microorganisms that do not depend on aerobic respiration.

Lake eutrophication, producing hypoxic conditions, began to be noticed as early as 1700 and records of hypoxic lakes began increasing exponentially before 1900, associated with the introduction of phosphorus fertiliser. This was some 70 years before a similar pattern became obvious for coastal waters, associated with increased land erosion and use of fertilisers after the Second World War (Figure 22.19a).

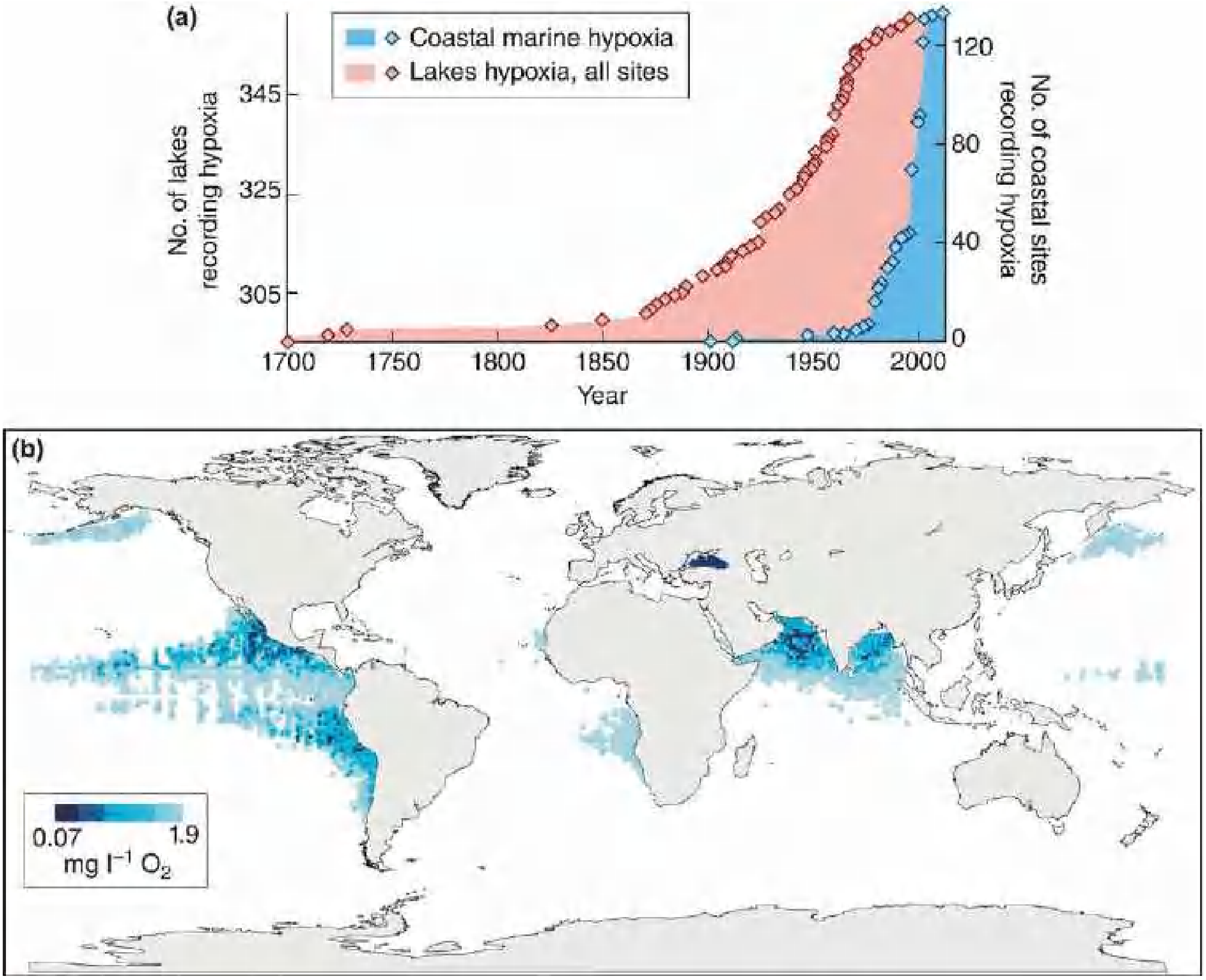


Figure 22.19 Global and regional patterns of increased incidence of eutrophication in lakes and oceans. (a) Patterns of increase in numbers of lakes and coastal sites where hypoxia has been reported worldwide. (b) Global map of ocean oxygen minimum zones.

Source: (a) Modified from Jenny *et al.* (2016), with coastal site data from Vaquer-Sunyer & Duarte (2008). (b) From Breitburg *et al.* (2018).

an interaction between eutrophication and climate warming

With an increase in temperature, water holds less oxygen and metabolic rates increase, accelerating the rate of oxygen consumption. Thus, projected global warming can be expected to exacerbate the eutrophication of lakes and coastal waters, providing a further example of an interaction between anthropogenic drivers of change. In the open oceans, much less affected by catchment runoff, there has already been a decline in oxygen associated with warming. Breitburg *et al*. (2018) estimate that 2% of open ocean oxygen has been lost in the last 50 years, and that so-called ‘open-ocean oxygen minimum zones' (<70 µmol O_2 kg^{-1} at 300 m depth) expanded by 4.5 million km^2, while the volume of water completely devoid of O_2 has more than quadrupled over the same period (Figure 22.19b). All these changes can be expected to have profound consequences for biodiversity and ecosystem functioning, and consequently for fisheries.

The general mitigation actions to counteract hypoxia and anoxia are obvious, though not necessarily easy to achieve, namely to reduce nutrient inputs and greenhouse gas emissions. Adaptation to restore and protect economically important species and fisheries could involve the creation of marine reserves in well-oxygenated areas to provide refugia when oxygen is low as well as reductions in fishing pressure on hypoxia-intolerant species (Breitburg *et al*., 2018).

APPLICATION 22.5 Strategies for catchment management

the Everglades

The Florida Everglades, one of the world's major wetlands, have been subject to ecological decline over the past 150 years because of profound land-use (agricultural and urban), water flow (drainage and flood control projects) and water quality changes (nutrient, sediment and pesticide inflows). Recognising that it has become a disconnected, compartmentalised, human-dominated landscape, it is now subject to what is claimed to be the largest restoration action in the world – the Comprehensive Everglades Restoration Plan – which is expected to need decades of effort and $13.5 billion to achieve (Schade-Poole & Möller, 2016). One of the primary ecological objectives is to reduce the rapid discharge of excess water, often with high nutrient and sediment loads, east to the St Lucie Estuary and west to the Caloosahatchee Estuary. This will be achieved by building storage reservoirs to reduce flow to these estuaries from the large Lake Okeechobee (to which they were not historically connected) and thus restore more natural flows towards the Everglades to the south. In addition, removal of dikes, levees and canals aims to restore some of the natural connectivity and flows in the Everglades too (Figure 22.20).

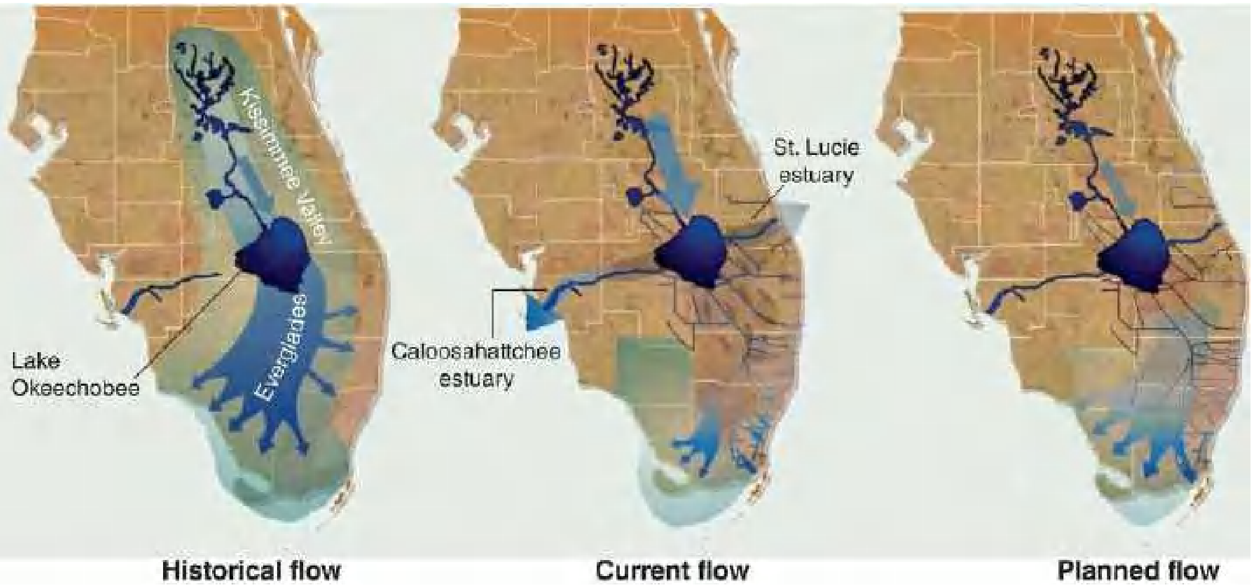


Figure 22.20 Comparison of historic, current and planned water flows in the Florida Everglades. Historically, neither the Caloosahatchee or St Lucie estuaries were connected to Lake Okeechobee.

Source: Modified from US Army Corps of Engineers, Comprehensive Everglades Restoration Plan (http://141.232.10.32/education/presentation_materials.aspx).

Increases to P concentrations have been particular problematic in the Everglades, responsible for widespread loss of the iconic sawgrass swamps (*Cladium jamaicense*), a plant that requires low nutrient conditions. There has been a two-pronged plan to reduce P concentrations (eventually to less than 0.010 mg l^{-1}). The first involves the adoption by farmers of best management practices (BMPs), for which they may gain tax credits, including erosion control, conservation buffers, improved fertiliser management and improved onsite water retention. The second involves construction of stormwater treatment areas (STAs), perhaps the largest constructed wetlands in the world. Each of six STAs, a total of 23 000 ha, is designed to operate so that a succession of aquatic plants (emergent, submerged and floating species), with their associated microbial communities, remove P from the water as it flows through. There has been considerable success in removing P, but the amount of legacy P that persists in the soil, sediments and water of the system continues to be a concern.

The Great Barrier Reef

The world's largest coral reef system, which extends over 2000 km along Australia's northeast coast, has in past decades experienced severe declines in habitats (e.g. 50% loss of coral cover) and species (including sharks, rays, marine turtles, sea snakes, seabirds, dolphins and dugongs), with further declines expected in future. Beyond the impacts of climate change and ocean acidification, dealt with in previous sections (e.g. Figure 22.13), another very influential driver has been human activity in the terrestrial catchment areas of eastern Australia that discharge into the Great Barrier Reef (GBR) ecosystem, leading to high nutrient, sediment and pesticide loads. Implementation of a series of reef management plans since 2003 has focused on achieving targets for water quality, and land and catchment management, by encouraging uptake, within existing agricultural practices, of a series of BMPs. These include maintenance of pasture cover during the dry season and drought years, increasing the proportion of erosion-resisting, deep-rooted perennial grasses in pasture, retaining riparian trees, reducing excess fertiliser application, particularly in intensive crops such as sugar cane, bananas and cotton, and better targeting of pesticides. While there has been progress in the uptake of agricultural BMPs and some improvements in river pollutant loads, it is highly unlikely that the Reef Plan's 2020 goal of ‘no detrimental impact' in the GBR ecosystem will be met. Kroon *et al.* (2016) argue that more substantial agricultural changes will be necessary, including some land retirement and switches to new agricultural products or land uses that reduce pollutant export, such as shifting from grazing by erosion-promoting hoofed animals to native kangaroos, more emphasis on low-input crops such as grains, cereals and macadamia trees, and hydrological restoration of wetlands that were drained long ago.

22.5 Pollution

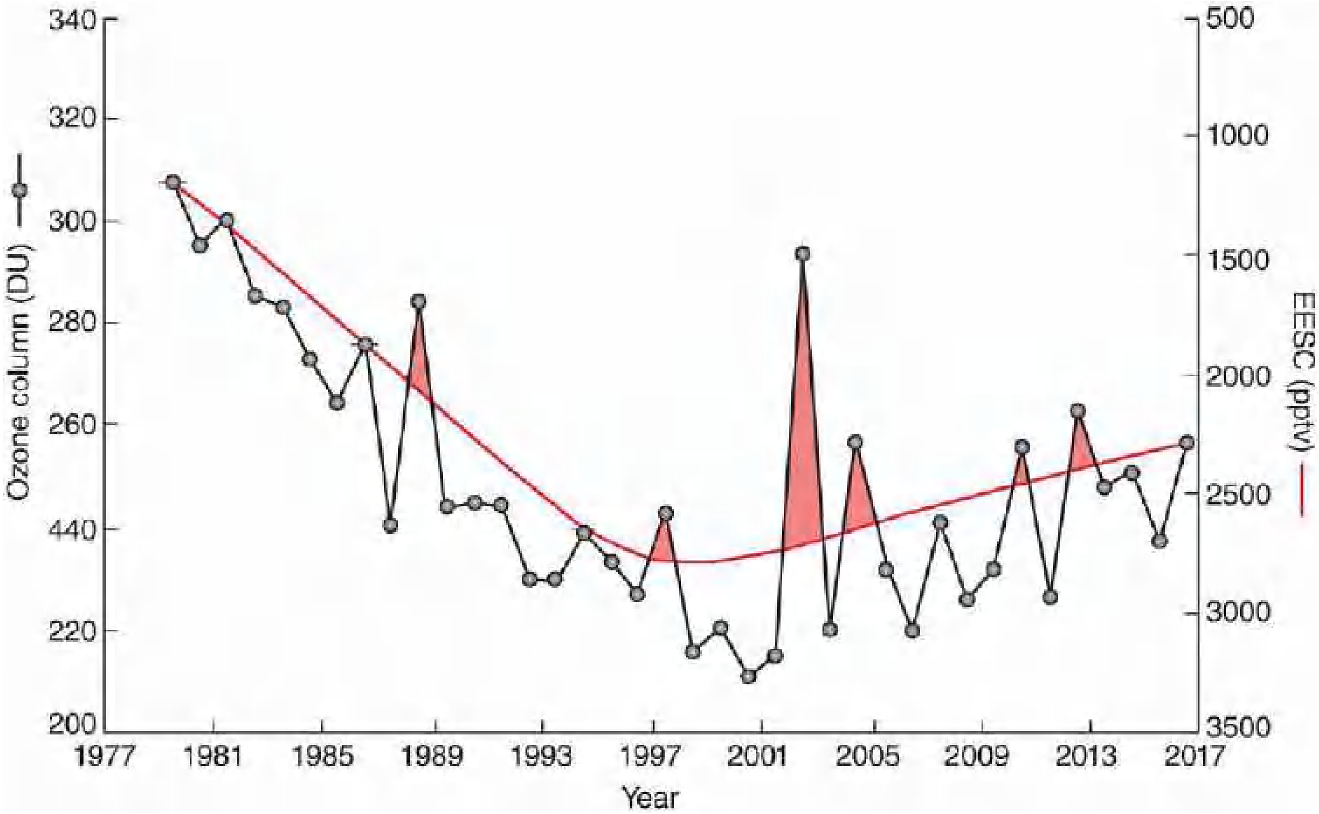
Pollution can be broadly defined as the anthropogenic addition to the natural environment of any solid, liquid, gas or energy (heat, sound, radioactivity) that has adverse environmental consequences. Thus, it should be obvious that much of this chapter has been about pollution (greenhouse gas emissions, acidification, discharge of P, N and sediment in rivers). In this section, we deal with three further examples of global ecological importance, involving damage to the stratospheric ozone layer, the biomagnification of persistent anthropogenic chemicals in food webs, and plastic waste.

22.5.1 Chlorofluorocarbons, ozone depletion and UVB radiation

The stratospheric ozone layer (located between 10 km and 50 km altitude) protects life from harmful solar ultraviolet radiation (particularly UV-B) which can damage biological molecules such as proteins, lipids and DNA. Higher levels of UV-B have been linked to increased human melanoma rates, deleterious effects on terrestrial microorganisms, mosses, tree growth rates, and food webs (generated by changes to plant physiology and growth form), and on marine microorganisms, macrophytes, phytoplankton, invertebrates and fish (Gouveia *et al.*, 2015). Eutrophic freshwaters and coastal waters are likely to be less affected, because dissolved and particulate matter in the water column can absorb and scatter incident radiation.

Antarctic ozone hole

Thinning of the ozone layer around the world is a classic example of anthropogenic pollution with substantial global consequences. First detected in the 1980s, a strong reduction in the thickness of the springtime Antarctic ozone layer became known as the ‘ozone hole’ and attracted considerable attention from scientists, the public and policy makers. Ozone thinning has not been so marked in Arctic regions, but very low ozone can occur there under particularly cold conditions in the lower stratosphere. It did not take long to unequivocally link stratospheric ozone depletion with emissions of chlorofluorocarbons (CFCs), and related substances, whose use in industrial and domestic applications (e.g. refrigerants, aerosol sprays) began rising sharply in the 1950s. UV radiation breaks down CFCs and the resulting chlorine radicals catalyse the destruction of stratospheric ozone. Concerted international action has resulted in a levelling off and reversal of springtime ozone losses since the late 1990s, although pre-1980 levels have not yet been reached. The Antarctic pattern in the amount of ozone in the ozone column (from the surface to the edge of the atmosphere) shows a remarkable fit to a combined measure of ozone depleting substances expressed as equivalent effective stratospheric chlorine (EESC). The outliers in Figure 22.21 are years when there was reduced polar stratographic cloud, which cause a reduction in catalytic destruction of ozone. The problem of ozone thinning provides an example of anthropogenic pollution that is being successfully reversed, and hopefully any ecological effects of UV damage should not remain a concern too far into the future.



[Figure 22.21](#) Ozone thinning in the Antarctic is being reversed as a result of international action. Average ozone column (in Dobson units (DU), grey circles and line) from multiple satellite measurements for the region south of 60°S based on days 220–280 of the year (the Antarctic spring). Also shown is equivalent effective stratospheric chlorine (EESC, in parts per trillion by volume (pptv), red line); this is a combined measure of anthropogenic ozone depleting substances in the atmosphere. Note the inverted scale for EESC: as EESC increases, the ozone column diminishes. The outliers shown as red shading are for years when there was reduced polar stratographic cloud.

Source: From de Laat *et al*. (2017).

22.5.2 Mercury and persistent organic pollutants

Mercury is a global concern for human and ecosystem health because of biomagnification in aquatic food webs of the methylated form, methylmercury, which binds with proteins. Anthropogenic mercury emissions to the atmosphere and ocean derive mainly from metal mining and coal-fired power stations. Methylation occurs in shallow, mesopelagic and deep waters and mercury concentrations in the oceans can be expected to continue increasing (Alava *et al*., 2017).

Two classes of persistent organic pollutant (POP) of particular concern have been the polychlorinated biphenyls (PCBs, used in hundreds of industrial products including electrical equipment, plasticisers and pigments) and dicholorodiphenyl-trichloroethanes (DDTs, used as pesticides). These fat-soluble POPs are bioaccumulated by many organisms and biomagnified in food chains, producing toxicological effects in top predators such as birds of prey and marine mammals. In much of the world their use can be considered historical because deleterious ecological (and human health) effects became obvious and they were largely phased out in the 1970s. However, some are still used in developing countries and cycling of POPs in the marine environment lingers in many industrial regions and beyond.

interaction with climate change

Moreover, the effects of POPs, and of mercury, can interact with climate change, via direct bioenergetic impacts of temperature on consumption and elimination rates, and also through changes in primary productivity and distributions of predators and their prey. While a few studies have documented declines in faunal POP concentrations, the majority of recent studies (post-2003) from polar regions have documented steady or increasing POP concentrations in polar bears and fish and mercury in sea birds (Alava *et al.*, 2017). In eggs of thick-billed murres (*Uria lomvia*, birds that breed in the Canadian Arctic), for example, mercury concentrations increased between 1975 and 2013 due to changes in diet and trophic position driven by climate change (Braune *et al.*, 2014), while POPs in East Greenland polar bears increased annually from 1984 to 2011 because of a climate-induced dietary shift from nearshore, ice-associated seals to more contaminated open water seals (McKinney *et al.*, 2013) and, finally, Adélie penguins (*Pygoscelis adeliae*) on the Western Antarctic Peninsula have shown no decrease in DDT concentrations in the last 30 years despite its hardly being used, most likely because of DDT release from melting glaciers (Geisz *et al.*, 2008).

22.5.3 Plastic waste

One of the most ubiquitous pollutants of the Anthropocene is plastic waste, with, for example, somewhere between 4.8 and 12.7 million tonnes entering the oceans each year. Macroplastics have been recognised for some time to have a negative influence on a very wide range of seabirds, crustaceans, fish, sea snakes, turtles and marine mammals as a result of ingestion (blocking the digestive track and leading to starvation) or entanglement. Of growing concern, however, are the microplastics (<5 mm) that result both from the degradation and fragmentation of macroplastics and the direct input of microplastic nurdles (raw plastic pellets and beads used in the production of plastics), microbeads used as abrasives in cosmetics and air blasting, and clothing fibres. The ecological consequences of microplastic ingestion is beginning to be understood, for example, causing reductions in fitness of earthworms (Huerta Lwanga *et al.*, 2016) and marine polychaetes (Wright *et al.*, 2013), facilitating the accumulation of sorbed organic pollutants in fish (Wardrop *et al.*, 2016), and even slowing the rate of sinking of zooplankton faecal pellets to the ocean bed and thus altering nutrient cycles (Cole *et al.*, 2016).

microplastics as a novel substrate for bacteria

Microplastics are abundant in the effluent of wastewater treatment plants, which act as point sources for rivers, estuaries and oceans. A comparison of microplastic concentrations upstream and downstream of treatment plants on nine rivers in Illinois, USA, showed a mean daily microplastic flux of 1.34 million pieces of mainly polypropylene, polyethylene and polystyrene (McCormick *et al.*, 2016). Of special interest was the finding that microplastic was associated with a particular bacterial assemblage, dominated by taxa that may degrade plastic polymers (e.g. *Pseudomonas*) or that represent common human intestinal pathogens (e.g. *Arcobacter*).

This ability of microplastics to serve as a novel substrate for the transport of certain bacteria has significance for the health of corals, already under threat from so many other drivers, because microplastics can promote colonisation by pathogenic bacteria in the genus *Vibrio*, one of a globally devastating group of coral diseases known as ‘white syndromes' (Lamb *et al.*, 2018). In a survey of 125 000 reef-building corals on 159 coral reefs in the Asia-Pacific region (which has more than half of the world's coral reefs), one third were recorded as having plastic waste (2–10.9 items per 100 m^2). In corals that were not in contact with microplastic, while all common coral diseases were detected, the likelihood of disease being present (its prevalence) was on average 4.4 ± 0.2%. In contrast, corals in contact with microplastic experienced a 20-fold increase in likelihood of disease to an average 89.1 ± 3.2%. The underlying mechanism is unclear, but is presumably linked with the presence on plastic debris of pathogenic bacteria, though plastic-induced physical injury facilitating pathogen invasion may also be involved. Given that 80% of marine plastic debris originates from the land, human population density in coastal regions and the quality of waste management systems largely determine which countries contribute the greatest plastic loads. By 2025, under a business-as-usual scenario, the cumulative amount of plastic waste entering the oceans from land is projected to increase by about 1% in developed countries such as Australia, but by as much as 200% in low-income countries (Figure 22.22).

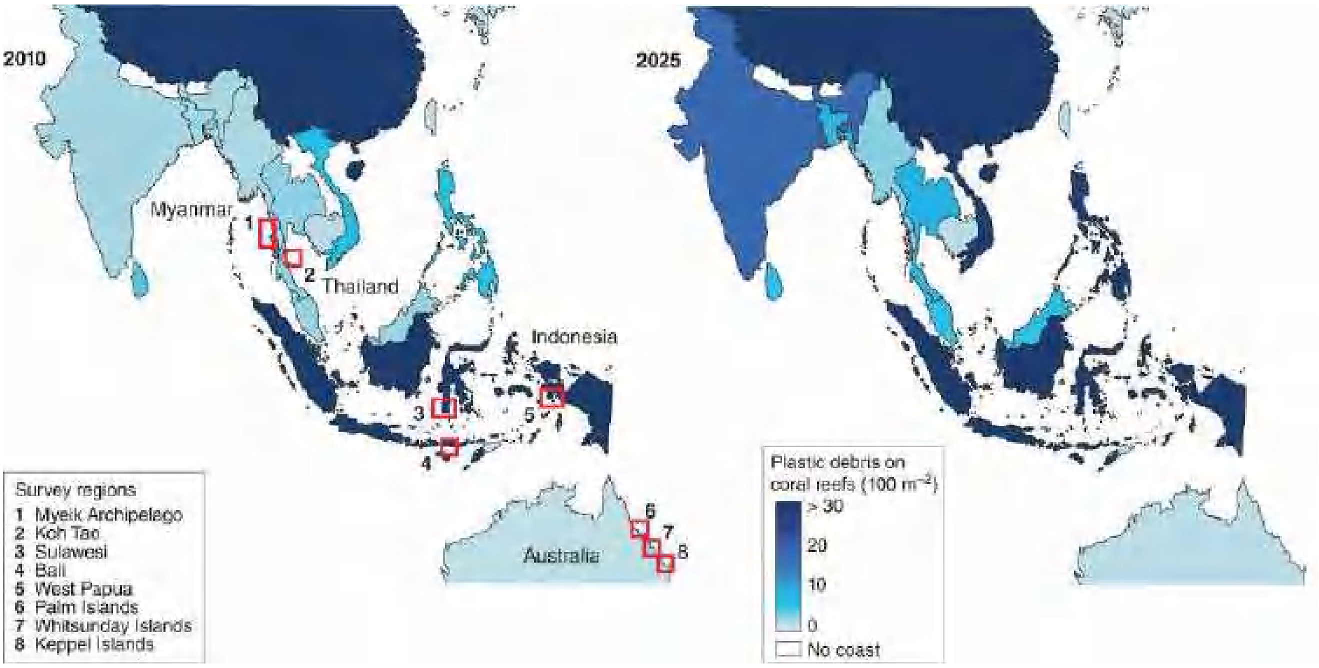


Figure 22.22 Current and projected distribution of plastic on coral reefs. Modelled distribution of plastic debris on coral reefs based on surveys of 159 coral reefs in eight survey regions in 2010, and projected under a business-as-usual scenario for 2025. The colour shading for each country indicates plastic debris loads on coral reefs associated with that country's coastline.

Source: From Lamb *et al.* (2018).

microplastics in marine food chains

Microplastics tend to aggregate in surface waters, where they are ingested by zooplankton and fish, but particles denser than sea water may sink directly toward the seabed. Mesopelagic fish that ascend to feed at the surface at night, following the vertical migration of their zooplankton prey, also transfer microplastics from the surface to deeper zones (Lusher *et al.*, 2016). In addition, trophic transfer of microplastic occurs from primary consumers to their predators at both the surface and in deeper waters, involving predatory fish, seals and even whales (Lusher *et al.*, 2015). Microplastic fibres that find their way to the seabed, either sinking directly or present in sinking faeces and dead bodies, become available to a wide range of consumers on the sea floor, whether this is shallow or deep (Taylor *et al.*, 2016), including filter feeders (anemones), deposit feeders (sea cucumbers), detritivores (hermit crabs) and predators (squat lobsters). The extent to which the ubiquitous presence of microplastics in marine fauna affects biodiversity or ecosystem functioning remains to be determined but is likely to be significant.

APPLICATION 22.6 International action to deal with global pollution problems

When pollution problems are global, due to persistence of the pollutants and their widespread dispersal through atmosphere and oceans, the solution demands international action.

ozone-depleting substances

The ozone hole in the southern hemisphere and ozone depletion in the northern hemisphere led to the Montreal Protocol, which entered into force in 1989 and is now ratified by 196 states plus the European Union, perhaps the single most successful international agreement to date. The agreement was for chlorofluorocarbons to be phased out completely by 1996 while the less damaging hydrochlorofluorocarbons are scheduled for phasing out by 2030 (UNEP, 2017a). The success of this international action can be gauged by measured increases in stratospheric ozone (see Figure 22.21), expected to return to pre-1980 levels by mid-century, and by noting that modelling suggests the protocol, with subsequent amendments and adjustments, will have saved 80% of global ozone by 2100. Consequently, in the absence of the protocol, UV radiation would have increased by factors of 4 to 8 in low and mid latitudes and as much as 16-fold in southern high latitudes, while with implementation of the protocol, UV should decrease by 2100 in comparison to pre-ozone-hole levels in mid latitudes by 5–10%, and in high latitudes by 20% in the northern hemisphere and 50% in the southern hemisphere (Egorova *et al.*, 2013).

mercury

The Minamata Convention (UNEP, 2017b) is named after the city in Japan where local communities were poisoned by mercury-tainted industrial wastewater in the late 1950s. Concerted international discussions about mercury pollution began in 2001 and the agreement, signed by 128 states, came into force in 2017. It requires parties to cap and subsequently eliminate primary mercury mining (over a 15-year period), forbids parties to export mercury or to manufacture or export specified mercury-added products after their phase-out dates, and to use best practice to control and where feasible reduce emissions of mercury and mercury compounds.

persistent organic pollutants

The analogous agreement for persistent organic pollutants, the Stockholm Convention, entered into force in 2004 and has now been ratified by 182 countries (Stockholm Convention Secretariat, 2010). This requires the parties to take measures necessary to eliminate production, use, export or import of a long list of chemicals, except where an 'acceptable purpose' has been specifically defined (e.g. DDT for ‘disease vector control in accordance with World Health Organization guidelines and where locally safe, effective and affordable alternatives are not available’).

microplastics

Microplastics have only recently been identified as an emerging issue, but already there has been some meaningful action (Sutherland *et al.*, 2010), with governments, including the USA, UK and New Zealand, enacting legislative bans on the use of microbeads in cosmetics and detergents, and many cosmetic companies committing to halt their use of microplastics by 2022. Policies are beginning to emerge to limit the use of plastic more generally, for example with decisions to phase out plastic bags and single-use plastic in supermarkets.

the Paris Climate Agreement

To respond to what is arguably the farthest-reaching of human-caused pollution problems, greenhouse gas emissions and climate warming, the Paris Climate Agreement was adopted by consensus on 12 December 2015 and 195 states had signed up and 176 had become party to it by April 2018. The agreement aims to respond to the threat of global warming by keeping the temperature increase by the end of this century to less than 2°C above preindustrial levels and to pursue efforts to keep the rise below 1.5°C. Each country has to determine, plan and regularly report on its contribution to mitigate warming. There is no mechanism to force countries to set a specific target by a particular date, but planning is currently underway around the world.

22.6 Overexploitation

wild fish as a human resource

Four hundred million people depend critically on wild fish for their food and, more generally, fisheries provide three billion people with nearly 20% of their average intake of animal protein (IPCC, 2014). Thus, historical and current overfishing of many fish stocks is of great regional and global concern. Total world catches from marine fisheries stabilised at about 90 million tonnes per year in the mid-1990s despite increasing fishing effort and technological capacity. Countries around the world have provided numerous examples of a pattern of fisheries development, growth, overexploitation and collapse since the 1950s, and indeed, with economic globalisation and inefficient management systems, it seems that the pattern of overexploitation may occur sequentially in geographically distant fisheries involving the same species: thus, Merino *et al.* (2011) argue the need for global institutions with broad authority.

overfishing joins a plethora of other threats

We have already seen how fisheries are under threat from a plethora of anthropogenic drivers, including climate change, mismatches between predators and prey, acidification, eutrophication, oxygen depletion, loss of coral habitat, ozone depletion, and pollution by persistent organic pesticides and plastic waste, among many others. Given the complexity of ensuring fisheries are sustainable (see Section 15.3), every shift to distributions of fish or their prey, or to growth, reproduction and mortality rates of target species, has profound implications for fisheries management. But in addition to these looming problems, a sustainable future will be beyond reach unless science-based approaches to limit catches are widely taken up.

To illustrate possible futures for fisheries, Serpetti *et al.* (2017) created an integrated ecosystem model for a 110 000 km^2 area off the west coast of Scotland based on food-web interactions and focusing on the effects of fishing pressure and the impact of rising temperatures relative to the thermal ranges of a variety of species. Six scenarios were considered: ‘status quo' where model drivers were set with fishing mortalities and water temperature as per the last year of the historical dataset; ‘maximum sustainable yield' (MSY) which set single-species fisheries mortalities at what would have been their estimated maximum sustainable yields for the last year of the historical dataset; and four ‘rising temperature scenarios', which kept fishing pressure constant at rates consistent with the maximum sustainable yield, but superimposed temperatures derived from increasing greenhouse emission scenarios RCP2.6, RCP4.5, RCP6.5 and RCP8.5.

changes to fishery species composition under climate change

Note that without any temperature increase, catches of all species except herring would be projected to increase in future if the fisheries are managed to MSY levels (i.e. the status quo is associated with smaller projected catches) (Figure 22.23). Species with the lowest optimum temperatures (cod, *Gadus morhua*, and herring, *Clupea harengus harengus*) were highly sensitive to relatively small changes in temperature predicted under the greenhouse gas emission scenarios. For these species, declines in biomass and catches are projected under all four scenarios. Haddock (*Melanogrammus aeglefinus*), with a slightly higher optimal temperature, showed a steady equilibrium under the best-case emissions scenario (RCP2.6) but declined under the other three. Saithe (also known as pollock, *Pollachius virens*), with its higher optimum temperature and broad temperature tolerance range, is predicted to be more resistant to rising temperatures, with a reduction in biomass only projected for the most extreme emissions scenario (RCP8.5). In stark contrast to all these, whiting (*Merlangius merlangus*) is projected to increase under all warming scenarios, as a result of its higher optimum temperature but also declines in its predators (cod and seals). In terms of overall catch, increases for saithe and whiting should compensate for projected declines in other species, highlighting the likelihood of significant changes in fish communities under climate change.

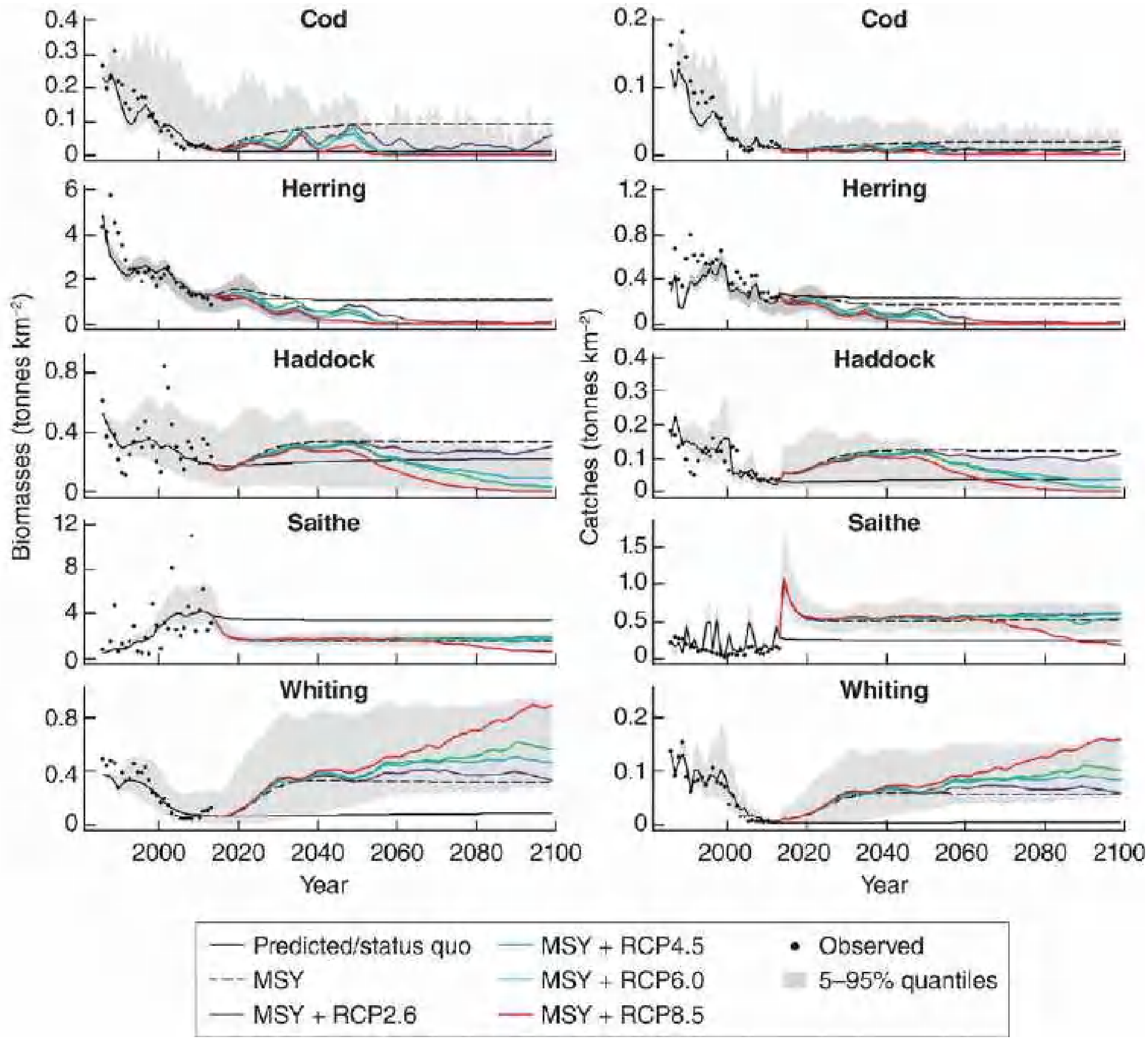


Figure 22.23 Projected changes in catch for five important fisheries. Fitted observed fish biomasses (left panels) and catches (right panels) for five fisheries species 1985–2013 (black dots) and hindcasted model output (solid black line, with 95% and 5% percentiles from Monte Carlo simulations, shaded area), together with future projections under six scenarios (see key). Hindcasting involves entering historical data into the model to test how well it ‘predicts' the known past.

Source: From Serpetti *et al.* (2017).

major fishing areas in regions of greatest climate change

Ramirez *et al.* (2017) seek a broader, global perspective on human impacts on marine fisheries by analysing changes since 1980 in 0.5° pixels across the globe in three sets of satellite remote-sensing records related to climate change – trends (slopes) in sea surface temperature (SST), ocean currents and primary productivity (chlorophyll a, CHL) – and combining them into a single environmental index with values from 0 to 1. They note first that hot spots of marine biodiversity (based on 2183 species of fish and marine mammals and birds) have all experienced environmental perturbations (i.e. have high environmental indexes), the most striking being in the central-western Pacific and south-western Atlantic as a consequence of significant changes in

SST and CHL. They then point to a worrying coincidence that the world's most biodiverse areas are not only likely to be substantially affected by climate change but are also among those experiencing the most intensive industrial fishing, and not only by nations operating within their own exclusive economic zones. It is therefore critical that fishing policies are designed from an international perspective to minimise potential synergistic effects of industrial fishing and climate change on both fisheries and marine biodiversity.

APPLICATION 22.7 Overfishing – the way forward

Barner *et al.* (2015) express some optimism that despite the many deleterious anthropogenic pressures on fisheries, several complementary strategies can provide for the capacity of marine ecosystems to sustain increased fish consumption, as long as fisheries are sustainably managed. They point to threefold needs. Firstly, reforms that mandate ending overfishing and rebuilding stocks using scientifically determined and strictly enforced annual catch limits. Secondly, greater use of rights-based fisheries (RBF) management. And thirdly, well-designed and managed networks of fully protected marine reserves. Rights-based fisheries (also known as 'catch shares') assign fishers, communities or cooperatives secure tenure rights to a fishery, which can be expected to incentivise environmental stewardship by those involved (Figure 22.24). There are two types of RBF: rights to a given fraction of a scientifically determined total allowable catch, or spatial rights to harvest in a specific region (known as territorial use rights in fisheries or TURFs). When properly designed, such strategies have shown some success in preventing fishery collapses, improving compliance with catch limits, stabilising catches and reversing some of the damage of overfishing (Barner *et al.*, 2015). Marine reserves, whether large open water reserves or networks of coastal reserves, not only protect biodiversity and ecosystem functioning, but can also lead to increases in fish abundance in neighbouring areas. Granting fishers TURF rights to fish in areas adjacent to reserves may optimise both conservation and fishing goals, and avoid conflicts over the establishment of reserves. Indeed, there have been examples where the granting of TURFs to fishers led them to create their own marine reserves (Ovando *et al.*, 2013).

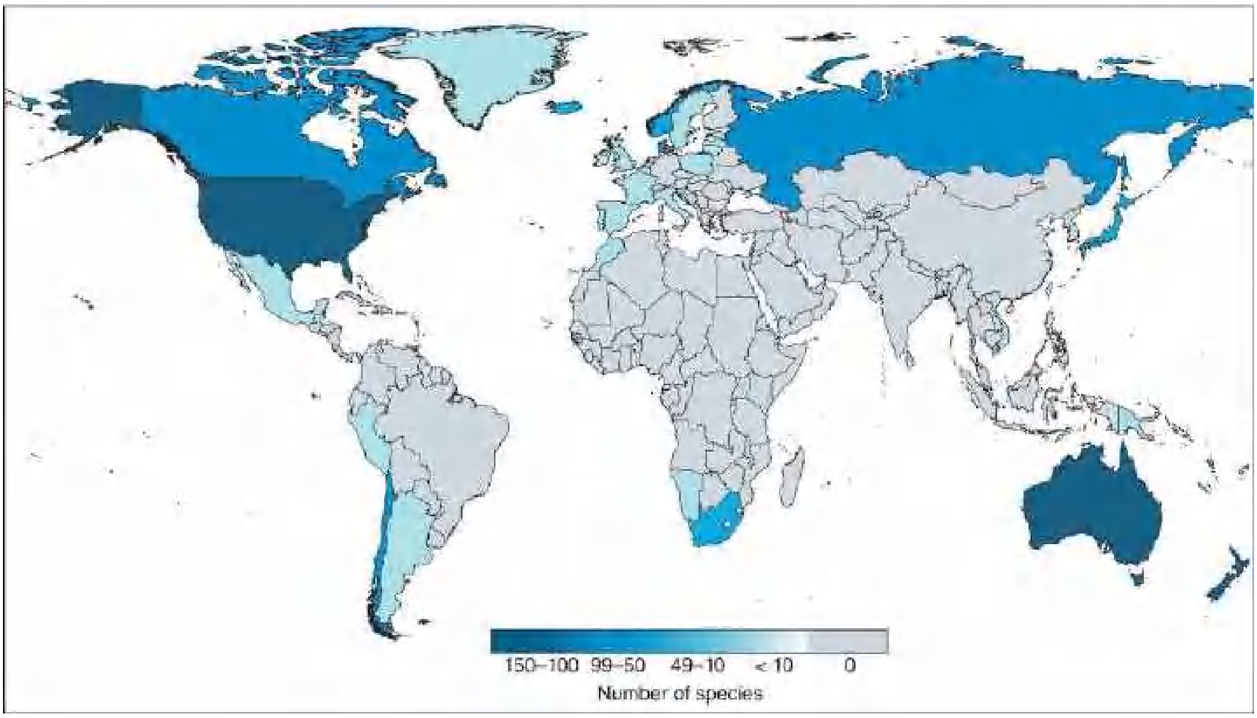


Figure 22.24 Map of global rights-based fishery (RBP) programmes. These are graded according to the number of species managed by RBP by country.

Source: From Barner *et al.* (2015).

Beyond the exploitation of natural populations, aquaculture will also play a very significant role in meeting the growing demand for fish and shellfish consumption (currently 63 million tonnes produced annually; IPPC, 2014), but extra attention also needs to be given to practicing aquaculture in a sustainable manner. Shellfish aquaculture appears particularly vulnerable to the effects of acidification.

22.7 Invasions

We have already discussed the history of the spread, by human agency, of invasive species around the world, many of which have devastating consequences (see Application 1.5), and discussed the role of ecological niche models to predict future potential distributions of damaging invaders (e.g. Figures 2.5, 6.5, 6.21 and 6.22).

22.7.1 Winners and losers among invaders under climate change

weeds ...

It should not be imagined, however, that climate change will inevitably facilitate the spread of destructive invaders. Merow *et al.* (2016) considered the establishment risk of two invasive Eurasian weed species in New England, USA. Both *Alliaria petiolata* (garlic mustard) and *Berberis thunbergii* (Japanese barberry) are present in New England but currently occupy only a fraction of potentially suitable locations. Using demographic data obtained from 21 experimental plots in diverse locations throughout the region, they compared predicted biogeographic ranges under current conditions with distributions projected using intermediate emission scenario RCP4.5. The model results suggest that garlic mustard is likely to have much lower establishment in New England under future climate, despite prolific success under current climate, whereas Japanese barberry is likely to do better in future (Figure 22.25).

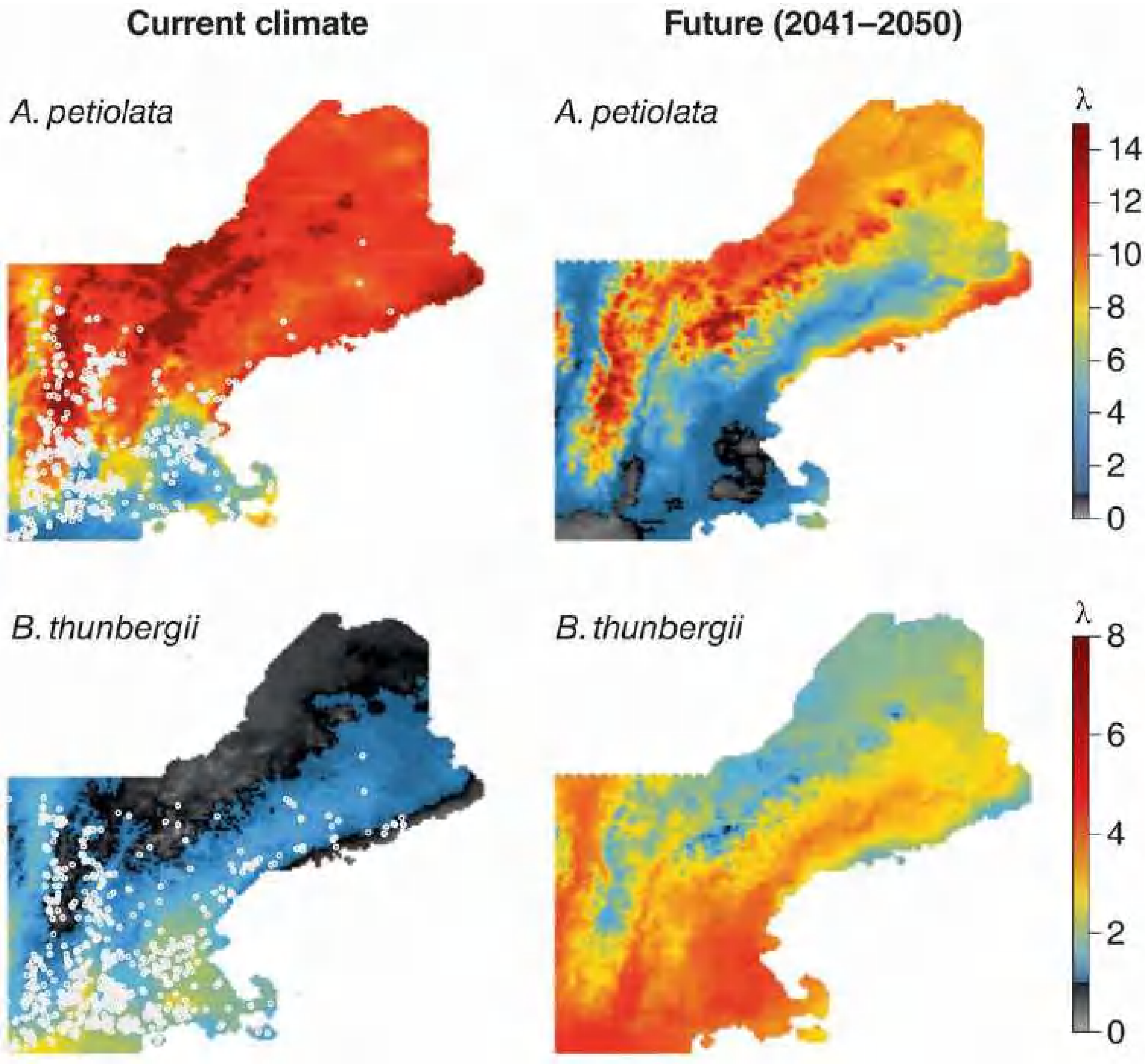


Figure 22.25 Projected changes in the distribution of two invasive plant species. Predicted distributions in New England under current climatic conditions and future conditions (2041–50) based on emission scenario RCP4.5 of invasive *Alliaria petiolata* (garlic mustard) and *Berberis thunbergii* (Japanese barberry) according to population growth at low density (λ). Higher densities of occurrences are expected in locations where $\lambda > 1$. White circles show all known presence points. Both species have the potential for considerable northward spread under current conditions, but *A. petiolata* is projected to be less invasive, and *B. thunbergii* more invasive, under future climate.

Source: From Merow *et al.* (2016).

… and freshwater fish

The aquarium trade provides the backdrop for multiple invasions of natural habitats by freshwater fish. In North America, for example, 10% of households possess ornamental fish, providing heavy propagule pressure for escape, establishment and impacts on native species. Venezia *et al.* (2018) estimated the magnitude of invasion risk across the USA and Quebec (Canada) in models that included propagule pressure, species traits and environmental variables, both currently and under high emission scenario RCP8.5. While the overall average risk of establishment for the entire suite of aquarium species is projected to increase by 40% by 2050 across North America, future establishment risk in northern regions, like Quebec, remains very low. Establishment risks in southern regions, in contrast, show much more pronounced increases, with Florida projected to have the greatest overall risk increase. This reflects both the already preferable habitats in southern climates, but also the novel conditions of temperature and precipitation expected in tropical and subtropical regions.

APPLICATION 22.8 Invasive plants in protected areas

The future potential risk of plant invasions into the world's protected areas (PAs) seems also to be strongly linked to changes in temperature and precipitation. Wang *et al.* (2017) used a version of niche modelling to estimate suitable habitats for 386 invasive plant species (IPSs) with at least 50 current distribution records. But they also included in their estimate the propensity of each protected area to be invaded according to the likelihood of the species arriving there, using a technique (corridor analysis) that serves to identify connected or unconnected pairs of suitable patches across a global grid of patches with and without suitable habitat. The model projections of the ability of PAs to support these potential pathways for the suite of IPSs for a low emission scenario are shown in Figure 22.26 (results for a high emission scenario were very similar). The main potential pathways for invasive plant species into protected areas appear to be in Europe, eastern Australia, New Zealand, southern Africa and eastern South America, and managers in these protected areas will need to be particularly vigilant.

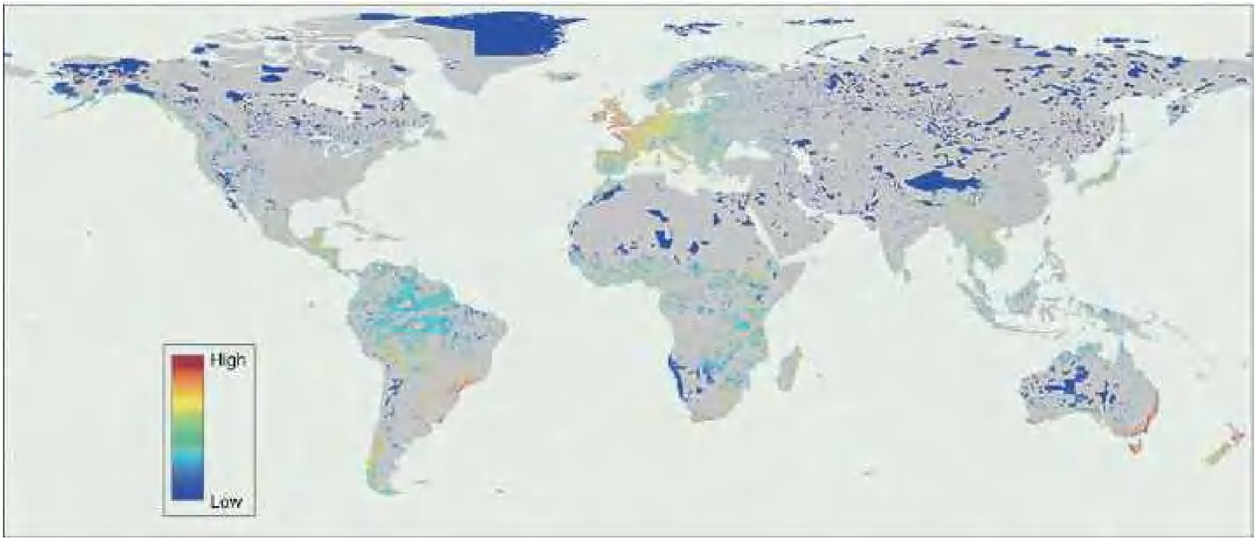


Figure 22.26 Global map of the propensity of the world's protected areas to be invaded by a suite of invasive plant species. This is according to future climate change (low emission scenario).

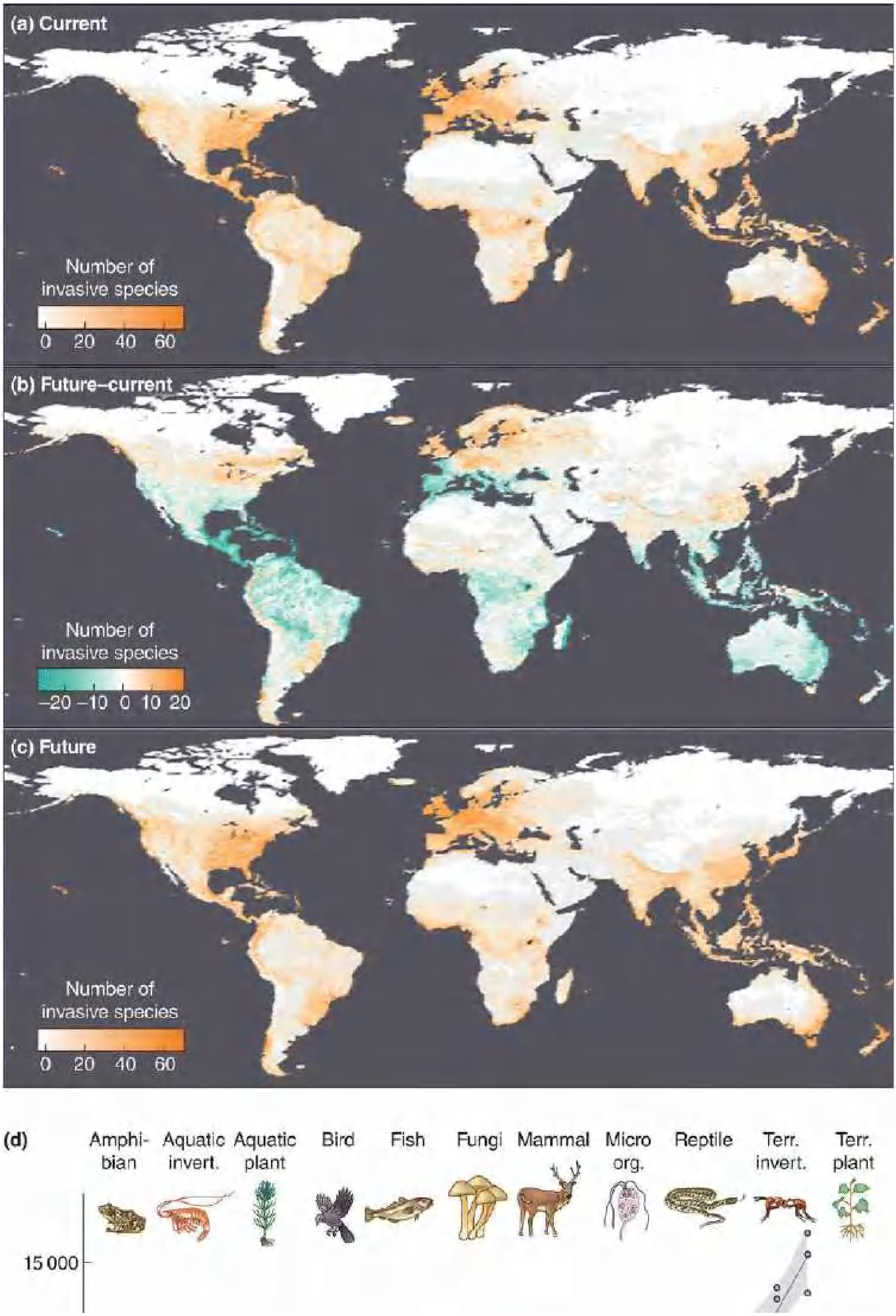
Source: From Wang *et al.* (2017).

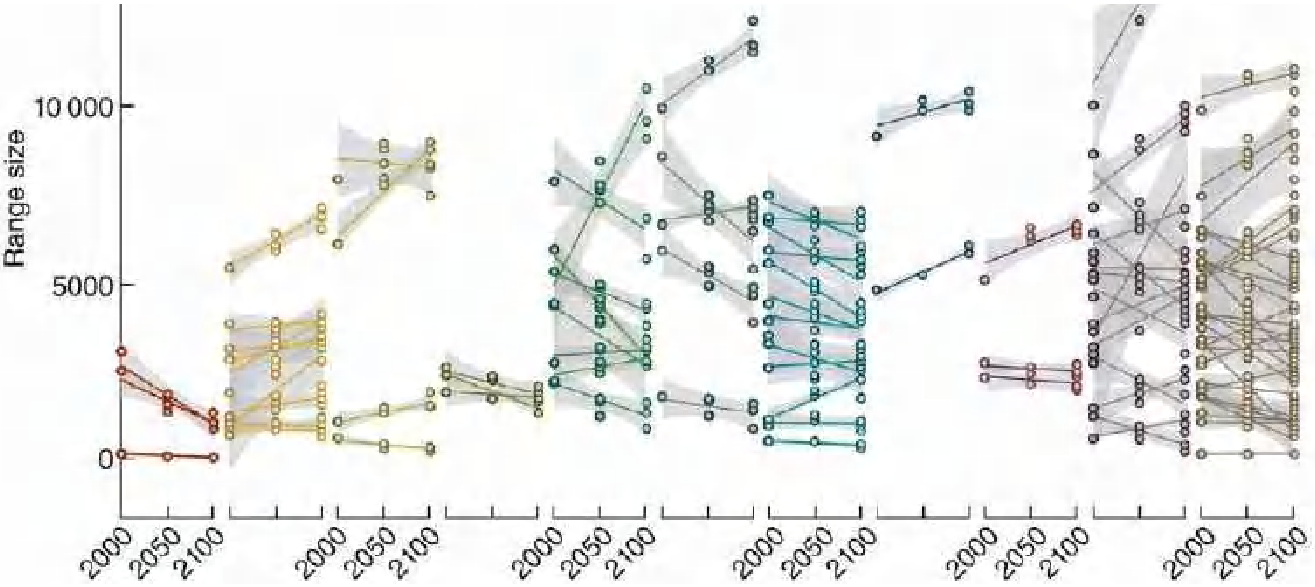
22.7.2 Climate change, land-use change and invasion risk

As well as changes to global patterns of temperature, climate change will also see more extreme climatic events such as floods, fires and cyclones, all of which have the potential to enhance invasion processes (introduction, establishment and spread). Until recently it has been non-climatic anthropogenic factors that have been the main driving force, especially transcontinental transportation, but also agricultural activities and other ecosystem disruption processes that provide pathways and prime locations for invaders to take hold. A pressing question, therefore, is how climate change and land-use change may interact in future.

To begin to address this knowledge gap, Bellard *et al.* (2013) used species distribution models to project future habitable areas of the world's ‘100 worst invasive species' as listed by the International Union for the Conservation of Nature (omitting one, the rinderpest virus, because it has been eradicated). The remaining 99 species include microorganisms, fungi, plants, invertebrates and vertebrates from terrestrial, freshwater and coastal marine environments. The authors' approach takes into account both climate change (using global climate models, and a lower and a higher emission scenario, similar to RCP6.0 and RCP4.5) and land-use change (using land cover change models that incorporate measures such as cultivated and managed areas, pasture, artificial surfaces, etc.).

Echoing several of the results shown in Figure 22.26, current (Figure 22.27a) and future (Figure 22.27c) hotspots of invasion are mostly in the eastern USA, Europe, south-west Australia and New Zealand (>60 invasive species), followed by Indonesian and Pacific island regions, central Africa and southern Brazil (20–40 invasive species). Increases in the number of invasive species from current to future conditions are projected for north-western Europe, north-eastern USA, India and China (Figure 22.27b). But note that a decrease in the number of invasive species is projected for Central and South America, south-western Europe, central Africa, Indonesian and Pacific islands regions and eastern Australia. In general, biomes subject to extreme climatic conditions (ice, hot desert, tundra) are not expected to be suitable for these invasive species by 2100. The highest increase in invasive species is expected for temperate and cool coniferous forest, while tropical forests and woodlands may become suitable for a lower number of the invaders in future. These patterns are essentially the same whether the lower or higher emission scenario is modelled.


(a) Current
Number of invasive species
0
20
40
60
(b) Future–current
Number of invasive species
−20
−10
0
10
20
(c) Future
Number of invasive species
0
20
40
60
(d)
Amphibian
Aquatic invert.
Aquatic plant
Bird
Fish
Fungi
Mammal
Micro org.
Reptile
Terr. invert.
Terr. plant
15 000



[Figure 22.27](#) Current and projected distributions of the world's most invasive species. (a) Global distribution of the world's 99 'most invasive species' predicted under current conditions. (b) Projected increases or decreases in the number of invasive species per grid cel in 2100 according to a future scenario of changed land-use and climate. (c) Projected richness of invasive species by 2100. (d) Projected changes in range size (number of suitable grid cells) for the 99 species among the different taxonomic groups. Results shown are for the higher emission scenario, but the pattern of results was the same whether the lower or higher emission scenario was modelled.

Source: From Bellard *et al.* (2013).

Turning to the projected fate of the 99 individual invasive species, a small increase overall in average potential range size is projected by 2100 (2–6%), but it is the considerable variation among taxa that is of particular interest (Figure 22.27d). Thus, among these 'most invasive species', the fungi (−11%), and particularly amphibians (−65%) and birds (−24%) are projected to experience strong range size shrinkage, freshwater fishes (−1%), mammals (−4%) and reptiles (−4%) look set to remain stable, while aquatic invertebrates (+59%) and terrestrial invertebrates (+17%), aquatic plants (+12%) and microorganisms (+17%) are predicted to generally expand their potential range distributions.

The finding that habitat suitability may decline for some taxa in some regions provide some grounds to be hopeful, if it means that current invaders are likely to decline or become extinct, making management easier and perhaps providing opportunities for restoration.

22.8 Planetary boundaries

relative importance of different human drivers of biodiversity loss

The Millennium Ecosystem Assessment (MEA, 2005b) brought together a panel of experts from around the world to evaluate and compare the principal drivers of biodiversity loss now and in future, and their conclusions for a variety of biomes are illustrated in Figure 22.28. They predicted that climate change and N and P pollution will become progressively more important causes of biodiversity loss across all ecosystem types, habitat change in most types, and invasive species and overexploitation in a smaller number of biomes. Of course, these are gross generalisations. For amphibians in Central America, for instance, habitat loss poses the greatest risk, followed by pollution and then introduced pathogens (Whitfield *et al.*, 2016), but for freshwater fish worldwide, it is invasive species that pose the biggest threat, followed by climate change, while overexploitation poses the biggest risk for marine fish (Arthington *et al.*, 2016). Nevertheless, the general picture in Figure 22.28 is useful background for the planning and prioritisation by ecosystem managers around the globe.

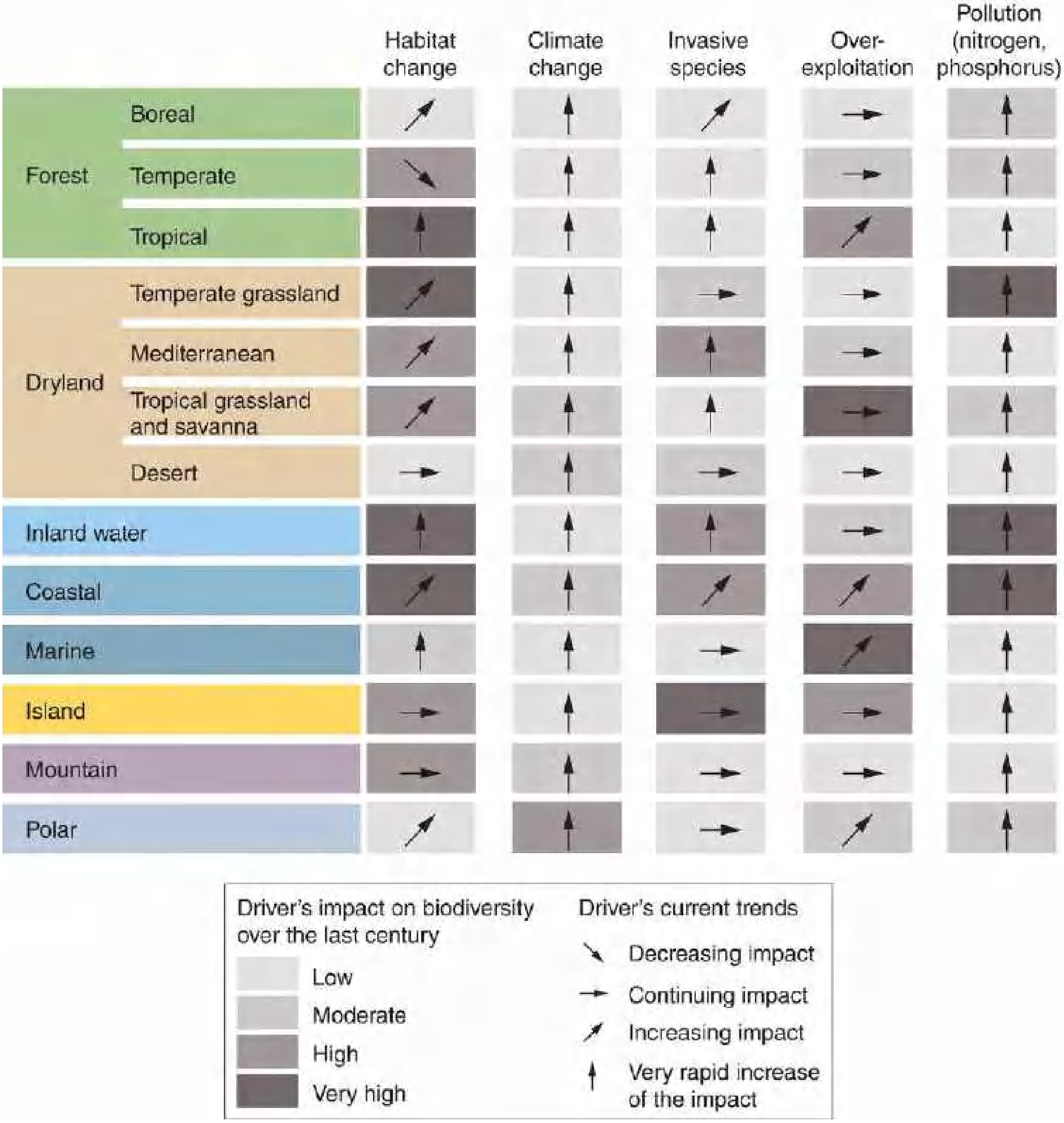


Figure 22.28 Principal drivers of biodiversity change in various terrestrial and aquatic biomes. Shading expresses expert panel opinion on the impact of drivers over the past century, while arrows indicate the predicted future trends.

Source: After MEA (2005b).

the planetary boundary approach

Taking a broader view, the planetary boundary approach is designed to provide a science-based analysis of the risk that human perturbations will destabilise the whole 'earth system', via changes to biochemical flows, stratospheric ozone, climate, land use, water use, atmospheric aerosol loading, as well as biosphere integrity. The rationale is that the relatively stable, 11 700 year-long Holocene epoch (see Figure 22.1) is the only state of the earth system known for certain to be able to support contemporary human societies. But now, in the Anthropocene epoch, human activities have reached levels that may damage the systems that keep the earth in a ‘desirable' state. The challenge then is to provide a framework based on planetary boundaries to maintain a Holocene state, or revert to it. For each planetary subsystem, Steffen *et al.* (2015) chose a relevant control variable and a threshold that, if crossed, could shift the system into a new state with deleterious consequences for humans. The boundary is then set at a safe distance from the threshold and a zone of uncertainty is also provided. These choices can be considered rather arbitrary, but they are based on discussions involving a number of leading earth-system and environmental scientists.

control variables and boundaries

For climate change, for example, the selected control variable is CO_2 concentration, the boundary is 350 ppm and the zone of uncertainty 350–450 ppm. We are currently above 400 ppm (Table 22.1). Anthropogenic perturbations to biosphere integrity, biochemical flows of P and N, and land-system change are already above the proposed boundaries (Figure 22.29). Some of the critical earth systems are yet to be quantified (aerosol loadings, novel entities, functional diversity of the biota) and increased sophistication will be required before many of them can be operationalised (particularly with respect to accounting for regional-level heterogeneity; Steffen *et al.*, 2015).

Table 22.1 Critical earth-system processes, their control variables, proposed planetary boundaries and current values. *Source:* From Steffen *et al.* (2015).

Earth-system process	Control variable	Planetary boundary (and zone of uncertainty)	Current value
Climate change	Atmospheric CO_2 concentration (ppm)	350 (350–450)	415[a]
Biosphere integrity	Extinction rate per million species years (E MSY^{-1})	<10 (10–100)	100-1000
Stratospheric ozone	Reduction from preindustrial level (290 Dobson units)	280 (261–280)	200[b]
Ocean acidification	% of preindustrial average carbonate saturation (aragonite)	≤80% (≤80%–≤70%)	84%
Biogeochemical flows P	P global from freshwater into ocean (Tg yr^{-1})	11 (11–100)	22
Biogeochemical flows N	N global (industrial and intentional biological fixation) (Tg yr^{-1})	62 (62–82)	150
Land-system change	Global as % of original forested land	75 (75–54)	62
Freshwater use	Max consumption[c] (from rivers, lakes, reservoirs and groundwater) (km^3 yr^{-1})	4000 (4000–6000)	2600
Atmospheric aerosol loading[d]	Aerosol optical depth, but much regional variation	Not yet implemented	-
Introduction of novel entities[e]	No control variable currently defined	Not yet implemented	-

[a] The ‘current value' for CO_2 concentration has been updated to June 2020.

[b] The value shown is for Antarctic in the Austral spring, but nowhere else has the boundary been transgressed.

[c] Based on the idea that freshwater systems need sufficient environmental water flows to avoid regime shifts.

[d] Emissions of black carbon, sulphates and nitrates from fossil fuel use and cooking and heating with biofuels. The control variable of aerosol optical depth is being considered, but is so variable regionally that a global boundary may not be implementable.

[e] New chemicals, new forms of existing substances and modified life forms that have the potential for unwanted geophysical and/or biological effects.

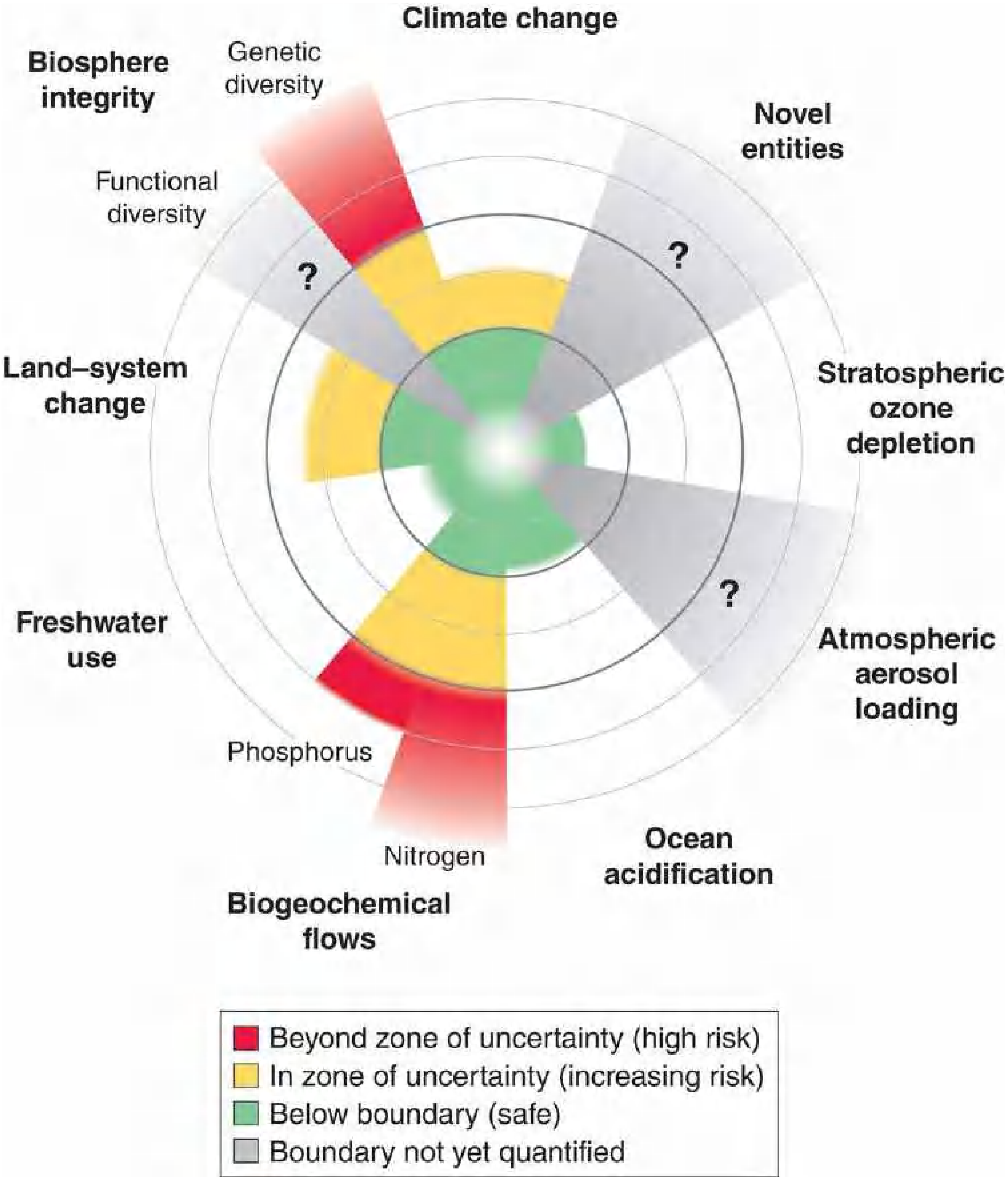


Figure 22.29 The current state of control variables for critical earth systems in relation to planetary boundaries. The authors use 'global extinction rate' as their measure of 'genetic diversity' (Table 22.1).

Source: From Steffen *et al.* (2015).

The planetary boundary framework should make a useful contribution for decision makers when charting desirable courses for societal development and environmental management in a changing world.

22.9 Finale

When we wrote the final paragraph of the first edition of this book in the mid-1980s, we concluded with a call to arms that was purely scientific: ‘Clear, unambiguous predictions and tests of *ideas* are often very difficult to devise and will require great ingenuity on the part of future generations of ecologists.’ (Italics added.) Thirty-five years on, that closing sentence remains true, but now our exhortation must be both more applied and more urgent. We have chosen to use this final chapter to emphasise the changing world in which we live and the threats those changes pose. Throughout the book, we have seen that ecologists are continually taking strides forward in understanding and, sometimes, in suggesting solutions or at least mitigation strategies, and we have tried to acknowledge studies that give us grounds for hope for the future. But the gaps in our understanding remain profound, the solutions that can be suggested do not come with guarantees, and even where we can offer grounds for hope, we must acknowledge the political practicalities and the range of social and economic drivers that, quite rightly, play their own role in any move towards policy change.

So, from now, future generations of ecologists will need more than great ingenuity in the testing of ideas. They – you – will need also to decide what the role of ecologists in policy formulation should be, and to do so in an intellectual climate where it has become fashionable to question the usefulness of ‘experts’ – yet another climate change for us to deplore. It would be quite wrong for ecologists to assume that society should simply do as we say, and therefore equally wrong for us to become frustrated and resentful whenever we appear to be ignored. But it may be wrong, too, for ecologists to carry on seeing themselves only as idealistic upholders of the ‘Enlightenment' tradition, explaining patiently the evidence and its interpretation and proceeding with caution, and imagining that eventually everyone will see the light. If, as Lakoff (2009) maintains, arguments in the public sphere are not won by those with the best evidence, but by those who appeal most effectively to the emotions, then to be truly effective in the public sphere, ecologists will need to learn to devise clear and persuasive messages (Begon, 2017) and frame our arguments, and repeat them time and time again, in ‘metaphors’ that people understand and relate to. These are often very difficult to devise but are desperately needed. This will require great ingenuity on the part of future generations of ecologists.

References

Abrams, P. (1976) Limiting similarity and the form of the competition coefficient. *Theoretical Population Biology*, 8, 356–375.

Abrams, P. (1983) The theory of limiting similarity. *Annual Review of Ecology and Systematics*, 14, 359–376.

Abrams, P.A. (2015) Why ratio dependence is (still) a bad model of predation. *Biological Reviews*, 90, 794–814.

Abramsky, Z. & Rosenzweig, M.L. (1983) Tilman's predicted productivity–diversity relationship shown by desert rodents. *Nature*, 309, 150–151.

Abrantes, K.G. & Sheaves, M. (2010) Importance of freshwater flow in terrestrial-aquatic energetic connectivity in intermittently connected estuaries of tropical Australia. *Marine Biology*, 157, 2017–2086.

Adams, E.S. (2001) Approaches to the study of territory size and shape. *Annual Review of Ecology and Systematics*, 32, 277–303.

Adem Esmail, B. & Geneletti, D. (2018) Multi-criteria decision analysis for nature conservation: a review of 20 years of applications. *Methods in Ecology and Evolution*, 9, 42–53.

Agrawal, A.A. (1998) Induced responses to herbivory and increased plant performance. *Science*, 279, 1201–1202.

Ainsworth, E.A. & Long, S.P. (2005) What have we learned from 15 years of free-air CO_2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO_2. *New Phytologist*, 164, 351–372.

Akçakaya, H.R. (1992) Population viability analysis and risk assessment. In: *Proceedings of Wildlife 2001: Populations* (D.R. McCullough, ed.), pp. 148–158. Elsevier, Amsterdam.

Akçakaya, H. & Root, W. (2013) *Ramas GIS: Linking Spatial Data with Population Viability Analysis (version 6.0)*. Applied Biomathematics, Setauket, New York.

Akman, G.E. & Douglas, A.E. (2009) Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proceedings of the Royal Society of London, Series B*, 276, 987–991.

Alava, J.J., Cheung, W.W.L., Ross, P.S. & Sumaila, U.R. (2017) Climate change–contaminant interactions in marine food webs: toward a conceptual framework. *Global Change Biology*, 23, 3984–4001.

Alberdi, A., Aizpurua, O., Bohmann, K., Zepeda-Mendoza, M. & Gilbert, M.T.P. (2016) Do vertebrate gut metagenomes confer rapid ecological adaptation? *Trends in Ecology and Evolution*, 31, 689–699.

Allan, B.F., Keesing, F. & Ostfeld, R.S. (2003) Effect of forest fragmentation of Lyme disease risk. *Conservation Biology*, 17, 267–272.

Allen, K.R. (1972) Further notes on the assessment of Antarctic fin whale stocks. *Report of the International Whaling Commission*, 22, 43–53.

Allen, L.J.S. & Allen, E.J. (2003) A comparison of three different stochastic population models with regard to persistence time. *Theoretical Population Biology*, 64, 439–449.

Alliende, M.C. & Harper, J.L. (1989) Demographic studies of a dioecious tree. I. Colonization, sex and age-structure of a population of *Salix cinerea*. *Journal of Ecology*, 77, 1029–1047.

Allison, M.J. (1984) Microbiology of the rumen and small and large intestines. In: *Dukes' Physiology of Domestic Animals*, 10th edn (M.J. Swenson, ed.), pp. 340–350. Cornell University Press, Ithaca, NY.

Alphey, T.W. (1970) Studies on the distribution and site location of *Nippostrongylus brasiliensis* within the small intestine of laboratory rats. *Parasitology*, 61, 449–460.

Amendt, J., Krettek, R. & Zehner, R. (2004) Forensic entomology. *Naturwissenschaften*, 91, 51–65.

Anav, A., Friedlingstein. P., Beer, C. *et al.* (2015) Spatiotemporal patterns of terrestrial gross primary production: a review. *Reviews of Geophysics*, 53, 785–818.

Anderson, J.M. (1978) Inter- and intrahabitat relationships between woodland Cryptostigmata species diversity and diversity of soil and litter micro-habitats. *Oecologia*, 32, 341–348.

Anderson, R.M. (1982) Epidemiology. In: *Modern Parasitology* (F.E.G. Cox, ed.), pp. 205–251. Blackwell Scientific Publications, Oxford.

Anderson, R.M. (1991) Populations and infectious diseases: ecology or epidemiology? *Journal of Animal Ecology*, 60, 1–50.

Anderson, R.M. & May, R.M. (1978) Regulation and stability of host– parasite population interactions. I. Regulatory processes. *Journal of Animal Ecology*, 47, 219–249.

Anderson, R.M. & May, R.M. (1991) *Infectious Diseases of Humans: dynamics and control*. Oxford University Press, Oxford.

Andreassen, H.P., Glorvigen, P., Rémy, A. & Ims, R.A. (2013) New views on how population-intrinsic and community-extrinsic processes interact during the vole population cycles. *Oikos*, 122, 507–515.

Andrewartha, H.G. & Birch, L.C. (1954) *The Distribution and Abundance of Animals*. University of Chicago Press, Chicago.

Antonovics, J. (2006) Evolution in closely adjacent plant populations X: long-term persistence of prereproductive isolation at a mine boundary. *Heredity*, 97, 33–37.

Arain, M.A., Black, T.A., Barr, A.G. *et al.* (2002) Effects of seasonal and interannual climate variability on net ecosystem productivity of boreal deciduous and conifer forests. *Canadian Journal of Forestry Research*, 32, 878–891.

Archer, M. (2014) Comparative analysis of insect succession data from Victoria (Australia) using summary statistics versus preceding mean ambient temperature models. *Journal of Forensic Science*, 59, 404–412.

Arditi, R. & Ginzburg, L.R. (1989) Coupling in predator–prey dynamics: ratio dependence. *Journal of Theoretical Biology*, 139, 311–326.

Arditi, R. & Ginzburg, L.R. (2012) *How Species Interact*. Oxford University Press, New York.

Arheimer, B., Dahne, J. & Donnelly, C. (2012) Climate change impact on riverine nutrient load and land-based measures of the Baltic Sea Action Plan. *Ambio*, 41, 600–612.

Armstrong, R.A. & McGehee, R. (1980) Competitive exclusion. *American Naturalist*, 115, 151–170.

Arora, A.K. & Douglas, A.E. (2017) Hype or opportunity? Using microbial symbionts in novel strategies for insect pest control. *Journal of Insect Physiology*, 103, 10–17.

Arsenault, R. & Owen-Smith, N. (2002) Facilitation versus competition in grazing herbivore assemblages. *Oikos*, 97, 313–318.

Arthington, A.H., Dulvy, N.K., Gladstone, W. & Winfield, I.J. (2016) Fish conservation in freshwater and marine realms: status, threats and management. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 26, 838–857.

Ashmole, N.P. (1971) Sea bird ecology and the marine environment. In: *Avian Biology*, Vol. 1 (D.S. Farner & J.R. King, eds), pp. 224–286. Academic Press, New York.

Asiimwe, P., Naranjo, S.E. & Ellsworth, P.C. (2014) Effects of irrigation levels on interactions among *Lygus hesperus* (Hemiptera: Miridae), insecticides, and predators in cotton. *Environmental Entomology*, 43, 263–273.

Astrup, R., Bernier, P.Y., Genet, H., Lutz, D.A. & Bright, R.M. (2018) A sensible climate solution for the boreal forest. *Nature Climate Change*, 8, 2–12.

Atkinson, D., Ciotti, B.J. & Montagnes, D.J.S. (2003) Protists decrease in size linearly with temperature: *ca.* 2.5% °C–1. *Proceedings of the Royal Society of London, Series B*, 270, 2605–2611.

Atkinson, W.D. & Shorrocks, B. (1981) Competition on a divided and ephemeral resource: a simulation model. *Journal of Animal Ecology*, 50, 461–471.

Azcárate, F.M. & Peco, B. (2006) Effects of seed predation by ants on Mediterranean grassland related to seed size. *Journal of Vegetation Science*, 17, 353–360.

Babikova, Z., Gilbert, L., Bruce, T.J.A. *et al.* (2014) Underground signals carried through common mycelial networks warn neighbouring plants of aphid attack. *Ecology Letters*, 16, 835–843.

Bailey, A., Chandler, D., Grant, W.P., Greaves, J., Prince, G. & Tatchell, M. (2010) *Biopesticides: pest management and regulation*. CABI, Wallingford.

Bajwa, A.A., Chauhan, B.S., Farooq, M., Shabbir, A. & Adkins, S.W. (2016) What do we really know about alien plant invasion? A review of the invasion mechanism of one of the world's worst weeds. *Planta*, 244, 39–57.

Bale, J.S. 2002 Insects and low temperatures: from molecular biology to distributions and abundance. *Philosophical Transactions of the Royal Society of London B*, 357, 849–862.

Ballaré, C.L. (2010) Jasmonate-induced defenses: a tale of intelligence, collaborators and rascals. *Trends in Plant Science*, 16, 249–257.

Balmford, A. & Bond, W. (2005) Trends in the state of nature and their implications for human well-being. *Ecology Letters*, 8, 1218–1234.

Balmford, A., Green, R.E. & Jenkins, M. (2003) Measuring the changing state of nature. *Trends in Ecology and Evolution*, 18, 326–330.

Baltensweiler, W., Benz, G., Bovey, P. & Delucchi, V. (1977) Dynamics of larch budmoth populations. *Annual Review of Ecology and Systematics*, 22, 79–100.

Banavar, J.R., Moses, M.E., Sibly, R.M. & Maritan, A. (2010) A general basis for quarter-power scaling in animals. *Proceedings of the National Academy of Sciences of the USA*, 107, 15816–15820.

Barluenga, M., Stolting, K.N., Salzburger, W., Muschick, M. & Meyer, A. (2006) Sympatric speciation in Nicaraguan crater lake cichlids fish. *Nature*, 439, 719–723.

Barner, A.K., Lubchenco, J., Costello, C. *et al.* (2015) Solutions for recovering and sustaining the bounty of the ocean: Combining fishery reforms, rights-based fisheries management, and marine reserves. *Oceanography*, 28, 252–263.

Barnes, M.D., Craigie, I.D., Harrison, L.B. *et al.* (2016) Wildlife population trends in protected areas predicted by national socio-economic metrics and body size. *Nature Communications*, 7, 12747.

Barnett, M.A. (1983) Species structure and temporal stability of mesopelagic fish assemblages in the Central Gyres of the North and South Pacific Ocean. *Marine Biology*, 74, 245–256.

Baross, J.A. & Deming, J.W. (1995) Growth at high temperatures: isolation and taxonomy, physiology, ecology. In: *Microbiology of Deep-Sea Hydrothermal Vent Habitats* (D.M. Karl, ed.), pp. 169–218. CRC Press, New York.

Barrett, K., Wait, D. & Anderson, W.B. (2003) Small island biogeography in the Gulf of California: lizards, the subsidized island biogeography hypothesis, and the small island effect. *Journal of Biogeography*, 30, 1575–1581.

Barrio, I.C., Hik, D.S., Bueno, C.G. & Cahill, J.F. (2013) Extending the stress-gradient hypothesis – is competition among animals less common in harsh environments? *Oikos*, 122, 516–523.

Bascompte, J. & Sole, R.V. (1996) Habitat fragmentation and extinction thresholds in spatially explicit models. *Journal of Animal Ecology*, 65, 465–473.

Bashey, F. (2008) Competition as a selective mechanism for larger offspring size in guppies. *Oikos*, 117, 104–113.

Bauerfeind, S.S., Theisen, A. & Fischer, K. (2009) Patch occupancy in the endangered butterfly *Lycaena helle* in a fragmented landscape: effects of habitat quality, patch size and isolation. *Journal of Insect Conservation*, 13, 271–277.

Baumann, P. (2005) Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual Review of Microbiology*, 59, 155–189.

Bayliss, P. (1987) Kangaroo dynamics. In: *Kangaroos, their Ecology and Management in the Sheep Rangelands of Australia* (G. Caughley, N. Shepherd & J. Short, eds), pp. 119–134. Cambridge University Press, Cambridge, UK.

Bazelet, C.S., Thompson, A.C. & Naskrecki, P. (2016) Testing the efficacy of global biodiversity hotspots for insect conservation: the case of South African katydids. *PLoS ONE*, 11, e0160630.

Bazzaz, F.A. & Williams, W.E. (1991) Atmospheric CO2 concentrations within a mixed forest: implications for seedling growth. *Ecology*, 72, 12–16.

Beaumont, L.J. & Hughes, L. (2002) Potential changes in the distributions of latitudinally restricted Australian butterfly species in response to climate change. *Global Change Biology*, 8, 954–971.

Beck, J., McCain, C.M., Axmacher, J.C. *et al.* (2017) Elevational species richness gradients in a hyperdiverse insect taxon: a global meta-study on geometrid moths. *Global Ecology and Biogeography*, 26, 412–424.

Becker, P. (1992) Colonization of islands by carnivorous and herbivorous Heteroptera and Coleoptera: effects of island area, plant species richness, and 'extinction' rates. *Journal of Biogeography*, 19, 163–171.

Becker, P. (2000) Competition in the regeneration niche between conifers and angiosperms: Bond's slow seedling hypothesis. *Functional Ecology*, 14, 401–412.

Bedford, H. & Elliman, D. (2018) Measles: neither gone nor forgotten. *British Medical Journal*, 362, k3976.

Beebee, T.J.C. (1991) Purification of an agent causing growth inhibition in anuran larvae and its identification as a unicellular unpigmented alga. *Canadian Journal of Zoology*, 69, 2146–2153.

Begon, M. (2017) Winning public arguments as ecologists: time for a new doctrine? *Trends in Ecology and Evolution*, 32, 394–396.

Begon, M., Bennett, M., Bowers, R.G., French, N.P., Hazel, S.M. & Turner, J. (2002) A clarification of transmission terms in host – microparasite models: numbers, densities and areas. *Epidemiology and Infection*, 129, 147–153.

Begon, M., Feore, S.M., Bown, K., Chantrey, J., Jones, T. & Bennett, M. (1998) Population and transmission dynamics of cowpox virus in bank voles: testing fundamental assumptions. *Ecology Letters*, 1, 82–86.

Begon, M., Firbank, L. & Wall, R. (1986) Is there a self-thinning rule for animal populations? *Oikos*, 46, 122–124.

Behmer, S.T. (2009) Insect herbivore nutrient regulation. *Annual Review of Entomology*, 54, 165–187.

Bekaert, M., Edger, P.P., Hudson, C.M., Pires, J.C. & Conant, G.C. (2012) Metabolic and evolutionary costs of herbivory defense: systems biology of glucosinolate synthesis. *New Phytologist*, 196, 596–605.

Bekessy, S.A., Runge, M.C., Kusmanoff, A.M., Keith, D.A. & Wintle, B.A. (2018) Ask not what nature can do for you: a critique of ecosystem services as a communication strategy. *Biological Conservation*, 224, 71–74.

Belassen, V. & Luyssaert, S. (2014) Managing forests in uncertain times. *Nature*, 506, 153–155.

Beldomenico, P.M. & Begon, M. (2010) Disease spread, susceptibility and infection intensity: vicious circles? *Trends in Ecology and Evolution*, 25, 21–27.

Bell, D.T. (2001) Ecological response syndromes in the flora of southwestern Western Austalia: fire resprouters versus reseeders. *The Botanical Review*, 67, 417–441.

Bellard, C., Thuiller, W., Leroy, B., Genovesi, P., Bakkenes, M. & Courchamp F. (2013) Will climate change promote future invasions? *Global Change Biology*, 19, 3740–3748.

Bellows, T.S. Jr (1981) The descriptive properties of some models for density dependence. *Journal of Animal Ecology*, 50, 139–156.

Belmaker, J. & Jetz, W. (2011) Cross-scale variation in species richness – environment associations. *Global Ecology and Biogeography*, 20, 464–474.

Belt, T. (1874) *The Naturalist in Nicaragua*. J.M. Dent, London.

Belyea, L.R. & Lancaster, J. (1999) Assembly rules within a contingent ecology. *Oikos*, 86, 402–416.

Ben-Hamo, M., Pinshow, B., McCue, M.D., McWilliams, S.R. & Bauchinger, U. (2010) Fasting triggers hypothermia, and ambient temperature modulates its depth in Japanese quail

Coturnix japonica. *Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology*, 156, 84–91.

Benot, M.-L., Bittebiere A.-K., Ernoult, A., Clement, B. & Mony, C. (2013) Fine-scale spatial patterns in grassland communities depend on species clonal dispersal ability and interactions with neighbours. *Journal of Ecology*, 101, 626–636.

Benson, E.J. & Hartnett, D.C. (2006) The role of seed and vegetative reproduction in plant recruitment and demography in tall-grass prairie. *Plant Ecology*, 187, 163–178.

Benson, J.F. (1973) The biology of Lepidoptera infesting stored products, with special reference to population dynamics. *Biological Reviews*, 48, 1–26.

Bergelson, J.M. (1985) A mechanistic interpretation of prey selection by *Anax junius* larvae (Odonata: Aeschnidae). *Ecology*, 66, 1699–1705.

Berger, J. (1990) Persistence of different-sized populations: an empirical assessment of rapid extinctions in bighorn sheep. *Conservation Biology*, 4, 91–98.

Bergman, C.M., Fryxell, J.M. & Gates, C.C. (2000) The effect of tissue complexity and sward height on the functional response of wood bison. *Functional Ecology*, 14, 61–69.

Bernasconi, G., Antonovics, J., Biere, A. *et al.* (2009) Silene as a model system in ecology and evolution. *Heredity*, 103, 5–14.

Berner, E.K. & Berner, R.A. (1987) *The Global Water Cycle: geochemistry and environment*. Prentice-Hall, Englewood Cliffs, NJ.

Bernstein, C., Kacelnik, A. & Krebs, J.R. (1988) Individual decisions and the distribution of predators in a patchy environment. *Journal of Animal Ecology*, 57, 1007–1026.

Bernstein, C., Kacelnik, A. & Krebs, J.R. (1991) Individual decisions and the distribution of predators in a patchy environment. II. The influence of travel costs and structure of the environment. *Journal of Animal Ecology*, 60, 205–225.

Berntson, G.M. & Wayne, P.M. (2000) Characterizing the size dependence of resource acquisition within crowded plant populations. *Ecology*, 8, 1072–1085.

Berry, J.A. & Björkman, O. (1980) Photosynthetic response and adaptation to temperature in higher plants. *Annual Review of Plant Physiology*, 31, 491–543.

Berryman, A.A. (ed.) (2002) *Population Cycles: the case for trophic interactions*. Oxford University Press, Oxford.

Berryman, A. & Turchin, P. (2001) Identifying the density-dependent structure underlying ecological time series. *Oikos*, 92, 265–270.

Bertness, M.D. & Callaway, R. (1994) Positive interactions in communities. *Trends in Ecology and Evolution*, 9, 191–193.

Bertness, M.D., Leonard, G.H., Levin, J.M., Schmidt, P.R. & Ingraham, A.O. (1999) Testing the relative contribution of positive and negative interactions in rocky intertidal communities. *Ecology*, 80, 2711–2726.

Bertuzzo, E., Carrara, F., Mari, L., Altermatt, F., Rodriguez-Iturbe, I. & Rinaldo, A. (2016) Geomorphic controls on elevational gradients of species richness. *Proceedings of the National Academy of Sciences of the USA*, 113, 1737–1742.

Bertzky, B., Corrigan, C., Kemsey, J. *et al.* (2012) *Protected Planet Report 2012: Tracking progress towards global targets for protected areas*. IUCN, Gland, Switzerland and UNEP-WCMC, Cambridge, UK.

Berven, K.A. (1995) Population regulation in the wood frog, *Rana sylvatica*, from three diverse geographic localities. *Australian Journal of Ecology*, 20, 385–392.

Beven, G. (1976) Changes in breeding bird populations of an oak-wood on Bookham Common, Surrey, over twenty-seven years. *London Naturalist*, 55, 23–42.

Beynon, S.A., Wainwright, W.A. & Christie, M. (2015) The application of an ecosystem services framework to estimate the economic value of dung beetles to the U.K. cattle industry. *Ecological Entomology*, 40 (Suppl. 1), 124–135.

Bezemer, T.M. & Mills, N.J. (2001) Host density responses of *Mastrus ridibundus*, a parasitoid of the codling moth, *Cydia pomonella*. *Biological Control*, 22, 169–175.

Bielby, J., Mace, G.M., Bininda-Emonds, O.R.P. *et al.* (2007) The fast-slow continuum in mammalian life history: an empirical reevaluation. *American Naturalist*, 169, 748–757.

Biological Records Centre (2017) Heracleum mantegazzianum. http://www.brc.ac.uk/plantatlas/plant/heracleum-mantegazzianum (last accessed November 2019).

Bird, C.E., Fernandez-Silva, I., Skillings, D.J. & Toonen, R.J. (2012) Sympatric speciation in the post ‘modern synthesis’ era of evolutionary biology. *Evolutionary Biology*, 39, 158–180.

Bishop, K.A., Leakey, A.D.B. & Ainsworth, E.A. (2014) How seasonal temperature or water inputs affect the relative response of C_3 crops to elevated [CO_2]: a global analysis of open top chamber and free air CO_2 enrichment studies. *Food and Energy Security*, 3, 33–45.

Bitume, E.V., Bonte, D., Ronce, O. *et al.* (2013) Density and genetic relatedness increase dispersal distance in a subsocial organism. *Ecology Letters*, 16, 430–437.

Björnhag, G. (1994) Adaptations in the large intestine allowing small animals to eat fibrous food. In: *The Digestive System in Mammals. Food, Form and Function* (D.J. Chivers & P. Langer, eds), pp. 287–312. Cambridge University Press, Cambridge, UK.

Bjørnstad, O.N. (2015) Nonlinearity and chaos in ecological dynamics revisited. *Proceedings of the National Academy of Sciences of the USA*, 112, 6252–6253.

Bjørnstad, O.N., Falck, W. & Stenseth, N.C. (1995) A geographic gradient in small rodent density fluctuations: a statistical modelling approach. *Proceedings of the Royal Society of London, Series B*, 262, 127–133.

Bjørnstad, O.N. & Grenfell, B.T. (2001) Noisy clockwork: time series analysis of population fluctuations in animals. *Science*, 293, 638–643.

Bjørnstad, O.N., Sait, S.M., Stenseth, N.C., Thompson, D.J. & Begon, M. (2001) The impact of specialized enemies on the dimensionality of host dynamics. *Nature*, 409, 1001–1006.

Black, A.J. & McKane, A.J. (2012) Stochastic formulation of ecological models and their applications. *Trends in Ecology and Evolution*, 27, 337–345.

Black, J.N. (1963) The interrelationship of solar radiation and leaf area index in determining the rate of dry matter production of swards of subterranean clover (*Trifolium subterraneum*). *Australian Journal of Agricultural Research*, 14, 20–38.

Blackwell, M. (2011) The fungi: 1, 2, 3...5.1 million species? *American Journal of Botany*, 98, 426–438.

Blaustein, A.R., Wake, D.B. & Sousa, W.P. (1994) Amphibian declines: judging stability, persistence, and susceptibility of populations to local and global extinctions. *Conservation Biology*, 8, 60–71.

Blomqvist, D., Pauliny, A., Larsson, M. & Flodin, L-A. (2010) Trapped in the extinction vortex? Strong genetic effects in a declining vertebrate population. *BMC Evolutionary Biology*, 10, 33.

Bloom, A.J. (2015) The increasing importance of distinguishing among plant nitrogen sources. *Current Opinion in Plant Biology*, 25, 10–16.

Bloom, A.J., Burger, M., Kimball, B.A. & Pinter, P.J. Jr (2014) Nitrate assimilation is inhibited by elevated CO_2 in field-grown wheat. *Nature Climate Change*, 4, 477–480.

Blouin, M., Hodson, M.E., Delgado, E.A. *et al.* (2013) A review of earthworm impact on soil function and ecosystem services*European Journal of Soil Science*, 64, 161–182.

Blueweiss, L., Fox, H., Kudzma, V., Nakashima, D., Peters, R. & Sams, S. (1978) Relationships between body size and some life history parameters. *Oecologia*, 37, 257–272.

Boddy, L. (1999) Saprotrophic cord-forming fungi: meeting the challenge of heterogeneous environments. *Mycologia*, 91, 13–32.

Boivin, G., Fauvergue, X. & Wajnberg, E. (2004) Optimal patch residence time in egg parasitoids: innate versus learned estimate of patch quality. *Oecologia*, 138, 640–647.

Boit, A., Sakschewski, B., Boysen, L. *et al.* (2016) Large-scale impact of climate change vs. land-use change on future biome shifts in Latin America. *Global Change Biology*, 22, 3689–3701.

Bonanomi, G., Incerti, G. & Mazzoleni, S. (2011) Assessing occurrence, specificity, and mechanisms of plant facilitation in terrestrial ecosystems. *Plant Ecology*, 212, 1777–1790.

Bond-Lamberty, B. & Thomson, A. (2010) A global database of soil respiration data. *Biogeosciences*, 7, 1915–1926.

Bonnet, X., Lourdais, O., Shine, R. & Naulleau, G. (2002) Reproduction in a typical capital breeder: costs, currencies and complications in the aspic viper. *Ecology*, 83, 2124–2135.

Bonsall, M.B., French, D.R. & Hassell, M.P. (2002) Metapopulation structure affects persistence of predator–prey interactions. *Journal of Animal Ecology*, 71, 1075–1084.

Boonstra, R. (2013) Reality as the leading cause of stress: rethinking the impact of chronic stress in nature. *Functional Ecology*, 27, 11–23.

Booth, B.D. & Swanton, C.J. (2002) Assembly theory applied to weed communities. *Weed Science*, 50, 2–13.

Boots, M. & Begon, M. (1993) Trade-offs with resistance to a granulosis virus in the Indian meal moth, examined by a laboratory evolution experiment. *Functional Ecology*, 7, 528–534.

Bordeau, P.E. (2012) Intraspecific trait cospecialization of constitutive and inducible morphological defences in a marine snail from habitats with different predation risk. *Journal of Animal Ecology*, 81, 849–858.

Borer, E.T., Halpern, B.S. & Seabloom, E.W. (2006) Asymmetry in community regulation: effects of predators and productivity. *Ecology*, 87, 2813–2820.

Börger, L. (2016) Stuck in motion? Reconnecting questions and tools in movement ecology. *Journal of Animal Ecology*, 85, 5–10.

Borland, A.M., Griffiths, H., Hartwell, J. & Smith, J.A.C. (2009) Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. *Journal of Experimental Botany*, 60, 2879–2896.

Borland, A.M., Hartwell, J., Weston, D.J. *et al.* (2014) Engineering crassulacean acid metabolism to improve water-use efficiency. *Trends in Plant Science*, 19, 327–338.

Bormann, B.T. & Gordon, J.C. (1984) Stand density effects in young red alder plantations: productivity, photosynthate partitioning and nitrogen fixation. *Ecology*, 65, 394–402.

Borregaard, M.K., Amorim, I.R., Borges, P.A.V. *et al.* (2017) Oceanic island biogeography through the lens of the general dynamic model: assessment and prospect. *Biological Reviews*, 92, 830–853.

Bossenbroek, J.M., Kraft, C.E. & Nekola, J.C. (2001) Prediction of long-distance dispersal using gravity models: zebra mussel invasion in inland lakes. *Ecological Applications*, 11, 1778–1788.

Boyce, M.S. (1984) Restitution of *r*- and *K*-selection as a model of density-dependent natural selection. *Annual Review of Ecology and Systematics*, 15, 427–447.

Boyd, P.W. (2002) The role of iron in the biogeochemistry of the Southern Ocean and equatorial Pacific: a comparison of *in situ* iron enrichments. *Deep-Sea Research II*, 49, 1803–1821.

Boyd, P.W., Bakker, D.C.E. & Chandler, C. (2012) A new database to explore the findings from large-scale ocean iron enrichments experiments. *Oceanography*, 25, 64–71.

Boyd, P.W., Jickells, T., Law, C.S. *et al.* (2007) Mesoscale iron enrichment experiments 1993–2005: synthesis and future directions. *Science*, 315, 612–617.

Boyer, E.W. & Howarth, R.W. (2008) Nitrogen fluxes from rivers to the coastal oceans. In: *Nitrogen in the Marine Environment*, 2nd edn (D. Capone, D.A. Bronk, M.R. Mulholland & E. J. Carpenter, eds), pp 1565–1587. Elsevier, Oxford.

Bradshaw, A.D. (1987) The reclamation of derelict land and the ecology of ecosystems. In: *Restoration Ecology* (W.R. Jordan III, M.E. Gilpin & J.D. Aber, eds), pp. 53–74. Cambridge University Press, Cambridge, UK.

Branch, T.A., Watson, R., Fulton, E.A. *et al.* (2010) The trophic fingerprint of marine fisheries. *Nature*, 468, 431–435.

Braune, B.M., Gaston, A.J., Hobson, K.A. *et al.* (2014) Changes in food web structure alter trends of mercury uptake at two seabird colonies in the Canadian Arctic. *Environmental Science and Technology*, 48, 13246–13252.

Breitburg, D., Levin, L.A., Oschlies, A. *et al.* (2018) Declining oxygen in the global ocean and coastal waters. *Science*, 359, 46.

Brewer, M.J. & Elliott, N.C. (2004) Biological control of cereal aphids in North America and mediating effects of host plant and habitat manipulations. *Annual Review of Entomology*, 49, 219–242.

Briand, F. (1983) Environmental control of food web structure. *Ecology*, 64, 253–263.

Briand, F. & Cohen, J.E. (1987) Environmental correlates of food chain length. *Science*, 238, 956–960.

Brigatti, E., Vieira, M.V., Kajin, M., Almeida, P.J.A.L., de Menezes, M.A. & Cerqueira, R. (2016) Detecting and modelling delayed density-dependence in abundance time series of a small mammal (*Didelphis aurita*). *Scientific Reports*, 6, 19553.

Briggs, C.J. & Hoopes, M.F (2004) Stabilizing effects in spatial parasitoid–host and predator–prey models: a review. *Journal of Theoretical Biology*, 65, 299–315.

Brockhurst, M.A., Morgan, A.D., Rainey, P.B. & Buckling, A. (2003) Population mixing accelerates coevolution. *Ecology Letters*, 6, 975–979.

Brokaw, N. & Busing, R.T. (2000) Niche versus chance and tree diversity in forest gaps. *Trends in Ecology and Evolution*, 15, 183–188.

Brooker, R.W., Karley, A.J., Newton, A.C., Pakeman, R.J. & Schob, C. (2016) Facilitation and sustainable agriculture: a mechanistic approach to reconciling crop production and conservation. *Functional Ecology*, 30, 98–107.

Brooker, R.W., Maestre, F.T., Callaway, R.M. *et al.* (2008) Facilitation in plant communities: the past, the present, and the future. *Journal of Ecology*, 96, 18–34.

Brower, L.P. & Corvinó, J.M. (1967) Plant poisons in a terrestrial food chain. *Proceedings of the National Academy of Science of the USA*, 57, 893–898.

Brown, J.H. & Davidson, D.W. (1977) Competition between seedeating rodents and ants in desert ecosystems. *Science*, 196, 880–882.

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage. V.A. & West, G.B. (2004) Toward a metabolic theory of ecology. *Ecology*, 85, 1771–1789.

Brown, J.H. & Kodric-Brown, A. (1977) Turnover rates in insular biogeography: effect of immigration and extinction. *Ecology*, 58, 445–449.

Brown, J.H. & Maurer, B.A. (1989) Macroecology: the division of food and space among species on continents. *Science*, 243, 1145–1150.

Brown, J.S., Kotler, B.P., Smith, R.J. & Wirtz, W.O. III (1988) The effects of owl predation on the foraging behaviour of heteromyid rodents. *Oecologia*, 76, 408–415.

Browne, R.A. (1981) Lakes as islands: biogeographic distribution, turnover rates, and species composition in the lakes of central New York. *Journal of Biogeography*, 8, 75–83.

Bruckmann, S.V., Krauss, J. & Steffan-Dewenter, I. (2010) Butterfly and plant specialists suffer from reduced connectivity in fragmented landscapes. *Journal of Applied Ecology*, 47, 799–809.

Brune, A. & Dietrich, C. (2015) The gut microbiota of termites: digesting the diversity in the light of ecology and evolution. *Annual Review of Microbiology*, 69, 145–166.

Brunet, A.K. & Medellín, R.A. (2001) The species–area relationship in bat assemblages of tropical caves. *Journal of Mammalogy*, 82, 1114–1122.

Bruno, J.F., Stachowicz, J.J. & Bertness, M.D. (2003) Inclusion of facilitation into ecological theory. *Trends in Ecology and Evolution*, 18, 119–125.

Bryant, D.A. & Frigaard, N-U. (2006) Prokaryotic photosynthesis and phototrophy illuminated. *Trends in Microbiology*, 14, 488–496.

Bryant, J.P., Chapin, F.S. III & Klein, D.R. (1983) Carbon/nutrient balance of boreal plants in relation to vertebrate herbivory. *Oikos*, 40, 357–368.

Brylinski, M. & Mann, K.H. (1973) An analysis of factors governing productivity in lakes and reservoirs. *Limnology and Oceanography*, 18, 1–14.

Buckling, A., Mclean, R.C., Brockhurst, M.A & Colgrave, N. (2009) The *Beagle* in a bottle. *Nature*, 457, 824–829.

Buesseler, K.O. (2012) The great iron dump. *Nature*, 487, 305–6.

Bulleri, F. & Malquori, F. (2015) High tolerance to simulated herbivory in the clonal seaweed, *Caulerpa cylindracea*. *Marine Environmental Research*, 107, 61–65.

Bullock, J.M., Mortimer, A.M. & Begon, M. (1994) Physiological integration among tillers of *Holcus lanatus*: age-dependence and responses to clipping and competition. *New Phytologist*, 128, 737–747.

Bullock, J.M., Moy, I.L., Pywell, R.F., Coulson, S.J., Nolan, A.M. & Caswell, H. (2002) Plant dispersal and colonization processes at local and landscape scales. In: *Dispersal Ecology* (J.M. Bullock, R.E. Kenward & R.S. Hails, eds), pp. 279–302. Blackwell Science, Oxford.

Burgin, L.E., Gloster, J., Sanders, C., Mellor, P.S., Gubbins, S. & Carpenter, S. (2012) Investigating incursions of bluetongue virus using a model of long-distance *Culicoides* biting midge dispersal. *Transboundary and Emerging Diseases*, 60, 263–272.

Burke, L., Reytar, K., Spalding, M. & Perry, A. (2011) *Reefs at Risk Revisited*. The World Resources Institute, Washington DC.

Burkhead, N.M (2012) Extinction rates of North American freshwater fishes, 1900–2010. *BioScience*, 62, 798–808.

Buscot, F., Munch, J.C., Charcosset, J.Y., Gardes, M., Nehls, U. & Hampp, R. (2000) Recent advances in exploring physiology and biodiversity of ectomycorrhizas highlight the functioning of these symbioses in ecosystems. *FEMS Microbiology Reviews*, 24, 601–614.

Busing, R.T. & Brokaw, N. (2002) Tree species diversity in temperate and tropical forest gaps: the role of lottery recruitment. *Folia Geobotanica*, 37, 33–43.

Buss, L.W. (1979) Byrozoan overgrowth interactions – the interdependence of competition for food and space. *Nature*, 281, 475–477.

Bustamante-Sanchez, M.A. & Armesto, J.J. (2012) Seed limitation during early forest succession in a rural landscape on Chiloe Island, Chile: implications for temperate forest restoration. *Journal of Applied Ecology*, 49, 1103–1112.

Butzin, M. & Pörtner, H-O. (2016) Thermal growth potential of Atlantic cod by the end of the 21st century. *Global Change Biology*, 22, 4162–4168.

Byrne, M., Gall, M., Wolfe, K. & Aguera, A. (2016) From pole to pole: the potential for the Arctic seastar *Asterias amurensis* to invade a warming Southern Ocean. *Global Change Biology*, 22, 3874–3887.

Caddy, J.F., Csirke, J., Garcia, S.M. & Grainger, R.J.R. (1998) How pervasive is 'fishing down marine food webs'? *Science*, 282, 1383a.

Cadotte, M.W., Carscadden, K. & Mirotchnick, N. (2011) Beyond species: functional diversity and the maintenance of ecological processes and services. *Journal of Applied Ecology*, 48, 1079–1087.

Cadotte, M.W., Mai, D.V., Jantz, S., Collins, M.D., Keele, M. & Drake, J.A. (2006) On testing the competition-colonization trade-off in a multispecies assemblage. *The American Naturalist*, 168, 704–709.

Cain, M.L., Pacala, S.W., Silander, J.A. & Fortin, M.-J. (1995) Neighbourhood models of clonal growth in the white clover *Trifolium repens*. *American Naturalist*, 145, 888–917.

Callaghan, T.V. (1976) Strategies of growth and population dynamics of plants: 3. *Growth and population dynamics of Carex bigelowii in an alpine environment*. *Oikos*, 27, 402–413.

Cameron, H., Monro, K., Malerba, M., Munch, S. & Marshall, D. (2016) Why do larger mothers produce larger offspring? *Ecology*, 97, 3452–3459.

Campbell, J., Antoine, D., Armstrong, R. *et al.* (2002) Comparison of algorithms for estimating ocean primary productivity from surface chlorophyll, temperature, and irradiance. *Global Biogeochemical Cycles*, 16, 91–96.

Campbell, J.L., Driscoll, C.T., Pourmokhtarian, A. & Hayhoe, K. (2011) Streamflow responses to past and projected future changes in climate at the Hubbard Brook Experimental Forest, New Hampshire, United States. *Water Resources Research*, 47, W02514.

Canham, C.D., Ruscoe, W.A., Wright, E.F. & Wilson, D.J. (2014) Spatial and temporal variation in tree seed production and dispersal in a New Zealand temperate rainforest. *Ecosphere*, 5, 49.

Canto, T., Aranda, M.A. & Fereres, A (2009) Climate change effects on physiology and population processes of hosts and vectors that influence the spread of hemipteran-borne plant viruses. *Global Change Biology*, 15, 1884–1894.

Caraco, T. & Kelly, C.K. (1991) On the adaptive value of physiological integration of clonal plants. *Ecology*, 72, 81–93.

Cardillo, M. & Bromham, L. (2001) Body size and risk of extinction in Australian mammals. *Conservation Biology*, 15, 1435–1440.

Cardillo, M., Mace, G.M., Gittleman, J.L., Jones, K.E., Bielby, J. & Purvis, A. (2008) The predictability of extinction: biological and external correlates of decline in mammals. *Proceedings of the Royal Society of London, Series B*, 275, 1441–1448.

Cardinale, B.J., Matulich, K.L., Hooper, D.U. *et al.* (2011) The functional role of producer diversity in ecosystems. *American Journal of Botany*, 98, 572–592.

Cardinale, B.J., Wright, J.P., Cadotte, M.W. *et al.* (2007) Impacts of plant diversity on biomass production increase through time because of species complementarity. *Proceedings of the National Academy of Sciences of the USA*, 104, 18123–18128.

Carignan, R., Planas, D. & Vis, C. (2000) Planktonic production and respiration in oligotrophic Shield lakes. *Limnology and Oceanography*, 45, 189–199.

Carmel, Y., Suprunenko, Y.F,. Kunin, W.E. *et al.* (2017) Using exclusion rate to unify niche and neutral perspectives on coexistence. *Oikos*, 126, 1451–1458.

Carne, P.B. (1969) On the population dynamics of the eucalyptdefoliating chrysomelid *Paropsis atomaria* OI. *Australian Journal of Zoology*, 14, 647–672.

Carpenter, S.R. & Biggs, R. (2009) Freshwaters: Managing across scales in space and time. In *Principles of Ecosystem Stewardship: Resilience-Based Natural Resource Management in a Changing World* (F.S. Chapin III, G.P. Kofinas, C. Folke, eds). Springer Verlag, New York.

Carson, W.P. & Root, R.B. (1999) Top-down effects of insect herbivores during early succession: influence on biomass and plant dominance. *Oecologia*, 121, 260–272.

Carvajal-Endara, S., Hendry, A.P., Emery, N.C. & Davies, T.J. (2017) Habitat filtering not dispersal limitation shapes oceanic island floras: species assembly of the Galápagos archipelago. *Ecology Letters*, 20, 495–204.

Cary, J.R. & Keith, L.B. (1979) Reproductive change in the 10-year cycle of snowshoe hares. *Canadian Journal of Zoology*, 57, 375–390.

Castellon, T.D. & Sieving, K.E. (2006) An experimental test of matrix permeability and corridor use by an endemic understory bird. *Conservation Biology*, 20, 135–145.

Castonguay, Y., Bertrand, A., Michaud, R. and Laberge, S. (2011) Cold-Induced Biochemical and Molecular Changes in Alfalfa Populations Selectively Improved for Freezing Tolerance. *Crop Science*, 51, 2132–2144.

Castro-Arellano, I. & Lacher, T.E. Jr (2009) Temporal niche segregation in two rodent assemblages of subtropical Mexico. *Journal of Tropical Ecology*, 25, 593–603.

Caswell, H. (1989) Analysis of life table response experiments. 1. Decomposition of effects on population growth rate. *Ecological Modelling*, 46, 221–237.

Caswell, H. (2001) *Matrix Population Models*, 2nd edn. Sinauer, Sunderland, MA.

Caughley, G. (1994) Directions in conservation biology. *Journal of Animal Ecology*, 63, 215–244.

Cavieres, L.A., Brooker, R.W., Butterfield, B.J. *et al.* (2014) Facilitative plant interactions and climate simultaneously drive alpine plant diversity. *Ecology Letters*, 17, 193–202.

Cayuela, L., Gotelli, N.J. & Colwell, R.K. (2015) Ecological and biogeographic null hypotheses for comparing rarefaction curves. *Ecological Monographs*, 85, 437–455.

Ceballos, G., Ehrlich, P.R. & Dirzo, R. (2017) Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences of the USA*, 114, E6089–E6096.

Cebrian, J. (1999) Patterns in the fate of production in plant communities. *American Naturalist*, 154, 449–468.

Cebrian, J. & Lartigue, J. (2004) Patterns of herbivory and decomposition in aquatic and terrestrial ecosystems. *Ecological Monographs*, 74, 237–259.

Chadburn, S.E., Burke, E.J., Cox, P.M., Friedlingstein, P., Hugelius, G. & Westermann, S. (2017) An observation-based constraint on permafrost loss as a function of global warming. *Nature Climate Change*, 7, 340–345.

Chantrey, J., Dale, T.D., Read, J.M. *et al.* (2014) European red squirrel population dynamics driven by squirrelpox at a gray squirrel invasion interface. *Ecology and Evolution*, 19, 3788–3799.

Chao, A., Gotelli, N.J., Hsieh, T.C. *et al.* (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, 84, 45–67.

Charlesworth, D. & Charlesworth, B. (1987) Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics*, 18, 237–268.

Charnov, E.L. (1976a) Optimal foraging: attack strategy of a mantid. *American Naturalist*, 110, 141–151.

Charnov, E.L. (1976b) Optimal foraging: the marginal value theorem. *Theoretical Population Biology*, 9, 129–136.

Charnov, E.L. & Krebs, J.R. (1974) On clutch size and fitness. *Ibis*, 116, 217–219.

Chase, J.M. & Leibold, M.A. (2002) Spatial scale dictates the productivity-biodiversity relationship. *Nature*, 416, 427–430.

Chasen, E.M., Undersander, D.J. & Cullen, E.M. (2015) Revisiting the economic injury level and economic threshold model for potato leafhopper (Hemiptera: Cicadellidae) in alfalfa. *Journal of Economic Entomology*, 108, 1748–1756.

Chapin, F.S., Matson, P.A. & Vitousek, P.M. (2011) *Principles of Terrestrial Ecosystem Ecology*, 2nd edn. Springer, New York.

Chapman, J.W., Nilsson, C., Lim, K.S., Backman, J., Reynolds, D.R. & Alerstam. T. (2015) Adaptive strategies in nocturnally migrating insects and songbirds: contrasting responses to wind. *Journal of Animal Ecology*, 85, 115–124.

Chavez, F.P., Messie, M. & Pennington, J.T. (2011) Marine primary production in relation to climate variability and change. *Annual Review of Marine Science*, 3, 227–260.

Chen, H-W., Shao, K-T., Liu, C.W-J., Lin, W-H. & Liu, W-C. (2011) The reduction of food web robustness by parasitism: fact and artefact. *International Journal of Parasitology*, 41, 627–634.

Cheney, K.L. & Côté, I.M. (2005) Mutualism or parasitism? The variable outcome of cleaning symbioses. *Biology Letters*, 1, 162–165.

Cherrett, J.M., Powell, R.J. & Stradling, D.J. (1989) The mutualism between leaf-cutting ants and their fungus. In: *Insect/Fungus Interactions* (N. Wilding, N.M. Collins, P.M. Hammond & J.F. Webber, eds), pp. 93–120. Royal Entomological Society Symposium No. 14. Academic Press, London.

Chesson, P.L. (2000) Mechanisms of maintenance of species diversity. *Annual Reviews of Ecology and Systematics*, 31, 343–366.

Chisholm, R.A. & Pacala, S.W. (2010) Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. *Proceedings of the National Academy of Sciences of the USA*, 36, 15821–15825.

Choquenot, D. (1998) Testing the relative influence of intrinsic and extrinsic variation in food availability on feral pig populations in Australia's rangelands. *Journal of Animal Ecology*, 67, 887–907.

Christian, J.J. (1980) Endocrine factors in population regulation. In: *Biosocial Mechanisms of Population Regulation* (M.N. Cohen, R.S. Malpass & H.G. Klein, eds), pp. 55–115. Yale University Press, New Haven, CT.

Clark, C.W. (2010) *Mathematical Bioeconomics; The Mathematics of Conservation*, 3rd edn. Wiley, Hoboken, NJ.

Clarke, A. (2017) *Principles of Thermal Ecology: Temperature, Energy and Life*. Oxford University Press, Oxford.

Clarke, A. & Crame, J.A. (2003) The importance of historical processes in global patterns of diversity. In: *Macroecology: concepts and consequences* (T.M. Blackburn & K.J. Gaston, eds), pp. 130–151. Blackwell Publishing, Oxford.

Clarke, B.C. & Partridge, L. (eds) (1988) Frequency dependent selection. *Philosophical Transactions of the Royal Society of London, Series B*, 319, 457–645.

Clauss, M., Dittmann, M.T., Müller, D.W.H., Zerbe, P. & Codron, D. (2014) Low scaling of a life history variable: analysing eutherian gestation periods with and without phylogeny-informed statistics. *Mammalian Biology*, 79, 9–16.

Clements, F.E. (1916) *Plant Succession: Analysis of the development of vegetation*. Carnegie Institute of Washington Publication No. 242. Washington, DC.

Cleveland, C.C. & Liptzin, D. (2006) C:N:P stoichiometry in soil: is there a 'Redfield ratio' for the microbial biomass? *Biogeochemistry*, 85, 235–252.

Clutton-Brock, T.H. & Harvey, P.H. (1979) Comparison and adaptation. *Proceedings of the Royal Society of London, Series B*, 205, 547–565.

Clutton-Brock, T.H., Major, M., Albon, S.D. & Guinness, F.E. (1987) Early development and population dynamics in red deer. I. Density-dependent effects on juvenile survival. *Journal of Animal Ecology*, 56, 53–67.

Clutton-Brock, T.H., Major, M. & Guinness, F.E. (1985) Population regulation in male and female red deer. *Journal of Animal Ecology*, 54, 831–836.

Cobb, A.H. & Reade, J.P.H. (2010) *Herbicides and Plant Physiology*, 2nd edn. Wiley Blackwell, Oxford.

Cockell, C.S. & Lee, P. (2002) The biology of impact craters – a review. *Biological Reviews*, 77, 279–310.

Cohen, J.E, (1995) *How Many People Can the Earth Support?* W.W. Norton & Co., New York.

Cohen, J.E. (2005) Human population grows up. *Scientific American*, 293, 48–55.

Cohen, J.E., Jonsson, T. & Carpenter, S.R. (2003) Ecological community description using food web, species abundance, and body-size. *Proceedings of the National Academy of Science of the USA*, 100, 1781–1786.

Colbourne, R., Bassett, S., Billing, T. *et al.* (2005) The development of Operation Nest Egg as a tool in the conservation management of kiwi. *Science for Conservation*, 259, 24.

Cole, J.J., Findlay, S. & Pace, M.L. (1988) Bacterial production in fresh and salt water ecosystems: a cross-system overview. *Marine Ecology Progress Series*, 4, 1–10.

Cole, L.C. (1954) The population consequences of life history phenomena. *Quarterly Review of Biology*, 25, 103–127.

Cole, M., Lindeque, P., Fileman, E. *et al.* (2016) Microplastics alter the properties and sinking rates of zooplankton faecal pellets. *Environmental Science and Technology*, 6, 3239–3246.

Coley, P.D. (1988) Effects of plant growth rate and leaf lifetime on the amount and type of anti-herbivore defense. *Oecologia*, 74, 531–536.

Coley, P.D., Bryant, J.P. & Chapin, F.S. III. (1985) Resource availability and plant antiherbivore defense. *Science*, 230, 895–899.

Collado-Vides, L. (2001) Clonal architecture in marine macroalgae: ecological and evolutionary perspectives. *Evolutionary Ecology*, 15, 531–545.

Collatz, G.J., Berry, J.A. & Clark, J.S. (1998) Effects of climate and atmospheric CO_2 partial pressure on the global distribution of C_4 grasses: present, past, and future. *Oecologia*, 114, 441–454.

Collins, M.D., Ward, S.A. & Dixon, A.F.G. (1981) Handling time and the functional response of *Aphelinus thomsoni*, a predator and parasite of the aphid, *Drepanosiphum platanoidis*. *Journal of Animal Ecology*, 50, 479–487.

Colwell, R.K., Chao, A., Gotelli, N.J. *et al.* (2012) Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology*, 5, 3–21.

Colwell, R.K. & Hurtt, G.C. (1994) Non-biological gradients in species richness and a spurious Rapoport effect. *American Naturalist*, 144, 570–595.

Comins, H.N., Hassell, M.P. & May, R.M. (1992) The spatial dynamics of host–parasitoid systems. *Journal of Animal Ecology*, 61, 735–748.

Comita, L.S., Queenborough, S.A., Murphy, S.J. *et al.* (2014) Testing predictions of the Janzen–Connell hypothesis: a meta-analysis of experimental evidence for distance- and density-dependent seed and seedling survival. *Journal of Ecology*, 102, 845–856.

Compton, S.G. (2001) Sailing with the wind: dispersal by small flying insects. In: *Dispersal Ecology* (J.M. Bullock, R.E. Kenward & R.S. Hails, eds), pp. 113–133. Blackwell Science, Oxford.

Connell, J.H. (1970) A predator–prey system in the marine intertidal region. I. *Balanus glandula* and several predatory species of *Thais*. *Ecological Monographs*, 40, 49–78.

Connell, J.H. (1971) On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. In: *Dynamics of Populations* (P.J. den Boer & G.R. Gradwell, eds), pp. 298–310. Proceedings of the Advanced Study Institute in Dynamics of Numbers in Populations, Oosterbeck. Centre for Agricultural Publishing and Documentation, Wageningen.

Connell, J.H. (1978) Diversity in tropical rainforests and coral reefs. *Science*, 199, 1302–1310.

Connell, J.H. (1979) Tropical rain forests and coral reefs as open nonequilibrium systems. In: *Population Dynamics* (R.M. Anderson, B.D. Turner & L.R. Taylor, eds), pp. 141–163. Blackwell Scientific Publications, Oxford.

Connell, J.H. (1980) Diversity and the coevolution of competitors, or the ghost of competition past. *Oikos*, 35, 131–138.

Connell, J.H. (1983) On the prevalence and relative importance of interspecific competition: evidence from field experiments. *American Naturalist*, 122, 661–696.

Connell, J.H. (1990) Apparent versus ‘real’ competition in plants. In: *Perspectives on Plant Competition* (J.B. Grace & D. Tilman, eds), pp. 9–26. Academic Press, New York.

Connor, E.H. & Simberloff, D. (1979) The assembly of species communities: chance or competition? *Ecology*, 60, 1132–1140.

Consoli, F.L. & Vinson, S.B. (2002) Clutch size, development and wing morph differentiation of *Melittobia digitata*. *Entomologia Experimentalis et Applicata*, 102, 135–143.

Cooch, E.G. & White, G. (2019) Program MARK. A Gentle Introduction. http://www.phidot.org/software/mark/docs/book/ (last accessed November 2019).

Cook, L.M. & Saccheri, I.J. (2013) The peppered moth and industrial melanism: evolution of a natural selection case study. *Heredity*, 110, 207–212.

Cooper, J.K., Li, J. & Montagnes, D.J.S. (2012) Intermediate fragmentation per se provides stable predator-prey metapopulation dynamics. *Ecology Letters*, 15, 856–863.

Cooper, N., Thomas, G.H. & FitzJohn, R.G. (2016) Shedding light on the 'dark side' of phylogenetic comparative methods. *Methods in Ecology and Evolution*, 7, 693–699.

Coops, N.C., Waring, R.H., Plowright, A., Lee, J. & Dilts, T.E. (2016) Using remotely-sensed land cover and distribution modeling to estimate tree species migration in the Pacific Northwest region of North America. *Remote Sensing*, 8, 65.

Corbari, L., Sorbe, J.C. & Massabuau, J.C. (2005) Video study of the caprellid amphipod *Parvipalpus major*: morpho-functional and behavioural adaptations to deep-sea bottoms. *Marine Biology*, 146, 363–371.

Coristine, L.E. & Kerr, J.T. (2015) Temperature-related geographical shifts among passerines: contrasting processes along poleward and equatorward range margins. *Ecology and Evolution*, 5, 5162–5176.

Cornwell, W.K. & Grubb, P.J. (2003) Regional and local patterns in plant species richness with respect to resource availability. *Oikos*, 100, 417–428.

Cortes, E. (2002) Incorporating uncertainty into demographic modeling: application to shark populations and their conservation. *Conservation Biology*, 16, 1048–1062.

Costantino, R.F., Desharnais, R.A., Cushing, J.M. & Dennis, B. (1997) Chaotic dynamics in an insect population. *Science*, 275, 389–391.

Costanza, R., de Groot, R., Braat, L. *et al.* (2017) Twenty years of ecosystem services: how far have we come and how far do we still need to go? *Ecosystem Services*, 28, 1–16.

Cotrufo, M.F., Ineson, P., Scott, A. *et al.* (1998) Elevated CO_2 reduces the nitrogen concentration of plant tissues. *Global Change Biology*, 4, 43–54.

Cottee-Jones, H.E.W. & Whittaker, R.J. (2012) The keystone species concept: a critical appraisal. *Frontiers of Biogeography*, 4, 117–127.

Cottingham, K.L., Brown, B.L. & Lennon, J.T. (2001) Biodiversity may regulate the temporal variability of ecological systems. *Ecology Letters*, 4, 72–85.

Coulson, T. (2012) Integral projections models, their construction and use in posing hypotheses in ecology. *Oikos*, 121, 1337–1350.

Courchamp, F., Fournier, A., Bellard, C. *et al.* (2017) Invasion biology: specific problems and possible solutions. *Trends in Ecology and Evolution*, 32, 13–22.

Cowling, R.M., Pressey, R.L., Rouget, M. & Lombard, A.T. (2003) A conservation plan for a global biodiversity hotspot – the Cape Floristic Region, South Africa. *Biological Conservation*, 112, 191–216.

Coyne, J.A. & Orr, H.A. (2004) *Speciation*. Sinauer Associates, Sunderland, MA.

Craine, J.M. (2005) Reconciling plant strategy theories of Grime and Tilman. *Journal of Ecology*, 93, 1041–1052.

Crawford, D.W., Purdie, D.A., Lockwood, A.P.M. & Weissman, P. (1997) Recurrent red-tides in the Southampton Water estuary caused by the phototrophic ciliate *Mesodinium rubrum*.

Estuarine, Coastal and Shelf Science, 45, 799–812.

Crawley, M.J. (1986) The structure of plant communities. In: *Plant Ecology* (M.J. Crawley, ed.), pp. 1–50. Blackwell Scientific Publications, Oxford.

Crawley, M.J. (1989) Insect herbivores and plant population dynamics. *Annual Review of Entomology*, 34, 531–564.

Crawley M.J., Johnston, A.E., Silvertown, J. *et al.* (2005) Determinants of species richness in the Park Grass experiment. *The American Naturalist*, 165, 179–192.

Creed, J.C., Kain, J.M. & Norton, T.A. (1998) An experimental evaluation of density and plant size in two large brown seaweeds. *Journal of Phycology*, 34, 39–52.

Cressler, C.E., McLeod, D.V., Rozins, C., Van Den Hoogen, J. & Day, T. (2016) The adaptive evolution of virulence: a review of theoretical predictions and empirical tests. *Parasitology*, 143, 915–930.

Cristofari, R., Liu, X., Bonadonna, F. *et al.* (2018) Climate-driven range shifts of the king penguin in a fragmented ecosystem. *Nature Climate Change*, 8, 245–251.

Crocker, R.L. & Major, J. (1955) Soil development in relation to vegetation and surface age at Glacier Bay, Alaska. *Journal of Ecology*, 43, 427–448.

Cropp, R. & Norbury, J. (2018) Linking obligate mutualism models in an extended consumer-resource framework. *Ecological Modelling*, 374, 1–13.

Cullen, J.J., Davis, R.F. & Huot, Y. (2012) Spectral model of depth-integrated water column photosynthesis and its inhibition by ultraviolet radiation. *Global Biogeochemical Cycles*, 26, GB1011, doi:10.1029/2010GB003914.

Cummings, J., Reid, N., Davies, I. & Grant, C. (2005) Adaptive restoration of sand-mined areas for biological conservation. *Journal of Applied Ecology*, 42, 160–170.

Cummins, K.W. (1974) Structure and function of stream ecosystems. *Bioscience*, 24, 631–641.

Currie, C.R. (2001) A community of ants, fungi, and bacteria: a multilateral approach to studying symbiosis. *Annual Review of Microbiology*, 55, 357–380.

Daan, S., Dijkstra, C. & Tinbergen, J.M. (1990) Family planning in the kestrel (*Falco tinnunculus*): the ultimate control of covariation of laying date and clutch size. *Behavior*, 114, 83–116.

Dainese, M., Aikio, S., Hulme, P.E., Bertolli, A., Prosser, F. & Marini, L. (2017) Human disturbance and upward expansion of plants in a warming climate. *Nature Climate Change*, 7, 577–582.

D'Amen, M., Mod, H.K., Gotelli, N.J. & Guisan, A. (2018) Disentangling biotic interactions, environmental filters, and dispersal limitation as drivers of species co-occurrence. *Ecography*, 41, 1233–1244.

Damos, P. (2015) Modular structure of web-based decision support systems for integrated pest management. *A review. Agronomy for Sustainable Development*, 35, 1347–1372.

Dangremond, E.M., Pardini, E.A. & Knight, T.M. (2010) Apparent competition with an invasive plant hastens the extinction of an endangered lupine. *Ecology*, 91, 2261–2271.

Dani, K.G.S. & Kodandaramaiah, U. (2017) Plant and animal reproductive strategies: lessons from offspring size and number tradeoffs. *Frontiers in Ecology and Evolution*, 5, 38.

Darwin, C. (1859) *The Origin of Species by Means of Natural Selection*, 1st edn. John Murray, London.

Darwin, C. (1888) *The Formation of Vegetable Mould Through the Action of Worms*. John Murray, London.

Daufresne, T. & Loreau, M. (2001) Ecological stoichiometry, primary producer–decomposer interactions, and ecosystem persistence. *Ecology*, 82, 3069–3082.

Davidian, E., Courtiol, A., Wachter, B., Hofer, H. &. Höner, O.P. (2016) Why do some males choose to breed at home when most other males disperse? *Scientific Advances*, 2, e1501236.

Davies, S.J., Palmiotto, P.A., Ashton, P.S., Lee, H.S. & Lafrankie, J.V. (1998) Comparative ecology of 11 sympatric species of *Macaranga* in Borneo: tree distribution in relation to horizontal and vertical resource heterogeneity. *Journal of Ecology*, 86, 662–673.

Davis, A.S., Hill, J.D., Chase, C.A., Johanns, A.M. & Liebman, M. (2012) Increasing cropping system diversity balances productivity, profitability and environmental health. *PLoS One*, 7, e47149.

Davis, M.B. (1976) Pleistocene biogeography of temperate deciduous forests. *Geoscience and Man*, 13, 13–26.

Davis, M.B. & Shaw, R.G. (2001) Range shifts and adaptive responses to quarternary climate change. *Science*, 292, 673–679.

Davison, A. & Blaxter, M. (2005) Ancient origin of glycosyl hydrolase family 9 cellulase genes. *Molecular Biology and Evolution*, 22, 1273–1284.

Dayan, T. & Simberloff, D. (2005) Ecological and community-wide character displacement: the next generation. *Ecology Letters*, 8, 875–894.

Dayan, T., Simberloff, E., Tchernov, E. & Yom-Tov, Y. (1989) Inter and intraspecific character displacement in mustelids. *Ecology*, 70, 1526–1539.

de Alencar, F.L.S., Navoni, J.A., do Amaral, V.S. (2017) The use of bacterial bioremediation of metals in aquatic environments in the twenty-first century: a systematic review. *Environmental Science and Pollution*, 24, 16545–16559.

de Brito Martins, A. & de Aguiar, A.M. (2016) Barriers to gene flow and ring species formation. *Evolution*, 71–2, 442–448.

De Jong, G. & van Noordwijk, A.J. (1992) Acquisition and allocation of resources: genetic (co)variances, selection, and life histories. *American Naturalist*, 139, 749–770.

De Kroon, H., Visser, E.J.W., Huber, H., Mommer, L. & Hutchings, M.J. (2009) A modular concept of plant foraging behaviour: the interplay between local responses and systemic control. *Plant, Cell and Environment*, 32, 704–712.

de Laat, A.T.J., van Weele, M. & van der A, R.J. (2017) Onset of stratospheric ozone recovery in the Antarctic ozone hole in assimilated daily total ozone columns. *Journal of Geophysical Research: Atmospheres*, 122, 11 880–11 899.

De Meester, N. & Bonte, D. (2010) Information use and density-dependent emigration in an agrobiont spider. *Behavioral Ecology*, 21, 992–998.

De Waal, C., Anderson, B. & Ellis, A.G. (2016) Dispersal, dormancy and life-history tradeoffs at the individual, population and species levels in southern African Asteraceae. *New Phytologist*, 210, 256–265.

de Wit, C.T. (1965) Photosynthesis of leaf canopies. *Verslagen van Landbouwkundige Onderzoekingen*, 663, 1–57.

de Wit, C.T., Tow, P.G. & Ennik, G.C. (1966) Competition between legumes and grasses. *Verslagen van landbouwkundige Onderzoekingen*, 112, 1017–1045.

Dearnaley, J.D.W., Martos, F. & Selosse, M-A. (2013) *Orchid mycorrhizas: molecular ecology, physiology, evolution and conservation aspects*. Springer, Berlin.

Debona, D., Rodrigues, F.A. & Datnoff, L.E. (2017) Silicon's role in abiotic and biotic plant stresses. *Annual Review of Phytopathology*, 55, 85–107.

Deevey, E.S. (1947) Life tables for natural populations of animals. *Quarterly Review of Biology*, 22, 283–314.

del Val, E. & Crawley, M.J. (2005) Are grazing increaser species better tolerators than decreasers? An experimental assessment of defoliation tolerance in eight British grassland species. *Journal of Ecology*, 93, 1005–1016.

Delavaux, C.S., Smith-Ramesh, L.M. & Kuebbing, S.E. (2017) Beyond nutrients: a meta-analysis of the diverse effects of arbuscular mycorrhizal fungi on plants and soils. *Ecology*, 98, 2111–2119.

D'Elia, J., Haig, S.M., Johnson, M., Marcot, B.G. & Young, R. (2015) Activity-specific ecological niche models for planning reintroductions of California condors (Gymnogyps californianus). *Biological Conservation*, 184, 90–99.

Delibes-Mateos, M., Smith, A.T., Slobodchikoff, C.N. & Swenson, J.E. (2011) The paradox of keystone species persecuted as pests: a call for the conservation of abundant small mammals in their native range. *Biological Conservation*, 14, 1335–1346.

Della Venezia, L., Samson, J. & Leung, B. (2018) The rich get richer: invasion risk across North America from the aquarium pathway under climate change. *Diversity and Distributions*, 24, 285–296.

DeLong, J.P., Okie, J.G., Moses, M.E., Sibly, R.M. & Brown, J.H. (2010) Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. *Proceedings of the National Academy of Sciences of the USA*, 107, 12941–12945.

Desbruyères, D., Segonzac, M. & Bright, M. (eds) (2006) *Handbook of Deep-Sea Vent Fauna: second completely revised edition*. *Denisia*, 18, 434–455.

Descamps, S., Boutin, S., McAdam, A.G., Berteaux, D. & Gaillard, J-M. (2009) Survival costs of reproduction vary with age in North American red squirrels. *Proceedings of the Royal Society of London, Series B*, 276, 1129–1135.

Descamps-Julien, B. & Gonzalez, A. (2005) Stable coexistence in a fluctuating environment: an experimental demonstration. *Ecology*, 86, 2815–2824.

Devine, W.D. & Harrington, T.B. (2011) Aboveground growth interactions of paired conifer seedlings in close proximity. *New Forests*, 41, 163–178.

Diamond, J.M. (1975) Assembly of species communities. In: *Ecology and Evolution of Communities* (M.L. Cody & J.M. Diamond, eds), pp. 342–444. Belknap, Cambridge, MA.

Diaz, R.J. & Rosenberg, R. (2008) Spreading dead zones and consequences for marine ecosystems. *Science*, 321, 926–929.

Dickey-Collas, M., Nash, R.D.M., Brunel, T. *et al.* (2010) Lessons learned from stock collapse and recovery of North Sea herring: a review. *ICES Journal of Marine Science*, 67, 1875–1886.

Dieckmann, U., Law, R. & Metz, J.A.J. (2000) *The Geometry of Ecological Interactions: simplifying spatial complexity*. Cambridge University Press, Cambridge, UK.

Djagoun, C.A.M.S., Codron, D., Sealy, J., Mensah, G.A. & Sinsin, B. (2016) Isotopic niche structure of a mammalian herbivore assemblage from a West African savanna: body mass and seasonality effect. *Mammalian Biology*, 81, 644–650.

Dobson, A.P. & Hudson, P.J. (1992) Regulation and stability of a freeliving host–parasite system: *Trichostrongylus tenuis* in red grouse. II. Population models. *Journal of Animal Ecology*, 61, 487–498.

Dobson, F.S. (2013) The enduring question of sex-biased dispersal: Paul J. Greenwood's (1980) seminal contribution. *Animal Behaviour*, 85, 299–304.

Dodds, P.N. & Rathjen, J.P. (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature Reviews Genetics*, 11, 539–548.

Dodson, S.I., Arnott, S.E. & Cottingham, K.L. (2000) The relationship in lake communities between primary productivity and species richness. *Ecology*, 81, 2662–2679.

Doledec, S., Phillips, N., Scarsbrook, M., Riley, R.H. & Townsend, C.R. (2006) Comparison of structural and functional approaches to determining landuse effects on grassland stream communities. *Journal of the North American Benthological Society*, 25, 44–60.

Donohue, I., Hillebrand, H., Montoya, J.M. *et al.* (2016) Navigating the complexity of ecological stability. *Ecology Letters*, 19, 1172–1185.

Doube, B.M. (1987) Spatial and temporal organization in communities associated with dung pads and carcasses. In: *Organization of Communities: past and present* (J.H.R. Gee & P.S. Giller, eds), pp. 255–280. Blackwell Scientific Publications, Oxford.

Douglas, A.E. (2014) The molecular basis of bacterial-insect symbiosis. *Journal of Molecular Biology*, 426, 3830–3837.

Douglas, E. (1998) Nutritional interactions in insect–microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. *Annual Review of Entomology*, 43, 17–37.

Drake, J.E., Davis, S.C., Raetz, L.M. and Delucia E.H. (2011) Mechanisms of age-related changes in forest production: the influence of physiological and successional changes. *Global Change Biology*, 17, 1522–1535.

Duan, L., Yu, Q., Zhang, Q. *et al.* (2016) Acid deposition in Asia: emissions, deposition, and ecosystem effects. *Atmospheric Environment*, 146, 55–69.

Duarte, C.M., Marbà, N., Gacia, E. *et al.* (2010) Seagrass community metabolism: assessing the carbon sink capacity of seagrass meadows. *Global Biogeochemical Cycles*, 24, GB4032, doi:10.1029/2010GB003793.

Ducrey, M. & Labbé, P. (1985) Étude de la régénération naturale contrôlée en fôret tropicale humide de Guadeloupe. I. Revue bibliographique, milieu naturel et élaboration d'un protocole expérimental. *Annales Scientifiques Forestière*, 42, 297–322.

Ducrey, M. & Labbé, P. (1986) Étude de la régénération naturale contrôlée en fôret tropicale humide de Guadeloupe. II. Installation et croissance des semis après les coupes d'ensemencement. *Annales Scientifiques Forestière*, 43, 299–326.

Dudley, N. (Ed.) (2008) *Guidelines for Applying Protected Area Management Categories*. IUCN, Gland, Switzerland.

Duggan, I.C., Green, J.D. & Shiel, R.J. (2002) Distribution of rotifer assemblages in North Island, New Zealand, lakes: relationships to environmental and historical factors. *Freshwater Biology*, 47, 195–206.

Dulvy, N.K. & Reynolds, J.D. (2002) Predicting extinction vulnerability in skates. *Conservation Biology*, 16, 440–450.

Duncan, R.P., Cassey, P. & Blackburn, T.M. (2009) Do climate envelope models transfer? A manipulative test using dung beetle introductions. *Proceedings of the Royal Society of London, Series B*, 267, 1449–1457.

Dunne, J.A., Lafferty, K.D., Dobson, A.P. *et al.* (2013) Parasites affect food web structure primarily through increased diversity and complexity. *PLoS Biology*, 11, e1001579.

Dunne, J.A. & Williams, R.J. (2009) Cascading extinctions and community collapse in model food webs. *Philosophical Transactions of the Royal Society of London, Series B*, 364, 1711–1723.

Dunne, J.A., Williams, R.J. & Martinez, N.J. (2002) Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology Letters*, 5, 558–567.

Dusenge, M.E., Wallin, G., Gårdesten, J. *et al.* (2015) Photosynthetic capacity of tropical montane tree species in relation to leaf nutrients, successional strategy and growth temperature. *Oecologia*, 177, 1183–1194.

Dybzinski, R. & Tilman, D. (2007) Resource use patterns predict long-term outcomes of plant competition for nutrients and light. *American Naturalist*, 170, 305–318.

Eamus, D. (1999) Ecophysiological traits of deciduous and evergreen woody species in the seasonally dry tropics. *Trends in Ecology and Evolution*, 14, 11–16.

Ebert, D., Zschokke-Rohringer, C.D. & Carius, H.J. (2000) Dose effects and density-dependent regulation in two microparasites of *Daphnia magna*. *Oecologia*, 122, 200–209.

Egorova, T., Rozanov, E., Gröbner, J., Hauser, M. and Schmutz, W. (2013) Montreal Protocol Benefits simulated with CCM SOCOL. *Atmospheric Chemistry and Physics*, 13, 3811–3823.

Ehrlich, P. & Raven, P.H. (1964) Butterflies and plants: a study in coevolution. *Evolution*, 18, 586–608.

Elith, J. & Graham, C.H. (2009) Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, 32, 66–77.

Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Reviews of Ecology Evolution and Systematics*, 40, 677–697.

Elkington, J. (2002) *Cannibals with Forks: The Triple Bottom Line of 21st Century Business*. Capstone, Oxford.

Elliott, J.K. & Mariscal, R.N. (2001) Coexistence of nine anenomefish species: differential host and habitat utilization, size and recruitment. *Marine Biology*, 138, 23–36.

Elliott, J.M. (1993) The self-thinning rule applied to juvenile sea-trout, *Salmo trutta*. *Journal of Animal Ecology*, 62, 371–379.

Elliott, J.M. (1994) *Quantitative Ecology and the Brown Trout*. Oxford University Press, Oxford.

Ellis, C.J. (2012) Lichen epiphyte diversity: a species, community and trait-based review. *Perspectives in Plant Ecology, Evolution and Systematics*, 14, 131–152.

Ellis, E.C., Klein Goldewijk, K., Siebert, S., Lightman, D. & Ramankutty, N. (2010) Anthropogenic transformation of the biomes, 1700 to 2000. *Global Ecology and Biogeography*, 19, 589–606.

Elmes, G.W., Akino, T., Thomas, J.A., Clarke, R.T. & Knapp, J.J. (2002) Interspecific differences in cuticular hydrocarbon profiles of *Myrmica* ants are sufficiently consistent to explain host specificity by *Maculinea* (large blue) butterflies. *Oecologia*, 130, 525–535.

Elser, J.J. & Urabe, J. (1999) The stoichiometry of consumer-driven nutrient recycling: theory, observations, and consequences. *Ecology*, 80, 735–751.

Elton, C.S. (1924) Periodic fluctuations in the number of animals: their causes and effects. *British Journal of Experimental Biology*, 2, 119–163.

Elton, C. (1927) *Animal Ecology*. Sidgwick & Jackson, London.

Elton, C. (1933) *The Ecology of Animals*. Methuen, London.

Elton, C.S. (1958) *The Ecology of Invasions by Animals and Plants*. Methuen, London.

Elwood, J.W., Newbold, J.D., O'Neill, R.V. & van Winkle, W. (1983) Resource spiralling: an operational paradigm for analyzing lotic ecosystems. In: *Dynamics of Lotic Ecosystems* (T.D. Fontaine & S.M. Bartell, eds), pp. 3–28. Ann Arbor Science Publishers, Ann Arbor, MI.

Emiliani, C. (1966) Isotopic palaeotemperatures. *Science*, 154, 851–857.

Endara, M-J., Coley, P.D., Ghabash, G. *et al.* (2017) Coevolutionary arms race versus host defense chase in a tropical herbivore–plant system. *Proceedings of the National Academy of Sciences of the USA*, 114, E7499–E7505.

Enever, R., Lewin, S., Reese, A. & Hooper, T. (2017) Mapping fishing effort: combining fishermen's knowledge with satellite monitoring data in English waters. *Fisheries Research*, 189, 67–76.

Engel, P. & Moran, N.A. (2013) The gut microbiota of insects – diversity in structure and function. *FEMS Microbiology Reviews*, 37, 699–735.

Enquist, B.J., Brown, J.H. & West, G.B. (1998) Allometric scaling of plant energetics and population density. *Nature*, 395, 163–165.

Enriquez, S., Duarte, C.M. & Sand-Jensen, K. (1993) Patterns in decomposition rates among photosynthetic organisms: the importance of detritus C : N : P content. *Oecologia*, 94, 457–471.

Erath, W., Bauer, E., Fowler, D.B. *et al.* (2017) Exploring new alleles for frost tolerance in winter rye. *Theoretical and Applied Genetics*, 130, 2151–2164.

Erickson, G.M., Makovicky, P.J., Inouye, B.D., Zhou, C-F & Gao, K-Q. (2009) A life table for *Psittacosaurus lujiatunensis*: initial insights into Ornithischian dinosaur population biology. *The Anatomical Record*, 292, 1514–1521.

Errington, P.L. (1934) Vulnerability of Bobwhite populations to predation. *Ecology*, 15, 110–127.

Essington, T.E., Beaudreau, A.H. & Wiedenmann, J. (2006) Fishing through marine food webs. *Proceedings of the National Academy of Sciences of the USA*, 103, 3171–3175.

Estes, J.A. (2015) Natural history, ecology, and the conservation and management of sea otters. In: *Sea Otter Conservation* (S.E. Larson, J.L. Bodkin & G.R. VanBlaricom, eds) pp. 19–41. Academic Press, London.

Ettema, C.H. & Wardle, D.A. (2002) Spatial soil ecology. *Trends in Ecology and Evolution*, 17, 177–183.

EU (European Union) (2013) Regulation (EU) No 1380/2013 of the European Parliament and of the Council of 11 December 2013 on the Common Fisheries Policy. Official Journal of the European Union.

European Commission (2018) Environment, Nature and Biodiversity: ecosystem services and green infrastructure. http://ec.europa.eu/environment/nature/ecosystems/index_en.htm (last accessed November 2019).

European Environment Agency (2002) Europe's Environment: the third assessment. Environmental assessment report No 10. 242 pp. Office for Official Publications of the European Communities, Luxembourg.

Ezenwa, V.O. & Jolles, A.E. (2015) Opposite effects of anthelmintic treatment on microbial infection at individual versus population scales. *Science*, 347, 175–177.

Fajer, E.D. (1989) The effects of enriched CO_2 atmospheres on plant– insect–herbivore interactions: growth responses of larvae of the specialist butterfly, *Junonia coenia* (Lepidoptera: Nymphalidae). *Oecologia*, 81, 514–520.

Falge, E., Baldocchi, D., Tenhunen, J. *et al.* (2002) Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements. *Agricultural and Forest Meteorology*, 113, 53–74.

FAO (2012) 3 December. Serious desert locust threat extends to NW and NE Africa. Desert Locust Briefs. http://www.fao.org/ag/locusts/en/archives/briefs/1810/1932/index.html (last accessed November 2019).

FAO (2016) *Climate Change and Food Security: risks and responses*. Food and Agriculture Organization of the United Nations, Rome.

Farinas, T.H., Bacher, C., Soudant, D., Belin, C. & Barille, L. (2015) Assessing phytoplankton realized niches using a French national phytoplankton monitoring network. *Estuarine, Coastal and Shelf Science*, 159, 15–27.

Fasham, M.J.R., Balino, B.M. & Bowles, M.C. (2001) A new vision of ocean biogeochemistry after a decade of the Joint Global Ocean Flux Study (JGOFS). *Ambio Special Report*, 10, 4–31.

Feeny, P. (1976) Plant apparency and chemical defence. *Recent Advances in Phytochemistry*, 10, 1–40.

Feller, C., Brown, G.G., Blanchart, E., Deleporte, P. & Chernyanskii, S.S. (2003) Charles Darwin, earthworms and the natural sciences: various lessons from past to future. *Agriculture, Ecosystems and Environment*, 99, 29–49.

Fei, T., Skidmore, A.K., Venus, V. *et al.* (2012) A body temperature model for lizards as estimated from the thermal environment. *Journal of Thermal Biology*, 37, 56–64.

Fenchel, T. (1975) Character displacement and coexistence in mud snails (Hydrobiidae) *Oecologia*, 20, 19–32.

Fenchel, T. (1987) *Ecology – Potentials and Limitations*. Ecology Institute, Federal Republic of Germany.

Fenner, F. (1983) Biological control, as exemplified by smallpox eradication and myxomatosis. *Proceedings of the Royal Society of London, Series B*, 218, 259–285.

Fenner, F. & Ratcliffe, R.N. (1965) *Myxomatosis*. Cambridge University Press, London.

Ferrol, N., Tamayo, E. & Vargas, P. (2016) The heavy metal paradox in arbuscular mycorrhizas: from mechanisms to biotechnological applications. *Journal of Experimental Botany*, 67, 6253–6265.

Field, R., O'Brien, E.M. & Whittaker, R.J. (2005) Global models for predicting woody plant richness from climate: development and evaluation. *Ecology*, 86, 2263–2277.

Field, R.D. & Reynolds, J.D. (2013) Ecological links between salmon, large carnivore predation, and scavenging birds. *Journal of Avian Biology*, 44, 9–16.

Fieldler, P.L. (1987) Life history and population dynamics of rare and common Mariposa lilies (*Calochortus* Pursh: Liliaceae). *Journal of Ecology*, 75, 977–995.

Fine, P.V.A., Mesones, I. & Coley, P.D. (2004) Herbivores promote habitat specialization by trees in Amazonian forests. *Science*, 305, 663–665.

Finlay, K.J. & Luck, J.E. (2011) Response of the bird cherry-oat aphid (*Rhopalosiphum padi*) to climate change in relation to its pest status, vectoring potential and function in a crop–vector–virus pathosystem. *Agriculture, Ecosystems and Environment*, 144, 405–421.

Firbank, L.G. & Watkinson, A.R. (1990) On the effects of competition: from monocultures to mixtures. In: *Perspectives on Plant Competition* (J.B. Grace & D. Tilman, eds), pp. 165–192. Academic Press, New York.

Fisher, J. (1954) Evolution and bird sociality. In: *Evolution as a Process* (J. Huxley, A.C. Hardy & E.B. Ford, eds), pp. 71–83. Allen & Unwin, London.

Fisher, M.C., Garner, T.W.J. & Walker, S.F. (2009) Global emergence of *Batrochochytrium dendrobatidis* and amphibian chytridiomycosis is space, time and host. *Annual Review of Microbiology*, 63, 291–310.

Fisher, R.A. (1930) *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Fitzpatrick, B.M., Fordyce, J.A. & Gavrilets, S. (2008) What, if anything, is sympatric speciation? *Journal of Evolutionary Biology*, 21, 1452–1459.

Fitzsimons, M.S., Miller, R.M. & Jastrow, J.D. (2008) Scale-dependent niche axes of arbuscular mycorrhizal fungi. *Oecologia*, 158, 117–127.

Flessa, K.W. & Jablonski, D. (1995) Biogeography of recent marine bivalve mollusks and its implications of paleobiogeography and the geography of extinction: a progress report. *Historical Biology*, 10, 25–47.

Flett, R.J., Schindler, D.W., Hamilton, R.D. & Campbell, E.R. (1980) Nitrogen fixation in Canadian Precambrian Shield lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 37, 494–505.

Flint, M.L. & van den Bosch, R. (1981) *Introduction to Integrated Pest Management*. Plenum Press, New York.

Flockhart, D.T.T., Martin, T.G. & Norris, D.R. (2012) Experimental examination of intraspecific density-dependent competition during the breeding period in Monarch butterflies (*Danaus plexippus*). *PLoS One*, 7, e45080.

Ford, E.B. (1940) Polymorphism and taxonomy. In: *The New Systematics* (J. Huxley, ed.), pp. 493–513. Clarendon Press, Oxford.

Foster, K.R. & Bell, T. (2012) Competition, not cooperation, dominates interactions among culturable microbial species. *Current Biology*, 22, 1845–1850.

Fowler, S.V. & Lawton, J.H. (1985) Rapidly induced defenses and talking trees: the devil's advocate position. *American Naturalist*, 126, 181–195.

Fox, N.J. & Beckley, L.E. (2005) Priority areas for conservation of Western Australian coastal fishes: a comparison of hotspot, biogeographical and complementarity approaches. *Biological Conservation*, 125, 399–410.

Franche, C., Lindström, K. & Elmerich, C. (2009) Nitrogen-fixing bacteria associated with leguminous and non-leguminous plants. *Plant and Soil*, 321, 35–59.

Frank, K.T., Petrie, B., Choi, J.S. & Leggett, W.C. (2005) Trophic cascades in a formerly cod-dominated ecosystem. *Science*, 308, 1621–1623.

Frankham, R., Bradshaw, C.J.A. & Brook, B.W. (2014) Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, 170, 56–63.

Franklin, I.A. (1980) Evolutionary change in small populations. In: *Conservation Biology, an Evolutionary–Ecological Perspective* (M.E. Soulé & B.A. Wilcox, eds), pp. 135–149. Sinauer Associates, Sunderland, MA.

Franzen, M. & Nilsson, M. (2010) Both population size and patch quality affect local extinctions and colonizations. *Proceedings of the Royal Society of London, Series B*, 277, 79–85.

Freckleton, R.P., Harvey, P.H. & Pagel, M. (2002) Phylogenetic analysis and comparative data: a test and review of evidence. *American Naturalist*, 160, 712–726.

Fredrickson, M.E. (2005) Ant species confer different partner benefits on two neotropical myrmecophytes. *Oecologia*, 143, 387–395.

Fredrickson, M.E., Greene, M.J. & Gordon, D.M. (2005) ‘Devil's gardens’ bedevilled by ants. *Nature*, 437, 495–496.

Free, C.A., Beddington, J.R. & Lawton, J.H. (1977) On the inadequacy of simple models of mutual interference for parasitism and predation. *Journal of Animal Ecology*, 46, 543–554.

Fretwell, S.D. & Lucas, H.L. (1970) On territorial behaviour and other factors influencing habitat distribution in birds. *Acta Biotheoretica*, 19, 16–36.

Frew, A., Allsopp, P.G., Gherlenda, A.N. & Johnson, S.N. (2017) Increased root herbivory under elevated atmospheric carbon dioxide concentrations is reversed by silicon-based plant defences. *Journal of Applied Ecology*, 54, 1310–1319.

Froese, R., Winker, H., Gascuel, D., Sumaila, U.R. & Pauly, D. (2016) Minimizing the impact of fishing. *Fish and Fisheries*, 17, 785–802.

Fuchs, J., Ericson, P.G.P., Bonillo, C., Couloux, A. & Pasquet, E. (2015) The complex phylogeography of the Into-Malayan Alophoixus bulbuls with the description of a putative new ring species complex. *Molecular Ecology*, 24, 5460–5474.

Fugère, V., Andino, P., Espinosa, R., Anthelme, F. Jacobsen, D. & Dangles, O. (2012) Testing the stress-gradient hypothesis with aquatic detritivorous invertebrates: insights for biodiversity-ecosystem functioning research. *Journal of Animal Ecology*, 81, 1259–1267.

Fujiwara, T., Septiana, R.M., Awang, S.A. *et al.* (2012) Changes in local social economy and forest management through the introduction of collaborative forest management (PHBM), and the challenges it poses on equitable partnership: a case study of KPH Pemalang, Central Java, Indonesia. *Tropics*, 20, 115–134.

Fukami, T. (2001) Sequence effects of disturbance on community structure. *Oikos*, 92, 215–224.

Fukami, T. (2015) Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution and Systematics*, 46, 1–23.

Gaillard, J-M., Nilsen, E.B., Odden, J., Andren, H. & Linnell, J.D.C. (2014) One size fits all: Eurasian lynx females share a common optimal litter size. *Journal of Animal Ecology*, 83, 107–115.

Gaines, M.S., Vivas, A.M. & Baker, C.L. (1979) An experimental analysis of dispersal in fluctuating vole populations: demographic parameters. *Ecology*, 60, 814–828.

Galligan, T.H., Amano, T., Prakash, V.M. *et al.* (2014) Have population declines in Egyptian vulture and red-headed vulture in India slowed since the 2006 ban on veterinary diclofenac? *Bird Conservation International*, 24, 272–281.

Galloway, J.N., Dentener, F.J., Capone, D.G., Boyer, E.W. & Howarth, R.W. (2004) Nitrogen cycles: past, present, and future. *Biogeochemistry*, 70, 153–226.

Galloway, L.F. & Fenster, C.B. (2000) Population differentiation in an annual legume: local adaptation. *Evolution*, 54, 1173–1181.

Game, E.T., Grantham, H.S., Hobday, A.J. *et al.* (2009) Pelagic protected areas: the missing dimension in ocean conservation. *Trends in Ecology and Evolution*, 24, 360–369.

Ganade, G. & Brown, V.K. (2002) Succession in old pastures of central Amazonia: role of soil fertility and plant litter. *Ecology*, 83, 743–754.

Gandon, S. & Michalakis, Y. (2001) Multiple causes of the evolution of dispersal. In: *Dispersal* (J. Clobert, E. Danchin, A.A. Dhondt & J.D. Nichols, eds), pp. 155–167. Oxford University Press, Oxford.

Garcia, K. & Zimmermann, S.D. (2014) The role of mycorrhizal associations in plant potassium nutrition. *Frontiers in Plant Science*, 5, 337.

Gaston, K.J. & Blackburn, T.M. (2000) *Pattern and Process in Macroecology*. Blackwell Science, Oxford.

Gauthier, G., Berteaux, D., Krebs, C.J. & Reid, D. (2009) Arctic lemmings are not simply food limited – a comment on Oksanen *et al. Evolutionary Ecology Research*, 11, 483–484.

Gavloski, J.E. & Lamb, R.J. (2000) Compensation for herbivory in cruciferous plants: specific responses in three defoliating insects. *Environmental Entomology*, 29, 1258–1267.

GBD 2013 Mortality and Causes of Death Collaborators (2015) Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, 385, 117–171.

Gbondo-Tugbawa, S.S., Driscoll, C.T., Aber, J.D. & Likens, G.E. (2001) Evaluation of an integrated biogeochemical model (PnET-BGC) at a northern hardwood forest ecosystem. *Water Resources Research*, 37, 1057–1070.

Geange, S.W. & Stier, A.C. (2009) Order of arrival affects competition in two reef fishes. *Ecology*, 90, 2868–2878.

Geider, R.J., Delucia, E.H., Falkowski, P.G. *et al.* (2001) Primary productivity of planet earth: biological determinants and physical constraints in terrestrial and aquatic habitats. *Global Change Biology*, 7, 849–882.

Geisz, H.N., Dickhut, R.M., Cochran, M.A., Fraser, W.R. & Ducklow, H.W. (2008) Melting glaciers: a probable source of DDT to the Antarctic marine ecosystem. *Environmental Science and Technology*, 42, 3958–3962.

Gende, S.M., Quinn, T.P. & Willson, M.F. (2001) Consumption choice by bears feeding on salmon. *Oecologia*, 127, 372–382.

Gerea, M., Perez, G.L., Unrein, F., Cardenas, C.S., Morris. D. & Queimalinos, C. (2017) CDOM and the underwater light climate in two shallow North Patagonian lakes: evaluating the effects on nano and microphytoplankton community structure. *Aquatic Sciences*, 79, 231–248.

Gilbert, K.J. & Whitlock, M.C. (2015) Evaluating methods for estimating local effective population size with and without migration. *Evolution*, 69, 2154–2166.

Gilg, O., Hanski, I. & Sittler, B. (2003) Cyclic dynamics in a simple vertebrate predator–prey community. *Science*, 302, 866–868.

Gill, J.A., Norris, K. & Sutherland, W.J. (2001) The effects of disturbance on habitat use by black-tailed godwits *Limosa limosa*. *Journal of Applied Ecology*, 38, 846–856.

Gillaranz, L.J. & Bascompte, J. (2012) Spatial network structure and metapopulation persistence. *Journal of Theoretical Biology*, 297, 11–16.

Gilligan, C.A. (1990) Comparison of disease progress curves. *New Phytologist*, 115, 223–242.

Gillman, L.N., Wright, S.D., Cusens, J., McBride, P.D., Malhi, Y. & Whittaker, R.J. (2015) Latitude, productivity and species richness. *Global Ecology and Biogeography*, 24, 107–117.

Gillooly, J.F., Brown, J.H., West, G.B., Savage, V.M. & Charnov, E.L. (2001) Effects of size and temperature on metabolic rate. *Science*, 293, 2248–2251.

Ginzburg, L.R. & Krebs, C.J (2015), Mammalian cycles: internally defined periods and interaction-driven amplitudes. *Peer Journal*, 3, e1180.

Glawe, G.A., Zavala, J.A., Kessler, A., Van Dam, N.M. & Baldwin, I.T. (2003) Ecological costs and benefits correlated with trypsin proteinase inhibitor production in *Nicotinia attenuata*. *Ecology*, 84, 79–90.

Glazier, D.S. (2005) Beyond the '3/4-power law': variation in the intra- and interspecific scaling of metabolic rate in animals. *Biological Reviews*, 80, 611–662.

Glazier, D.S. (2010) A unifying explanation for diverse metabolic scaling in animals and plants. *Biological Reviews*, 85, 111–138.

Glazier, D.S. (2014) Metabolic scaling in complex living systems. *Systems*, 2, 451–440.

Gleason, H.A. (1926) The individualistic concept of the plant association. *Torrey Botanical Club Bulletin*, 53, 7–26.

Glikson, A. (2016) Cenozoic mean greenhouse gases and temperature changes with reference to the Anthropocene. *Global Change Biology*, 22, 3843–3858.

Głowacka, K., Ahmed, A., Sharma, S. *et al.* (2016) Can chilling tolerance of C_4 photosynthesis in Miscanthus be transferred to sugarcane? *Global Change Biology: Bioenergy*, 8, 407–418.

Godfray, H.C.J. (1994) *Parasitoids: behavioral and evolutionary ecology*. Princeton University Press, Princeton, NJ.

Goffredi, S.K., Orphan, V.J., Rouse, G.W. *et al.* (2005) Evolutionary innovation: a bone-eating marine symbiosis. *Environmental Microbiology*, 7, 1369–1378.

Goldenberg, S.U., Nagelkerken, I., Ferreira, C.M., Ullah, H. & Connell, S.D. (2017) Boosted food web productivity through ocean acidification collapses under warming. *Global Change Biology*, 23, 4177–4184.

Goldson, S.L., Barron, M.C., Kean, J.M. & van Koten, C. (2011) Argentine stem weevil (*Listronotus bonariensis*, Coleoptera: Curculionidae) population dynamics in Canterbury, New Zealand dryland pasture. *Bulletin of Entomological Research*, 101, 295–303.

Gómez-Aparicio, L., Zamora, R., Gómez, J.M., Hódar, J.A., Castro, J. & Baraza, E. (2004) Applying plant facilitation to forest restoration: a meta-analysis of the use of shrubs as nurse plants. *Ecological Applications*, 14, 1128–1138.

Gonzalez, A. (2009) Metacommunities: spatial community ecology. In: *Encyclopedia of Life Sciences*. John Wiley & Sons, Chichester.

Gonzalez-Andujar, J.L., Fernandez-Quintanilla, C. & Navarrete, L. (2006) Population cycles produced by delayed density dependence in an annual plant. *American Naturalist*, 168, 318–322.

González-Olivares, E., González-Yañez, B., Becerra-Klix, R. & Ramos-Jiliberto, R. (2017) Multiple stable states in a model based on predator-induced defenses. *Ecological Complexity*, 32, 111–120.

Gorman, M.L. (1979) *Island Ecology*. Chapman & Hall, London.

Gosper, C.R., Yates, C.J. & Prober, S.M. (2012) Changes in plant species and functional composition with time since fire in two Mediterranean climate plant communities. *Journal of Vegetation Science*, 23, 1071–1081.

Gotelli, N.J. & Colwell, R.K. (2011) Estimating species richness. In: *Frontiers in Measuring Biodiversity* (A.E. Magurran & B.J. McGill, eds), pp. 39–54. Oxford University Press, Oxford.

Gotelli, N.J. & McCabe, D.J. (2002) Species co-occurrence: a metaanalysis of J.M. Diamond's assembly rules model. *Ecology*, 83, 2091–2096.

Gotelli, N.J. & Ulrich, W. (2012) Statistical challenges in null model analysis. *Oikos*, 121, 171–180.

Götzenberger, L., de Bello, F., Brathen, K.A. *et al.* (2012) Ecological assembly rules in plant communities – approaches, patterns and prospects. *Biological Reviews*, 87, 111–127.

Gough, C.M. (2011) Terrestrial primary production: fuel for life. *Nature Education Knowledge*, 3, 28–34.

Gough, L. Shaver, G.R., Carroll, J., Royer, D.L. & Laundre, J.A. (2000) Vascular plant species richness in Alaskan arctic tundra: the importance of soil pH. *Journal of Ecology*, 88, 54–66.

Gouveia, G.R., Trindade, G.S., Maia Nery, L.E. & Muelbert, J.H. (2015) UVA and UVB penetration in the water column of a south west atlantic warm temperate estuary and its effects on cells and fish larvae. *Estuaries and Coasts*, 38, 1147–1162.

Graff, P., Aguiar, M.R. & Chaneton, E.J. (2007) Shifts in positive and negative plant interactions along a grazing intensity gradient. *Ecology*, 88, 188–199.

Graham, A.L. (2008) Ecological rules governing helminth-microparasite coinfection. *Proceedings of the National Academy of Sciences of the USA*, 105, 566–570.

Graham, C.H., Moritz, C. & Williams, S.E. (2006) Habitat history improves prediction of biodiversity in rainforest fauna. *Proceedings of the National Academy of Sciences of the USA*, 103, 632–636.

Graham, I. & Lambin, X. (2002) The impact of weasel predation on cyclic field-vole survival: the specialist predator hypothesis contradicted. *Journal of Animal Ecology*, 71, 946–956.

Grant, B.S., Cook, A.D., Clarke, C.A. & Owen, D.F. (1998) Geographic and temporal variation in the incidence of melanism in peppered moth populations in America and Britain. *The Journal of Heredity*, 89, 465–471.

Grant, P.R. (2018) Cold tolerance allows a selected few anole lizards to survive a cold snap. *Science*, 357, 451–452.

Grant, P.R. & Grant, B.R. (2010) Conspecific versus heterospecific gene exchange between populations of Darwin's finches. *Philosophical Transactions of the Royal Society of London.Series B*, 365, 1065–1076.

Grant, P.R., Grant, R., Keller, L.F. & Petren, K. (2000) Effects of El Nino events on Darwin's finch productivity. *Ecology*, 81, 2442–2457.

Grassberger, M. & Reiter, C. (2002) Effect of temperature on development of the forensically important holarctic blowfly *Protophormia terraenovae* (Robineau-Desvoidy) (Diptera: Calliphoridae). *Forensic Science International*, 128, 177–182.

Gray, M.E., Sappington, T.W., Miller, N.J., Moeser, J. & Bohn, M.O. (2009) Adaptation and invasiveness of western corn rootworm: intensifying research on a worsening pest. *Annual Review of Entomology*, 54, 303–321.

Gray, S.M. & Robinson, B.W. (2001) Experimental evidence that competition between stickleback species favours adaptive character divergence. *Ecology Letters*, 5, 264–272.

Greene, D.F. & Calogeropoulos, C. (2001) Measuring and modelling seed dispersal of terrestrial plants. In: *Dispersal Ecology* (J.M. Bullock, R.E. Kenward & R.S. Hails, eds), pp. 3–23. Blackwell Science, Oxford.

Greenwood, P.J. (1980) Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*, 28, 1140–1162.

Greenwood, P.J., Harvey, P.H. & Perrins, C.M. (1978) Inbreeding and dispersal in the great tit. *Nature*, 271, 52–54.

Griffith, D.M, Anderson, T.M., Osborne, C.P., Strömberg, C.A.E., Forrestel, E.J. & Still, C.J. (2015) Biogeographically distinct controls on C_3 and C_4 grass distributions: merging community and physiological ecology. *Global Ecology and Biogeography*, 24, 304–313.

Griffith, D.M. & Poulson, T.M. (1993) Mechanisms and consequences of intraspecific competition in a carabid cave beetle. *Ecology*, 74, 1373–1383.

Griffiths, H., Meyer, M.T. & Rickaby, R.E.M. (2017) Overcoming adversity through diversity: aquatic carbon concentrating mechanisms. *Journal of Experimental Botany*, 68, 3689–3695.

Griffiths, R.A., Denton, J. & Wong, A.L.-C. (1993) The effect of food level on competition in tadpoles: interference mediated by protothecan algae? *Journal of Animal Ecology*, 62, 274–279.

Griffiths, S., Wingen, L., Pietragalla, J. *et al.* (2015) Genetic dissection of grain size and grain number trade-offs in CIMMYT wheat germplasm. *PLoS One*, 10, e0118847.

Grime, J.P. (1974) Vegetation classification by reference to strategies. *Nature*, 250, 26–31.

Grime, J.P. (2007) Plant strategy theories: a comment on Craine (2005) *Journal of Ecology*, 95, 227–230.

Grime, J.P. & Pierce, S. (2012) *The Evolutionary Strategies that Shape Ecosystems*. Wiley Blackwell, Oxford.

Grimes, J.E.T., Croll, D., Harrison, W.E., Utzinger, J., Freeman, M.C. & Templeton, M.R. (2015) The roles of water, sanitation and hygiene in reducing schistosomiasis: a review. *Parasites and Vectors*, 8, 156.

Gripenberg, S., Mayhew, P.J., Parnell, M. & Roslin, T. (2010) A meta-analysis of preference–performance relationships in phytophagous insects. *Ecology Letters*, 13, 383–393.

Groenteman, R., Kelly, D., Fowler, S.V. & Bourdot, G.W. (2011) Abundance, phenology and impact of biocontrol agents on nodding thistle (*Carduus nutans*) in Canterbury 35 years into a biocontrol programme. *New Zealand Journal of Agricultural Research*, 54, 1–11.

Grubb, P. (1977) The maintenance of species richness in plant communities: the importance of the regeneration niche. *Biological Reviews*, 52, 107–145.

Grubb, P.J. (1986) The ecology of establishment. In: *Ecology and Design in Landscape* (A.D. Bradshaw, D.A. Goode & E. Thorpe, eds), pp. 83–97. Symposia of the British Ecological Society, No. 24. Blackwell Scientific Publications, Oxford.

Grube, M. & Wedin, M. (2016) Lichenized fungi and the evolution of symbiotic organization. *Microbiology Spectrum*, 4. doi: 10.1128/microbiolspec.FUNK-0011-2016.

Gruber, N. & Deutsch, C.A. (2014) Redfield's evolving legacy. *Nature Geoscience*, 7, 853–855.

Grutter, A.S. (1999) Cleaner fish really do clean. *Nature*, 398, 672–673.

Grutter, A.S. & Bshary, R. (2003) Cleaner wrasse prefer client mucus: support for partner control mechanisms in cleaning interactions. *Proceedings of the Royal Society of London, Series B*, 270, S242–S244.

Grytnes, J.A. & Vetaas, O.R. (2002) Species richness and altitude: a comparison between null models and interpolated plant species richness along the Himalayan altitudinal gradient, Nepal. *American Naturalist*, 159, 294–304.

Guedes, R.N.C. & Cutler, G.C. (2014) Insecticide-induced hormesis and arthropod pest management. *Pest Management Science*, 70, 690–697.

Gündüz, E.A. & Douglas, A.E. (2009) Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proceedings of the Royal Society of London, Series B*, 276, 987–991.

Guiñez, R. & Castilla, J.C. (2001) An allometric tridimensional model of self-thinning for a gregarious tunicate. *Ecology*, 82, 2331–2341.

Gurr, G. & Wrattem, S. (2010) *Biological Control: Measures of Success*. Springer-Science, Dordrecht.

Haberl, H., Erb, K-H. & Krausmann, F. (2014) Human appropriation of net primary production: patterns, trends, and planetary boundaries. *Annual Review of Environment and Resources*, 39, 363–391.

Haddon, M. (2002) *Modelling and Quantitative Methods in Fisheries*, 2nd Edition. CRC Press, Boca Raton, FL.

Haefner, P.A. (1970) The effect of low dissolved oxygen concentrations on temperature–salinity tolerance of the sand shrimp, *Crangon septemspinosa*. *Physiological Zoology*, 43, 30–37.

Hainsworth, F.R. (1981) *Animal Physiology*. Addison-Wesley, Reading, MA.

Hairston, N.G., Smith, F.E. & Slobodkin, L.B. (1960) Community structure, population control, and competition. *American Naturalist*, 44, 421–425.

Haldane, J.B.S. (1949) Disease and evolution. *La Ricerca Scienza*, 19 (Suppl.), 3–11.

Hall, S.J. & Raffaelli, D.G. (1993) Food webs: theory and reality. *Advances in Ecological Research*, 24, 187–239.

Hamel, S., Côté, S.D., Gaillard, J-M. & Festa-Bianchet, M. (2009) Individual variation in reproductive costs of reproduction: high-quality females always do better. *Journal of Animal Ecology*, 78, 143–151.

Hamilton, W.D. (1971) Geometry for the selfish herd. *Journal of Theoretical Biology*, 31, 295–311.

Hamilton, W.D. & May, R.M. (1977) Dispersal in stable habitats. *Nature*, 269, 578–581.

Hampson, K., Dushoff, J., Cleaveland, S. *et al.* (2009) Transmission dynamics and prospects for the elimination of canine rabies. *PLoS Biology*, 7, e1000053.

Hansen, J., Ruedy, R., Glasgoe, J. & Sato, M. (1999) GISS analysis of surface temperature change. *Journal of Geophysical Research*, 104, 30997–31022.

Hansen, M.C., Stehman, S.V. & Potapov, P.V. (2010) Quantification of global gross forest cover loss. *Proceedings of the National Academy of Sciences of the USA*, 107, 8650–8655.

Hanski, I. (1991) Single-species metapopulation dynamics: concepts, models and observations. In: *Metapopulation Dynamics* (M.E. Gilpin & I. Hanski, eds), pp. 17–38. Academic Press, London.

Hanski, I. (1999) *Metapopulation Ecology*. Oxford University Press, Oxford.

Hanski, I. & Gyllenberg, M. (1993) Two general metapopulation models and the core-satellite hypothesis. *American Naturalist*, 142, 17–41.

Hanski, I., Hansson, L. & Henttonen, H. (1991) Specialist predators, generalist predators, and the microtine rodent cycle. *Journal of Animal Ecology*, 60, 353–367.

Hanski, I. & Henttonen, H. (2002) Population cycles of small rodents in Fennoscandia. In: *Population Cycles: The Case for Trophic Interactions* (A.A. Berryman, ed.), pp. 44–68. Oxford University Press, Oxford.

Hanski, I., Henttonen, H., Korpimaki, E., Oksanen, L. & Turchin, P. (2001) Small-rodent dynamics and predation. *Ecology*, 82, 1505–1520.

Hanski, I. & Ovaskainen, O. (2000) The metapopulation capacity of a fragmented landscape. *Nature*, 404, 755–758.

Hanski, I., Pakkala, T., Kuussaari, M. & Lei, G. (1995) Metapopulation persistence of an endangered butterfly in a fragmented landscape. *Oikos*, 72, 21–28.

Hanski, I., Schulz, T., Wong, S.C., Ahola, V., Ruokolainen, A. & Ojanen, S.P. (2017) Ecological and genetic basis of metapopulation persistence of the Glanville fritillary butterfly in fragmented landscapes. *Nature Communications*, 81, 14504.

Härkönen, T., Dietz, R., Reijnders, P. *et al.* (2006) A review of the 1988 and 2002 phocine distemper virus epidemics in European harbour seals. *Diseases of Aquatic Organisms*, 68, 115–130.

Harper, D.G.C. (1982) Competitive foraging in mallards: 'ideal free ducks'. *Animal Behaviour*, 30, 575–584.

Harper, J.L. (1955) The influence of the environment on seed and seedling mortality. VI. The effects of the interaction of soil moisture content and temperature on the mortality of maize grains. *Annals of Applied Biology*, 43, 696–708.

Harper, J.L. (1977) *The Population Biology of Plants*. Academic Press, London.

Harper, J.L., Rosen, R.B. & White, J. (eds) (1986) The growth and form of modular organisms. *Philosophical Transactions of the Royal Society of London, Series B*, 313, 1–250.

Harris, G. & Pimm, S.L. (2008) Range size and extinction risk in forest birds. *Conservation Biology*, 22, 163–171.

Harris, G.P. (2001) Biogeochemistry of nitrogen and phosphorus in Australian catchments, rivers and estuaries: effects of land use and flow regulation and comparisons with global patterns. *Marine and Freshwater Research*, 52, 139–149.

Harrison, J.F. (2017) Do performance–safety tradeoffs cause hypometric metabolic scaling in animals? *Trends in Ecology and Evolution*, 32, 653–664.

Harrison, S. & Taylor, A.D. (1997) Empirical evidence for metapopulations. In: *Metapopulation Biology* (I.A. Hanski & M.E. Gilpin, eds), pp. 27–42. Academic Press, San Diego, CA.

Harsch, M.A., Phillips, A., Zhou, Y., Leung, M-R., Rinnan, D.S. & Kot, M. (2017) Moving forward: insights and applications of moving-habitat models for climate change ecology. *Journal of Ecology*, 105, 1169–1181.

Hart, A.J., Bale, J.S., Tullett, A.G., Worland, M.R. & Walters, K.F.A. (2002) Effects of temperature on the establishment potential of the predatory mite *Amblyseius californicus* McGregor (Acari: Phytoseiidae) in the UK. *Journal of Insect Physiology*, 48, 593–599.

Hart, S.P. & Marshall, D.J. (2009) Spatial arrangement affects population dynamics and competition independent of community composition. *Ecology*, 90, 1485–1491.

Hartley, S.E. & DeGabriel, J.L. (2016) The ecology of herbivore-induced silicon defences in grasses. *Functional Ecology*, 30, 1311–1322.

Hartley, S.E. & Jones, C.G. (1997) Plant chemistry and herbivory; or why the world is green. In: *Plant Ecology*, 2nd edn (M.J. Crawley, ed.), pp 284–324. Wiley Blackwell, Oxford.

Hartmann, T. (2007) From waste products to ecochemicals: fifty years research of plant secondary metabolism. *Phytochemistry*, 68, 2831–2846.

Harvey, P.H. & Pagel, M.D. (1991) *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford.

Hassell, M.P. (1978) *The Dynamics of Arthropod Predator–Prey Systems*. Princeton University Press, Princeton, NJ.

Hassell, M.P. (1982) Patterns of parasitism by insect parasitoids in patchy environments. *Ecological Entomology*, 7, 365–377.

Hassell, M.P. (1985) Insect natural enemies as regulating factors. *Journal of Animal Ecology*, 54, 323–334.

Hastings, A., Hom, C.L., Ellner, S., Turchin, P. & Godfray, H.C.J. (1993) Chaos in ecology: is mother nature a strange attractor? *Annual Review of Ecology and Systematics*, 24, 1–33.

Hatchwell, B.J. (2009) The evolution of cooperative breeding in birds: kinship, dispersal and life history. *Philosophical Transactions of the Royal Society of London, Series B*, 364, 3217–3227.

Hatton, I.A., McCann, K.S., Fryxell, J.M. *et al.* (2015) The predator-prey power law: biomass scaling across terrestrial and aquatic biomes. *Science*, 349(6252), aac6284.

Hawke, D.J. & Vallance, J.R. (2015) Microbial carbon concentration in samples of seabird and non-seabird forest soil: implications for leaf litter cycling. *Pedobiologia*, 58, 33–39.

Hawkins, B.A. & Cornell, H.V. (eds) (1999) *Theoretical Approaches to Biological Control*. Cambridge University Press, Cambridge, UK.

Hawkins, B.A., Field, R., Cornell, H.V. *et al.* (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, 84, 3105–3117.

Hay, M.E. & Taylor, P.R (1985) Competition between herbivorous fishes and urchins on Caribbean reefs. *Oecologia*, 65, 591–598.

Hay, W.T., Bihmidine, S., Mutluc, N. *et al.* (2017) Enhancing soybean photosynthetic CO_2 assimilation using a cyanobacterial membrane protein, ictB. *Journal of Plant Physiology*, 212, 58–68.

He, Q., Bertness, M.D. & Altieri, A.H. (2013) Global shifts towards positive species interactions with increasing environmental stress. *Ecology Letters*, 16, 695–706.

Heal, O.W. & MacLean, S.F. (1975) Comparative productivity in ecosystems – secondary productivity. In: *Unifying Concepts in Ecology* (W.H. van Dobben & R. H. Lowe–McConnell, eds), pp. 89–108. Junk, The Hague.

Heap, I. (2018) Chronological increase in resistant weeds globally. http://weedscience.org/Graphs/ChronologicalIncrease.aspx (last accessed June 2020).

Hector, A., Hautier, Y., Saner, P. *et al.* (2010) General stabilizing effects of plant diversity on grassland productivity through population asynchrony and overyielding. *Ecology*, 91, 2213–2220.

Hector, A., Shmid, B., Beierkuhnlein, C. *et al.* (1999) Plant diversity and productivity experiments in European grasslands. *Science*, 286, 1123–1127.

Heed, W.B. (1968) Ecology of Hawaiian Drosophiladae. *University of Texas Publications*, 6861, 387–419.

Henderson, P.A. & Southwood, T.R.E. (2016) *Ecological Methods*, 4th edn. Wiley Blackwell, Oxford.

Henttonen, H. & Kaikusalo, A. (1993) Lemming movements. In: *The Biology of Lemmings* (N.C. Stenseth & R.A. Ims, eds), pp. 157–186. Academic Press, London.

Hereford, J. (2009) A quantitative survey of local adaptation and fitness trade-offs. *American Naturalist*, 173, 579–588.

Herms, D.A. & Mattson, W.J. (1992) The dilemma of plants: to grow or defend. *Quarterly Review of Biology*, 67, 283–335.

Hernandez-Agreda, A., Gates, R.D. & Ainsworth, T.D. (2016) Defining the core microbiome in corals' microbial soup. *Trends in Microbiology*, 25, 125–140.

Hernandez-Ordonez, O., Urbina-Cardona, N. & Martınez-Ramos, M. (2015) Recovery of amphibian and reptile assemblages during old-field succession of tropical rain forests. *Biotropica*, 47, 377–388.

Herre, E.A. & West, S.A. (1997) Conflict of interest in a mutualism: documenting the elusive fig wasp–seed trade-off. *Proceedings of the Royal Society of London, Series B*, 264, 1501–1507.

Hessen, D.O., Elser, J.J., Sterner, R.W. & Urabe, J. (2013) Ecological stoichiometry: an elementary approach using basic principles. *Limnology and Oceanography*, 58, 2219–2236.

Hestbeck, J.B. (1982) Population regulation of cyclic mammals: the social fence hypothesis. *Oikos*, 39, 157–163.

Hilborn, R. & Walters, C.J. (1992) *Quantitative Fisheries Stock Assessment*. Chapman & Hall, New York.

Hildrew, A.G. (2018) Freshwater Acidification: natural history, ecology and environmental policy. Excellence in Ecology, Volume 27. Ecology Institute, Oldendorf/Luhe, Germany.

Hildrew, A.G. & Townsend, C.R. (1980) Aggregation, interference and the foraging by larvae of *Plectrocnemia conspersa* (Trichoptera: Polycentropodidae). *Animal Behaviour*, 28, 553–560.

Hilker, T., Coops, N.C., Wulder, M.A., Black, T.A. & Guy, R.D. (2008) The use of remote sensing in light use efficiency based models of gross primary production: a review of current status and future requirements. *Science of the Total Environment*, 404, 411–423.

Hill, M. (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology*, 54, 427–432.

Hill, M.F., Witman, J.D. & Caswell, H. (2002) Spatio-temporal variation in Markov chain models of subtidal community succession. *Ecology Letters*, 5, 665–675.

Hill, W.R., Mulholland, P.J. & Marzolf, E.R. (2001) Stream ecosystem responses to forest leaf emergence in spring. *Ecology*, 82, 2306–2319.

Hilton, J., O'Hare, M., Bowes, M.J. & Jones, I. (2006) How green is my river? A new paradigm of eutrophication in rivers. *Science of the Total Environment*, 365, 66–83.

Hirakawa, H. (2001) Coprophagy in leporids and other mammalian herbivores. *Mammal Review*, 31, 61–80.

Hirst, A.G., Lilley, M.K.S., Glazier, D.S. & Atkinson, D. (2016) Ontogenetic body-mass scaling of nitrogen excretion relates to body surface area in diverse pelagic invertebrates. *Limnology and Oceanography*, 62, 311–319.

Hirzel, A.H., Posse, B., Oggier, P-A., Crettenand, Y., Glenz, C. & Arlettaz, R. (2004) Ecological requirements of reintroduced species and the implications for release policy: the case of the bearded vulture. *Journal of Applied Ecology*, 41, 1103–1116.

Hitchcock, J.N. & Mitrovic, S.M. (2013) Different resource limitation by carbon, nitrogen and phosphorus between base flow and high flow conditions for estuarine bacteria and phytoplankton. *Estuarine, Coastal and Shelf Science*, 135, 106–115.

Hodge, A. (2004) The plastic plant: root responses to heterogeneous supplies of nutrients. *New Phytologist*, 162, 9–24.

Hodgson, D., McDonald, J.L. & Hosken, D.J. (2015) What do you mean, 'resilient'? *Trends in Ecology and Evolution*, 30, 503–506.

Hodgson, J.A., Kunin, W.E., Thomas, C.D., Benton, T.G. & Gabriel, D. (2010) Comparing organic farming and land sparing: optimizing yield and butterfly populations at a landscape scale. *Ecology Letters*, 13, 1358–1367.

Hodgson, J.A., Moilanen, A., Wintle, B.A. & Thomas, C.D. (2011) Habitat area, quality and connectivity: striking the balance for efficient conservation. *Journal of Applied Ecology*, 48, 148–152.

Hodgson, J.A., Thomas, C.D., Wintle, B.A. & Moilanen, A. (2009) Climate change, connectivity and conservation decision making: back to basics. *Journal of Applied Ecology*, 46, 964–969.

Hodgson, J.A., Wallis, D.W., Krishna, R. & Cornell, S.J. (2016) How to manipulate landscapes to improve the potential for range expansion. *Methods in Ecology and Evolution*, 7, 1558–1566.

Hoellein, T.J., Bruesewitz, D.A. & Richardson, D.C. (2013) Revisiting Odum (1956): a synthesis of aquatic ecosystem metabolism. *Limnology and Oceanography*, 58, 2089–2100.

Hoffmann, A.A., Montgomery, B.L., Popovici, J. *et al.* (2011) Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. *Nature*, 476, 454–457.

Hogan, M.E., Veivers, P.C., Slaytor, M. & Czolij, R.T. (1988) The site of cellulose breakdown in higher termites (*Nasutitermes walkeri* and *Nasutitermes exitosus*). *Journal of Insect Physiology*, 34, 891–899.

Holbrook, S.J. & Schmitt, R.J. (2005) Growth, reproduction and survival of a tropical sea anemone (Actiniaria): benefits of hosting anemonefish. *Coral Reefs*, 24, 67–73.

Holling, C.S. (1959) Some characteristics of simple types of predation and parasitism. *Canadian Entomologist*, 91, 385–398.

Holloway, J.D. (1977) *The Lepidoptera of Norfolk Island, their Biogeography and Ecology*. Junk, The Hague.

Holm, J.A., Shugart, H.H., Van Bloem, S.J. & Larocque, G.R. (2012) Gap model development, validation, and application to succession of secondary subtropical dry forests of Puerto Rico. *Ecological Modelling*, 233, 70–82.

Holmer, M. & Storkholm, P. (2001) Sulphate reduction and sulphur cycling in lake sediments: a review. *Freshwater Biology*, 46, 431–451.

Holmgren, M. & Scheffer, M. (2010) Strong facilitation in mild environments: the stress gradient hypothesis revisited. *Journal of Ecology*, 98, 1269–1275.

Holmgren, M., Scheffer, M., Ezcurra, E., Gutierrez, J.R. & Mohren, M.J. (2001) El Nino effects on the dynamics of terrestrial ecosystems. *Trends in Ecology and Evolution*, 16, 89–94.

Holt, R.D. (1977) Predation, apparent competition and the structure of prey communities. *Theoretical Population Biology*, 12, 197–229.

Holt, R.D. (1984) Spatial heterogeneity, indirect interactions, and the coexistence of prey species. *American Naturalist*, 124, 377–406.

Holt, R.D. (1997) Community modules. In: *Multitrophic Interactions in Terrestrial Ecosystems* (A.C. Gange & V.K. Brown, eds), pp. 333–349. Blackwell Science, Oxford.

Holyoak, M. & Lawler, S.P. (1996) Persistence of an extinction-prone predator–prey interaction through metapopulation dynamics. *Ecology*, 77, 1867–1879.

Hone, D.W.E. & Benton, M.J. (2004) The evolution of large size: how does Cope'srule work? *Trends in Ecology and Evolution*, 20, 4–6.

Honnay, O., Jacquemyn, H., Van Looy, K., Vandepitte, K. & Breyne, P. (2009) Temporal and spatial genetic variation in a metapopulation of the annual *Erysimum cheiranthoides* on stony river banks. *Journal of Ecology*, 97, 131–141.

Hook, P.B. & Burke, I.C. (2000) Biogeochemistry in a shortgrass landscape: control by topography, soil texture, and microclimate. *Ecology*, 81, 2686–2703.

Hooper, D.U., Adair, E.C., Cardinale, B.J. *et al.* (2012) A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, 486, 105–108.

Hoorn, C., Wesselingh, F.P., ter Steege, H. *et al.* (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, 330, 927–931.

Horn, D.J. (1988) *Ecological Approach to Pest Management*. Elsevier, London.

Horn, H.S. (1975) Markovian processes of forest succession. In: *Ecology and Evolution of Communities* (M.L. Cody & J.M. Diamond, eds), pp. 196–213. Belknap, Cambridge, MA.

Horn, H.S. (1981) Succession. In: *Theoretical Ecology: principles and applications* (R.M. May, ed.), pp. 253–271. Blackwell Scientific Publications, Oxford.

Hotez, P.J., Brindley, P.J., Bethony, J.M., King, C.H., Pearce, E.J. & Jacobson, J. (2008) Helminth infections: the great neglected tropical diseases. *Journal of Clinical Investigation*, 118, 1311–1321.

Houghton, R.A. (2000) Interannual variability in the global carbon cycle. *Journal of Geophysical Research*, 105, 20121–20130.

Hoving, C.I., Harrison, D.J., Krohn, W.B., Jakubas, W.J. & McCollough, M.A. (2004) Canada lynx *Lynx canadensis* habitat and forest succession in northern Maine, USA. *Wildlife Biology*, 10, 285–294.

Howarth, R.W. (2002) The nitrogen cycle. In: *Encyclopedia of Global Environmental Change. The Earth System: Biological and Ecological Dimensions of Global Environmental Change, Vol. 2* (H.A. Mooney & J.G.Canadell, eds), pp. 429–435. Wiley, Chichester.

Hoy, S.R., Petty, S.J., Millon, A. *et al.* (2015) Age and sex-selective predation moderate the overall impact of predators. *Journal of Animal Ecology*, 84, 692–701.

Hoyer, M.V. & Canfield, D.E. (1994) Bird abundance and species richness on Florida lakes: influence of trophic status, lake morphology and aquatic macrophytes. *Hydrobiologia*, 297, 107–119.

Huang, S-C. & Reinhard, J. (2012) Color change from male-mimic to gynomorphic: a new aspect of signaling sexual status in damselflies (Odonata, Zygoptera). *Behavioral Ecology*, 23, 1269–1275.

Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.

Hudson, P.J., Dobson, A.P. & Newborn, D. (1992a) Do parasites make prey vulnerable to predation? Red grouse and parasites. *Journal of Animal Ecology*, 61, 681–692.

Hudson, P.J., Dobson, A.P. & Newborn, D. (1998) Prevention of population cycles by parasite removal. *Science*, 282, 2256–2258.

Hudson, P.J., Newborn, D. & Dobson, A.P. (1992b) Regulation and stability of a free-living host–parasite system: *Trichostrongylus tenuis* in red grouse. I. Monitoring and parasite reduction experiments. *Journal of Animal Ecology*, 61, 477–486.

Huerta Lwanga, E., Gertsen, H., Gooren, H. *et al.* (2016) Microplastics in the terrestrial ecosystem: implications for Lumbricus terrestris (Oligochaeta, Lumbricidae). *Environmental Science and Technology*, 50, 2685–2691.

Huffaker, C.B. (1958) Experimental studies on predation: dispersion factors and predator–prey oscillations. *Hilgardia*, 27, 343–383.

Huffaker, C.B., Shea, K.P. & Herman, S.G. (1963) Experimental studies on predation. *Hilgardia*, 34, 305–330.

Hughes, A.R., Byrnes, J.F., Kimbro, D.L. & Stachowicz, J.J. (2007) Reciprocal relationships and potential feedbacks between biodiversity and disturbance. *Ecology Letters*, 10, 849–864.

Hughes, J.E., Deegan, L.A., Peterson, B.J., Holmes, R.M. & Fry, B. (2000) Nitrogen flow through the food web in the oligohaline zone of a New England estuary. *Ecology*, 81, 433–452.

Hughes, L. (2000) Biological consequences of global warming: is the signal already apparent. *Trends in Ecology and Evolution*, 15, 56–61.

Hughes, P.W. & Simons, A.S. (2014) The continuum between semelparity and iteroparity: plastic expression of parity in response to season length manipulation in *Lobelia inflata*. *BMC Evolutionary Biology*, 14, 90.

Hughes, R.N. (1989) *A Functional Biology of Clonal Animals*. Chapman & Hall, London.

Hughes, R.N. & Griffiths, C.L. (1988) Self-thinning in barnacles and mussels: the geometry of packing. *American Naturalist*, 132, 484–491.

Hughes, T.P., Ayre, D. & Connell, J.H. (1992) The evolutionary ecology of corals. *Trends in Ecology and Evolution*, 7, 292–295.

Hughes, T.P., Linares, C., Dakos, V., van de Leemput, I.A. & van Nes, E.H. (2013) Living dangerously on borrowed time during slow, unrecognized regime shifts. *Trends in Ecology and Evolution*, 28, 149–155.

Hughes, W.O.H., Petersen, K.S., Ugelvig, L.V. *et al.* (2004) Density-dependence and within-host competition in a semelparous parasite of leaf-cutting ants. *BMC Evolutionary Biology*, 4, 45.

Hui, C. & McGeoch, M.A. (2014) Zeta diversity as a concept and metric that unifies incidence-based biodiversity patterns. *American Naturalist*, 184, 684–694.

Huisman, J. (1999) Population dynamics of light-limited phytoplankton: microcosm experiments. *Ecology*, 80, 202–210.

Huisman, J. & Weissing, F.J. (1999) Biodiversity of plankton by species oscillations and chaos. *Nature*, 402, 407–410.

Humborg, C., Blomqvist, S., Avsan, E., Bergensund, Y. & Smedberg, E. (2002) Hydrological alterations with river damming in northern Sweden: implications for weathering and river biogeochemistry. *Global Biogeochemical Cycles*, 16, 1–13.

Hummel, C, Honkoop, P. & van der Meer, J. (2011) Small is profitable: no support for the optimal foraging theory in sea stars *Asterias rubens* foraging on the blue edible mussel *Mytilus edulis*. *Estuarine, Coastal and Shelf Science*, 94, 89–92.

Hunter, M.D. & Price, P.W. (1992) Playing chutes and ladders: heterogeneity and the relative roles of bottom-up and top-down forces in natural communities. *Ecology*, 73, 724–732.

Hunter, M.L. & Yonzon, P. (1992) Altitudinal distributions of birds, mammals, people, forests, and parks in Nepal. *Conservation Biology*, 7, 420–423.

Hurwitz, I., Fieck, A. & Durvasula, R. (2012) Antimicrobial peptide delivery strategies: use of recombinant antimicrobial peptides in paratransgenic control systems. *Current Drug Targets*, 13, 1173–1180.

Husband, B.C. & Barrett, S.C.H. (1996) A metapopulation perspective in plant population biology. *Journal of Ecology*, 84, 461–469.

Hussner, A., Mettler-Altmann, T., Weber, A.P.M. & Sand-Jensen, K. (2016) Acclimation of photosynthesis to supersaturated CO_2 in aquatic plant bicarbonate users. *Freshwater Biology*, 61, 1720–1732.

Hussner, A., Stiers, I., Verhofstad, M.J.J.M. *et al.* (2017) Management and control methods of invasive alien freshwater aquatic plants: a review. *Aquatic Botany*, 136, 112–137.

Huston, M.A. & Wolverton, S. (2009) The global distribution of net primary production: resolving the paradox. *Ecological Monographs*, 79, 343–377.

Hutchings, J.A. (2014) Renaissance of a caveat: Allee effects in marine fish. *ICES Journal of Marine Science*, 71, 2152–2157.

Hutchings, M.J. (1983) Ecology's law in search of a theory. *New Scientist*, 98, 765–767.

Hutchings, M.J. & de Kroon, H. (1994) Foraging in plants: the role of morphological plasticity in resource acquisition. *Advances in Ecological Research*, 25, 159–238.

Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbour Symposium on Quantitative Biology*, 22, 415–427.

Hutchinson, G.E. (1959) Homage to Santa Rosalia, or why are there so many kinds of animals? *American Naturalist*, 93, 145–159.

Hutchinson, G.E. (1961) The paradox of the plankton. *American Naturalist*, 95, 137–145.

Iluz, D. & Dubinsky, Z. (2015) Coral photobiology: new light on old views. *Zoology*, 118, 71–78.

Ims, R. & Yoccoz, N. (1997) Studying transfer processes in metapopulations: emigration, migration and colonization. In: *Metapopulation Biology* (I. Hanski & M. Gilpin, eds), pp. 247–265. Academic Press, San Diego, CA.

Inderjit, Wardle, D.A., Karban, R. & Callaway, R.M. (2011) The ecosystem and evolutionary contexts of allelopathy. *Trends in Ecology and Evolution*, 26, 655–662.

Interlandi, S.J. & Kilham, S.S. (2001) Limiting resources and the regulation of diversity in phytoplankton communities. *Ecology*, 82, 1270–1282.

IPBES (2019) Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services' 2019 Global Assessment Report on Biodiversity and Ecosystem Services. Summary for Policy Makers. https://www.ipbes.net (last accessed November 2019).

IPCC (Intergovernmental Panel on Climate Change) (2014) *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Core Writing Team, R.K. Pachauri & L.A. Meyer, eds). IPCC, Geneva, Switzerland.

IUCN (2012) *IUCN Red List Categories and Criteria: Version 3.1*. 2nd edn. IUCN, Cambridge.

IUCN (2015) IUCN Red List of Threatened Species, Version 2015.2. Revised 10 January 2017. www.iucnredlist.org (last accessed November 2019).

IUCN/UNEP/WWF (1991) *Caring for the Earth. A Strategy for Sustainable Living*. Gland, Switzerland.

Ives, A.R. (1992) Continuous-time models of host–parasitoid interactions. *American Naturalist*, 140, 1–29.

Ives, A.R. & Carpenter, S.R. (2007) Stability and diversity of ecosystems. *Science*, 317, 58–62.

Ives, A.R. & Dobson, A.P. (1987) Behavior and the population dynamics of simple predator–prey systems. *American Naturalist*, 130, 431–447.

Izaguirre, M.M., Mazza, C.A., Astgueta, M.S., Ciarla, A.M. & Ballaré, C.L. (2013) No time for candy: passionfruit (*Passiflora edulis)* plants down-regulate damage-induced extra floral nectar production in response to light signals of competition. *Oecologia*, 173, 213–221.

Jack, S.B. & Long, J.N. (1996) Linkages between silviculture and ecology: an analysis of density management diagrams. *Forest Ecology and Management*, 86, 205–220.

Jackson, S.T. & Weng, C. (1999) Late Quaternary extinction of a tree species in eastern North America. *Proceedings of the National Academy of Sciences of the USA*, 96, 13847–13852.

Jackson, T. (2017) *Prosperity without Growth – Foundations for the Economy of Tomorrow*. Routledge, London.

Jamali, H., Livesley, S.J., Dawes, T.Z. *et al.* (2011) Termite mound emissions of CH_4 and CO_2 are primarily determined by seasonal changes in termite biomass and behavior. *Oecologia*, 167, 525–534.

Jamieson, I.G. & Ryan, C.J. (2001) Island takahe: closure of the debate over the merits of introducing Fiordland takahe to predator-free islands. In: *The Takahe: fifty years of conservation management and research* (W.G. Lee & I.G. Jamieson, eds), pp. 96–113. University of Otago Press, Dunedin, New Zealand.

Janis, C.M. (1993) Tertiary mammal evolution in the context of changing climates, vegetation and tectonic events. *Annual Review of Ecology and Systematics*, 24, 467–500.

Janssen, A., van Gool, E., Lingeman, R., Jacas, J. & van de Klashorst, G. (1997) Metapopulation dynamics of a persisting predator–prey system in the laboratory: time series analysis. *Experimental and Applied Acarology*, 21, 415–430.

Janzen, D.H. (1967) Interaction of the bull's-horn acacia (*Acacia cornigera* L.) with an ant inhabitant (*Pseudomyrmex ferruginea* F. Smith) in eastern Mexico. *University of Kansas Science Bulletin*, 47, 315–558.

Janzen, D.H. (1970) Herbivores and the number of tree species in tropical forests. *American Naturalist*, 104, 501–528.

Janzen, D.H. (1981) Evolutionary physiology of personal defence. In: *Physiological Ecology: an evolutionary approach to resource use* (C.R. Townsend & P. Calow, eds), pp. 145–164.

Blackwell Scientific Publications, Oxford.

Janzen, D.H., Juster, H.B. & Bell, E.A. (1977) Toxicity of secondary compounds to the seed-eating larvae of the bruchid beetle *Callosobruchus maculatus*. *Phytochemistry*, 16, 223–227.

Javot, H., Pumplin, N. & Harrison, M.J. (2007) Phosphate in the arbuscular mycorrhizal symbiosis: transport properties and regulatory roles. *Plant, Cell and Environment*, 30, 310–322.

Jaworski, C.C., Bompard, A., Genies, L., Amiens-Desneux, E. & Desneux, N. (2013) Preference and prey switching in a generalist predator attacking local and invasive alien pests. *PLOS One*, 8, e82231.

Jeffries, M.J. & Lawton, J.H. (1985) Predator–prey ratios in communities of freshwater invertebrates: the role of enemy free space. *Freshwater Biology*, 15, 105–112.

Jenkins, B., Kitching, R.L. & Pimm, S.L. (1992) Productivity, disturbance and food web structure at a local spatial scale in experimental container habitats. *Oikos*, 65, 249–255.

Jenkins, S.R. Murua, J. & Burrows, M.T. (2008) Temporal changes in the strength of density-dependent mortality and growth in intertidal barnacles. *Journal of Animal Ecology*, 77, 573–584.

Jenny, J-P., Francus, P., Normandeau, A. *et al.* (2016) Global spread of hypoxia in freshwater ecosystems during the last three centuries is caused by rising local human pressure. *Global Change Biology*, 22, 1481–1489.

Jeschke, J.M. and Strayer, D.L. (2008) Usefulness of bioclimatic models for studying climate change and invasive species. *Annals of the New York Academy of Sciences*, 1134, 1–24.

Jin, H.R., Liu, J, Liu, J. & Huang, X.W. (2012) Forms of nitrogen uptake, translocation, and transfer via arbuscular mycorrhizal fungi: a review. *Science China Life Sciences*, 55, 474–482.

Jobbagy, E.G., Sala, O.E. & Paruelo, J.M. (2002) Patterns and controls of primary production in the Patagonian steppe: a remote sensing approach. *Ecology*, 83, 307–319.

Johannesson, K. (2015) What can be learnt from a snail? *Evolutionary Applications*, 9, 153–165.

John, K., Jauker, F., Marxsen, J. *et al.* (2015) Earthworm bioturbation stabilizes carbon in non-flooded paddy soil at the risk of increasing methane emissions under wet soil conditions. *Soil Biology and Biochemistry*, 91, 127–132.

Johnson, C.M., Zarin, D.J. & Johnson, A.H. (2000) Post-disturbance aboveground biomass accumulation in global secondary forests. *Ecology*, 81, 1395–1401.

Johnson, C.N., Isaac, J.L. & Fisher, D.O. (2007) Rarity of a top predator triggers continent-wide collapse of mammal prey: dingoes and marsupials in Australia. *Proceedings of the Royal Society of London, Series B*, 274, 341–346.

Johnson, E.A. & Miyanishi, K. (2008) Testing the assumptions of chronosequences in succession. *Ecology Letters*, 11, 419–431.

Johnson, N.C., Zak, D.R., Tilman, D. & Pfleger, F.L. (1991) Dynamics of vesicular–arbuscular mycorrhizae during old field succession. *Oecologia*, 86, 349–358.

Johnson, P.T.J., Dobson, A., Lafferty, A.D. *et al.* (2010) When parasites become prey: ecological and epidemiological significance of eating parasites. *Trends in Ecology and Evolution*, 25, 362–371.

Johnston, D.W. & Odum, E.P. (1956) Breeding bird populations in relation to plant succession on the piedmont of Georgia. *Ecology*, 37, 50–62.

Johnston, K. & Robson, B.K. (2009) Commensalism used by freshwater crayfish species to survive drying in seasonal habitats. *Invertebrate Biology*, 128, 269–275.

Johnston, S.E., Gratten, J., Berenos, C. *et al.* (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*, 502, 93–96.

Jolles, A.E., Ezenwa, V.O., Etienne, R.S., Turner, W.C. & Olff, H. (2008) Interactions between macroparasites and microparasites drive infection patterns in free-ranging African buffalo. *Ecology*, 89, 2239–2250.

Jolliffe, P.A. (2000) The replacement series. *Journal of Ecology*, 88, 371–385.

Jones, C.G., Lawton, J.H. & Shachak, M. (1994) Organisms as ecosystem engineers. *Oikos*, 69, 373–386.

Jones, E.I., Afkhami, M.E., Akcay, E. *et al.* (2015) Cheaters must prosper: reconciling theoretical and empirical perspectives on cheating in mutualisms. *Ecology Letters*, 18, 1270–1284.

Jones, N.T., Germain, R.M., Grainger, T.N., Hall, A.M., Baldwin, L. & Gilbert, B. (2015) Dispersal mode mediates the effect of patch size and patch connectivity on metacommunity diversity. *Journal of Ecology*, 103, 935–944.

Jones, O.R., Scheuerlein, A., Salguero-Gomez, R. *et al.* (2014) Diversity of ageing across the tree of life. *Nature*, 505, 169–173.

Jonsson, T. (2017) Metabolic theory predicts animal self-thinning. *Journal of Animal Ecology*, 86, 645–653.

Jordan, R.S. (1991) Impact of ENSO events on the southeastern Pacific region with special reference to the interaction of fishing and climatic variability. In: *ENSO Teleconnections Linking Worldwide Climate Anomalies: scientific basis and societal impacts* (M. Glantz, ed.), pp. 401–430. Cambridge University Press, Cambridge.

Jordano, P., Garcia, C., Godoy, J.A. & Garcia-Castano, J.L. (2007) Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings of the National Academy of Sciences of the USA*, 104, 3278–3282.

Juniper, S.K., Tunnicliffe, V. & Southward, E.C. (1992) Hydrothermal vents in turbidite sediments on a Northeast Pacific spreading centre: organisms and substratum at an ocean drilling site. *Canadian Journal of Zoology*, 70, 1792–1809.

Jutila, H.M. (2003) Germination in Baltic coastal wetland meadows: similarities and differences between vegetation and seed bank. *Plant Ecology*, 166, 275–293.

Kadlec, R.H. (2012) Constructed marshes for nitrate removal. *Critical Reviews in Environmental Science and Technology*, 42, 934–1005.

Kaelke, C.M., Kruger, E.L. & Reich, P.B. (2001) Trade-offs in seedling survival, growth, and physiology among hardwood species of contrasting successional status along a light availability gradient. *Canadian Journal of Forestry Research*, 31, 1602–1616.

Kaiser-Bunbury, C.N., Mougal, J., Whittington, A.E. *et al.* (2017) Ecosystem restoration strengthens pollination network resilience and function. *Nature*, 542, 223–227.

Kamijo, T., Kitayama, K., Sugawara, A., Urushimichi, S. & Sasai, K. (2002) Primary succession of the warm-temperate broad-leaved forest on a volcanic island, Miyake-jima, Japan. *Folia*

Geobotanica, 37, 71–91.

Kang, L., Settlage, R., McMahon, W. *et al.* (2016) Genomic signatures of speciation in sympatric and allopatric Hawaiian picture-winged *Drosophila*. *Genome Biology and Evolution*, 8, 1482–1488.

Kant, M.R., Jonckheere, W., Knegt, B. *et al.* (2015) Mechanisms and ecological consequences of plant defence induction and suppression in herbivore communities. *Annals of Botany*, 115, 1015–1051.

Kaplan, I. & Denno, R.F. (2007) Interspecific interactions in phytophagous insects revisited: a quantitative assessment of competition theory. *Ecology Letters*, 10, 977–994.

Kaplan, I., Halitschke, R., Kessler, A., Sardanelli, S. & Denno, R.F. (2008) Constitutive and induced defenses to herbivory in above- and below ground plant tissues. *Ecology*, 89, 392–406.

Kaplan, R.H. & Salthe, S.N. (1979) The allometry of reproduction: an empirical view in salamanders. *American Naturalist*, 113, 671–689.

Karban, R. (2011) The ecology and evolution of induced resistance against herbivores. *Functional Ecology*, 25, 339–347.

Karban, R., Agrawal, A.A., Thaler, J.S. & Adler, L.S. (1999) Induced plant responses and information content about risk of herbivory. *Trends in Ecology and Evolution*, 14, 443–447.

Karell, P., Ahola, K., Karstinen, T., Zolei, A. & Brommer, J.E. (2009) Population dynamics in a cyclic environment: consequences of cyclic food abundance on tawny owl reproduction and survival. *Journal of Animal Ecology*, 78, 1050–1062.

Karels, T.J. & Boonstra, R. (2000) Concurrent density dependence and independence in populations of arctic ground squirrels. *Nature*, 408, 460–463.

Karels, T.J., Byrom, A.E., Boonstra, R. & Krebs, C.J. (2000) The interactive effects of food and predators on reproduction and overwinter survival of arctic ground squirrels. *Journal of Animal Ecology*, 69, 235–247.

Karl, D.M., Bjorkman, K.M., Dore, J.E. *et al.* (2001) Ecological nitrogen-to-phosphorus stoichiometry at station ALOHA. *Deep-Sea Research, II* 48, 1529–1566.

Karubian, J., Fabara, J., Yunes, D., Jorgeson, J.P., Romo, D. & Smith, T.B. (2005) Temporal and spatial patterns of macaw abundance in the Ecuadorian Amazon. *The Condor*, 107, 617–626.

Kathiresan, K. & Rajendran, N. (2005) Coastal mangrove forests mitigated tsunami. *Estuarine, Coastal and Shelf Science*, 65, 601–606.

Kattge, J., Knorr, W., Raddatz, T. & Wirth, C. (2009) Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models. *Global Change Biology*, 15, 976–991.

Kautz, G., Zimmer, M. & Topp, W. (2002) Does *Porcellio scabar* (Isopoda: Oniscidea) gain from coprophagy? *Soil Biology and Biochemistry*, 34, 1253–1259.

Kawano, K. (2002) Character displacement in giant rhinoceros beetles. *American Naturalist*, 159, 255–271.

Kays, S. & Harper, J.L. (1974) The regulation of plant and tiller density in a grass sward. *Journal of Ecology*, 62, 97–105.

Kearns, P.J., Angell, J.H., Howard, E.H., Deegan, L.A., Stanley, R.H.R. & Bowen, J.L. (2016) Nutrient enrichment induces dormancy and decreases diversity of active bacteria in salt marsh sediments. *Nature Communications*, 7, 12881.

Keddy, P.A. (1982) Experimental demography of the sand-dune annual, *Cakile edentula*, growing along an environmental gradient in Nova Scotia. *Journal of Ecology*, 69, 615–630.

Keeling, M. (1999) Spatial models of interacting populations. In: *Advanced Ecological Theory* (J. McGlade, ed.), pp. 64–99. Blackwell Science, Oxford.

Keesing, F., Belden, L.K., Daszak, P. *et al.* (2010) Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature*, 468, 647–652.

Keesing, F., Holt, R.D. & Ostfeld, R.S. (2006) Effects of species diversity on disease risk. *Ecology Letters*, 9, 485–498.

Kelly, C.A. & Dyer, R.J. (2002) Demographic consequences of inflorescence-feeding insects for *Liatris cylindrica*, an iteroparous perennial. *Oecologia*, 132, 350–360.

Kelly, D. & Sork, V. (2002) Mast seeding in perennial plants: why, how, where? *Annual Review of Ecology and Systematics*, 33, 427–447.

Kerkhoff, A.J., Moriarty, P.E. & Weiser, M.D. (2014) The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. *Proceedings of the National Academy of Sciences of the USA*, 111, 8125–8130.

Kersch-Becker, M.F. & Lewinsohn, T.M. (2012) Bottom-up multitrophic effects in resprouting plants. *Ecology*, 93, 9–16.

Kery, M., Matthies, D. & Fischer, M. (2001) The effect of plant population size on the interactions between the rare plant *Gentiana cruciata* and its specialized herbivore *Maculinea rebeli*. *Journal of Ecology*, 89, 418–427.

Kessler, A. & Baldwin, I.T. (2004) Herbivore-induced plant vaccination. Part I. The orchestration of plant defences in nature and their fitness consequences in the wild tobacco. *Plant Journal*, 38, 639–649.

Kettlewell, H.B.D. (1955) Selection experiments on industrial melanism in the Lepidoptera. *Heredity*, 9, 323–342.

Khaliq, I., Hof, C., Prinzinger, R., Bohning-Gaese, K. & Pfenninger, M. (2014) Global variation in thermal tolerances and vulnerability of endotherms to climate change. *Proceedings of the Royal Society of London, Series B*, 281, 20141097.

Kiær, L.P., Weisbach, A.N. & Weoner, J. (2013) Root and shoot competition: a meta-analysis. *Journal of Ecology*, 101, 1298–1312.

Kiba, T. & Krapp, A. (2016) Plant nitrogen acquisition under low availability: regulation of uptake and root architecture. *Plant Cell Physiology*, 57, 707–714.

Kiers, E.T., Rousseau, R.A., West, S.A. & Denison, R.F. (2003) Host sanctions and the legume–rhizobium mutualism. *Nature*, 425, 78–81.

Kim, T.N., Underwood, N. & Inouye, B.D. (2013) Insect herbivores change the outcome of plant competition through both inter- and intraspecific processes. *Ecology*, 94, 1753–1763.

Kimball, B.A., Morris, C.F., Pinter, P.J. Jr *et al.* (2001) Elevated CO_2, drought and soil nitrogen effects on wheat grain quality. *New Phytologist*, 150, 295–303.

Kingsland, S.E. (1985) *Modeling Nature*. University of Chicago Press, Chicago.

Kingsolver, J.G. & Buckley, L.B. (2017) Evolution of plasticity and adaptive responses to climate change along climate gradients. *Proceedings of the Royal Society of London, Series B*, 284, 20170386.

Kingsolver, J.G. & Pfennig, D.W. (2008) Patterns and power of phenotypic selection in nature. *BioScience*, 57, 561–572.

Kirchman, D.L. (2012) *Processes in Microbial Ecology*. Oxford University Press, Oxford.

Kirk, J.T.O. (1994) *Light and Photosynthesis in Aquatic Ecosystems*. Cambridge University Press, Cambridge, UK.

Kittle, A.M., Anderson, M., Avgar, T. *et al.* (2015) Wolves adapt territory size, not pack size to local habitat quality. *Journal of Animal Ecology*, 84, 1177–1184.

Kleiber, M. (1932) Body size and metabolism. *Hilgardia*, 6, 315–353.

Kleintjes, P.K., Jacobs, B.F. & Fettig, S.M. (2004) Initial response of butterflies to an overstory reduction and slash mulching treatment of a degraded pinon-juniper woodland. *Restoration Ecology*, 12, 231–238.

Klemola, T., Koivula, M., Korpimaki, E. & Norrdahl, K. (2000) Experimental tests of predation and food hypotheses for population cycles of voles. *Proceedings of the Royal Society of London, Series B*, 267, 352–356.

Klemola, T., Tanhuanpää, M., Korpimäki, E. & Ruohomäki, K. (2002) Specialist and generalist natural enemies as an explanation for geographical gradients in population cycles of northern herbivores. *Oikos*, 99, 83–94.

Klemow, K.M. & Raynal, D.J. (1981) Population ecology of *Melilotus alba* in a limestone quarry. *Journal of Ecology*, 69, 33–44.

Kling, G.W., Kipphut, G.W., Miller, M.M. & O'Brien, W.J. (2000) Integration of lakes and streams in a landscape perspective: the importance of material processing on spatial patterns and temporal coherence. *Freshwater Biology*, 43, 477–497.

Kneip, C., Lockhart, P., Voß, C. and Maier. U-G. (2007) Nitrogen fixation in eukaryotes – new models for symbiosis. *BMC Evolutionary Biology*, 7, 55.

Knell, R.J. (1998) Generation cycles. *Trends in Ecology and Evolution*, 13, 186–190.

Knutson, A.E., Butler, J., Bernal, J., Bográn, C. & Campos, M. (2011) Impact of area-wide malathion on predatory arthropods and secondary pests in cotton during boll weevil eradication in Texas. *Crop Protection*, 30, 456–467.

Koenig, W.D., Knops, J.M.H., Carmen, W.J. & Pesendorfer, M.B. (2017) Testing the terminal investment hypothesis in California oaks. *American Naturalist*, 189, 564–569.

Koenig, W.D. & Liebhold, A.M. (2005) Effects of periodical cicada emergences on abundance and synchrony of avian populations. *Ecology*, 86, 1873–1882.

Koike, T., Kitao, M., Maruyama, Y., Mori, S. & Lei, T.T. (2001) Leaf morphology and photosynthetic adjustments among deciduous broad-leaved trees within the vertical canopy profile. *Tree Physiology*, 21, 951–958.

Kolokotrones, T., Savage, V., Deeds, E.J. & Fontana, W. (2010) Curvature in metabolic scaling. *Nature*, 464, 753–756.

Kondoh, M. & Ninomiya, K. (2009) Food-chain length and adaptive foraging. *Proceedings of the Royal Society of London, Series B*, 276, 3113–3121.

Koop, J.A.H., Huber, S.K., Laverty, S.M. & Clayton, D.H. (2011) Experimental demonstration of the fitness consequences of an introduced parasite of Darwin's finches. *PLoS One*, 6, e19706.

Koop, J.A.H., Kim, P.S., Knutie, S.A., Adler, F. & Clayton, D.H. (2016) An introduced parasitic fly may lead to local extinction of Darwin's finch populations. *Journal of Applied Ecology*, 53, 511–518.

Körner, C. & Paulsen, J. (2004) A world-wide study of high altitude treeline temperatures. *Journal of Biogeography*, 31, 713–732.

Korpimaki, E., Norrdahl, K., Klemola, T., Pettersen, T. & Stenseth, N.C. (2002) Dynamic effects of predators on cyclic voles: field experimentation and model extrapolation. *Proceedings of the Royal Society of London, Series B*, 269, 991–997.

Kozlowski, J. (1993) Measuring fitness in life-history studies. *Trends in Ecology and Evolution*, 7, 155–174.

Kraft, C.E. & Johnson, L.E. (2000) Regional differences in rates and patterns of North American inland lake invasions by zebra mussels (*Dreissena polymorpha*). *Canadian Journal of Fisheries and Aquatic Sciences*, 57, 993–1001.

Kramer, D.L. & Chapman, M.R. (1999) Implications of fish home range size and relocation for marine reserve function. *Environmental Biology of Fishes*, 55, 65–79.

Kratz, T.K., Webster, K.E., Bowser, C.J., Magnuson, J.J. & Benson, B.J. (1997) The influence of landscape position on lakes in northern Wisconsin. *Freshwater Biology*, 37, 209–217.

Krause, J. & Ruxton, G.D. (2002) *Living in Groups*. Oxford University Press, Oxford.

Krebs, C.J. (1972) *Ecology*. Harper & Row, New York.

Krebs, C.J. (1999) *Ecological Methodology*, 2nd edn. Addison-Welsey Educational, Menlo Park, CA.

Krebs, C.J. (2011) Of lemmings and snowshoe hares: the ecology of northern Canada. *Proceedings of the Royal Society of London, Series B*, 278, 481–489.

Krebs, C.J., Boonstra, R & Boutin, S. (2017) Using experimentation to understand the 10-year snowshoe hare cycle in the boreal forest of North America. *Journal of Animal Ecology*, 87, 87–100.

Krebs, C.J., Boonstra, R., Boutin, S. & Sinclair, A.R.E. (2001) What drives the 10-year cycle of snowshoe hares? *Bioscience*, 51, 25–35.

Krebs, C.J., Sinclair, A.R.E., Boonstra, R., Boutin, S., Martin, K. & Smith, J.N.M. (1999) Community dynamics of vertebrate herbivores: how can we untangle the web? In: *Herbivores: between plants and predators* (H. Olff, V.K. Brown & R.H. Drent, eds), pp. 447–473. Blackwell Science, Oxford.

Krebs, J.R. (1978) Optimal foraging: decision rules for predators. In: *Behavioural Ecology: an evolutionary approach* (J.R. Krebs & N.B. Davies, eds), pp. 23–63. Blackwell Scientific Publications, Oxford.

Krebs, J.R. & Davies, N.B. (1993) *An Introduction to Behavioural Ecology*, 3rd edn. Blackwell Scientific Publications, Oxford.

Krebs, J.R., Erichsen, J.T., Webber, M.I. & Charnov, E.L. (1977) Optimal prey selection in the great tit (*Parus major*). *Animal Behaviour*, 25, 30–38.

Krebs, J.R. & Kacelnik, A. (1991) Decision-making. In: *Behavioural Ecology: an evolutionary approach*, 3rd edn (J.R. Krebs & N.B. Davies, eds), pp. 105–136. Blackwell Scientific Publications, Oxford.

Krebs, J.R., Stephens, D.W. & Sutherland, W.J. (1983) Perspectives in optimal foraging. In: *Perspectives in Ornithology* (A.H. Brush & G.A. Clarke, Jr, eds), pp. 165–216. Cambridge University Press, New York.

Krivan, V. (2007) The Lotka-Volterra predator-prey model withforaging–predation risk trade-offs. *American Naturalist*, 170, 771–782.

Kromdijk, J., Głowacka, K., Leonelli, L., Gabilly, S.T., Iwai, M., Niyogi, K.K. & Long, S.P. (2016) Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science*, 354, 857–861.

Kroon, F.J., Thorburn, P., Schaffelke, B. & Whitten, S. (2016) Towards protecting the Great Barrier Reef from land-based pollution. *Global Change Biology*, 22, 1985–2002.

Kubanek, J., Whalen, K.E., Engel, S. *et al.* (2002) Multiple defensive roles for triterpene glycosides from two Caribbean sponges. *Oecologia*, 131, 125–136.

Kuhar, T.P., Wright, M.G. Hoffmann, M.P. & Chenus, S.A. (2002) Life table studies of European corn borer (Lepidoptera: Crambidae) with and without inoculative releases of *Trichogramma ostriniae* (Hymenoptera: Trichogrammatidae). *Environmental Entomology*, 31, 482–489.

Kumar, J. & Atri, N.S. (2018) Studies on ectomyccorhiza: an appraisal. *Botanical Reviews*, 84, 108–155.

Kunert, G. & Weisser, W.W. (2003) The interplay between densityand trait-mediated effects in predator–prey interactions: a case study in aphid wing polymorphism. *Oecologia*, 135, 304–312.

Kunkel, C.M., Hallberg, R.W. & Oppenheimer, M. (2006) Coral reefs reduce tsunami impact in model simulations. *Geophysical Research Letters*, 33, L23612.

Kuno, E. (1991) Some strange properties of the logistic equation defined with *r* and *K*: inherent defects or artifacts? *Researches on Population Ecology*, 33, 33–39.

Kunstler, G., Albert, C.H., Courbaud, B. *et al.* (2011) Effects of competition on tree radial-growth vary in importance but not in intensity along climatic gradients. *Journal of Ecology*, 99, 300–312.

Laas, A., Cremola, F., Meinson, P., Room, E-I., Noges, T. & Noges, P. (2016) Summer depth distribution profiles of CO_2 and O_2 in shallow temperate lakes reveal trophic state and lake type specific differences. *Science of the Total Environment*, 566–567, 63–75.

Labbé, P. (1994) Régénération après passage du cyclone Hugo en for*b*t dense humide de Guadeloupe. *Acta Ecologica*, 15, 301–315.

Lack, D. (1947) The significance of clutch size. *Ibis*, 89, 302–352.

Lack, D. (1963) Cuckoo hosts in England. (With an appendix on the cuckoo hosts in Japan, by T. Royama.) *Bird Study*, 10, 185–203.

Lacy, R. & Pollak, J. (2013) *VORTEX: A stochastic simulation of the extinction process, version 10.0*. Chicago Zoological Society, Brookfield, IL.

Lafferty, K.D., DeLeo, G., Briggs, C.J., Dobson, A.P., Gross, T. & Kuris, A.M. (2015) A general consumer-resource population model. *Science*, 349, 854–857.

Lafferty, K.D., Dobson, A.P. & Kuris, A.M. (2006) Parasites dominate food web links. *Proceedings of the National Academy of Sciences of the USA*, 103, 11211–11216.

Lakoff, G. (2009) *The Political Mind: A Cognitive Scientist's Guide to Your Brain and Its Politics*. Penguin Books, London.

Lai, P.C.C. & Wong, B.S.F. (2005) Effects of tree guards and weed mats on the establishment of native tree seedlings: implications for forest restoration in Hong Kong, China. *Restoration Ecology*, 13, 138–143.

Lai, Y.T., Chen, J.H. & Lee, L.L. (2011) Prey selection of a shell-invading leech as predicted by optimal foraging theory with consumption success incorporated into estimation of prey profitability. *Functional Ecology*, 25, 147–157.

Laing, A. & Evans, J-L. (2011) *Introduction to Tropical Meteorology*, 2nd edn. Comet, University Program for Atmospheric Research.

Lamb, J.B., Willis, B.L., Fiorenza, E.A. *et al.* (2018) Plastic waste associated with disease on coral reefs. *Science*, 359(6374), 460–462.

Lambers, H., Martinoia, E. & Renton, M. (2015) Plant adaptations to severely phosphorus-impoverished soils. *Current Opinion in Plant Biology*, 25, 23–31.

Lambin, X. (2018) The population dynamics of bite-sized predators: prey dependence, territoriality, and mobility. In: *Biology and Conservation of Musteloids* (D.W. Macdonald. C. Newman & L.A. Harrington, eds), pp. 129–148. Oxford University Press, Oxford.

Lambin, X., Aars, J. & Piertney, S.B. (2001) Dispersal, intraspecific competition, kin competition and kin facilitation: a review of the empirical evidence. In: *Dispersal* (J. Clobert, E. Danchin, A.A. Dhondt & J.D. Nichols, eds), pp. 110–122. Oxford University Press, Oxford.

Lambin, X., Bretagnolle, V. & Yoccoz, N.G. (2006) Vole population cycles in northern and southern Europe: is there a need for different explanations for single pattern? *Journal of Animal Ecology*, 75, 340–349.

Lambin, X. & Krebs, C.J. (1993) Influence of female relatedness on the demography of Townsend's vole populations in spring. *Journal of Animal Ecology*, 62, 536–550.

Lambin, X., Krebs, C.J., Moss, R., Stenseth, N.C. & Yoccoz, N.G. (1999) Population cycles and parasitism (technical comment). *Science*, 286, 2425a.

Lambin, X., Petty, S.J. & MacKinnon, J.L. (2000) Cyclic dynamics in field vole populations and generalist predation. *Journal of Animal Ecology*, 69, 106–118.

Lambin, X. & Yoccoz, N.G. (1998) The impact of population kinstructure on nestling survival in Townsend's voles, *Microtus townsendii*. *Journal of Animal Ecology*, 67, 1–16.

Lande, R. & Barrowclough, G.F. (1987) Effective population size, genetic variation, and their use in population management. In: *Viable Populations for Conservation* (M.E. Soulé, ed.), pp. 87–123. Cambridge University Press, Cambridge.

Landeweert, R., Hoffland, E., Finlay, R.D., Kuyper, T.W. & van Breemen, N. (2001) Linking plants to rocks: ectomycorrhizal fungi mobilize nutrients from minerals. *Trends in Ecology and Evolution*, 16, 248–254.

Lang, M., Prestele, J., Fischer, C., Kollmann, J. & Albrecht, H. (2016) Reintroduction of rare arable plants by seed transfer. What are the optimal sowing rates? *Ecology and Evolution*, 6, 5506–5516.

Lange, O.L., Nobel, P.S., Osmond, C.B. & Ziegler, H. (2012) *Physiological Plant Ecology I: Responses to the Physical Environment*. Springer Science & Business Media, Berlin.

Lankester, F., Hampson, K., Lembo, T., Palmer, G., Taylor, L. & Cleaveland, S. (2014) Implementing Pasteur's vision for rabies elimination. *Science*, 345, 1562–1564.

Lanzas, M., Hermoso, V., de-Miguel, S, Bota, G & Brotons, L. (2019) Designing a network of green infrastructure to enhance the conservation value of protected areas and maintain ecosystem services. *Science of the Total Environment*, 651, 541–550.

Larcher, W. (1980) *Physiological Plant Ecology*, 2nd edn. Springer-Verlag, Berlin.

LaRue, M.A. & Nielsen, C.K. (2016) Population viability of recolonizing cougars in midwestern North America. *Ecological Modelling*, 321, 121–129.

Lau, J.A. (2008) Beyond the ecological: biological invasions alter natural selection on a native plant species. *Ecology*, 89, 1023–1031.

Lauber, C.L., Metcalf, J.L, Keepers, K., Ackermann, G., Carter, D.O. & Knight, R. (2014) Vertebrate decomposition is accelerated by soil microbes. *Applied and Environmental Microbiology*, 80, 4920–4929.

Laugen, A.T., Laurila, A. & Rasanen, K. *et al.* (2003) Latitudinal countergradient variation in the common frog (Rana temporaria) development rates – evidence for local adaptation. *Journal of Evolutionary Biology*, 16, 996–1005.

Law, B.E., Thornton, P.E., Irvine, J., Anthoni, P.M. & van Tuyl, S. (2001) Carbon storage and fluxes in ponderosa pine forests at different developmental stages. *Global Climate Change*, 7, 755–777.

Lawler, I.R., Foley, W.J. & Eschler, B.M. (2000) Foliar concentration of a single toxin creates habitat patchiness for a marsupial folivore. *Ecology*, 81, 1327–1338.

Lawlor, L.R. (1980) Structure and stability in natural and randomly constructed competitive communities. *American Naturalist*, 116, 394–408.

Lawton, J.H. (1989) Food webs. In: *Ecological Concepts* (J.M. Cherrett, ed.), pp. 43–78. Blackwell Scientific Publications, Oxford.

Lawton, J.H. & Woodroffe, G.L. (1991) Habitat and the distribution of water voles: why are there gaps in a species' range? *Journal of Animal Ecology*, 60, 79–91.

Le Quéré, C., Andrew, R.M., Canadell, J.G., Sitch, S. & Korsbakken, J.I. (2016) Global Carbon Budget 2016. *Earth System Science Data*, doi:10.5194/essd-8-605-2016.

Lee, J.H., Kubo, Y., Fujiwara, T., Septiana, R.M., Riyanto, S. & Iwasa, Y. (2018) Profit sharing as a management strategy for a state-owned teak plantation at high risk for illegal logging. *Ecological Economics*, 149, 140–148.

Lee, W.G. & Jamieson, I.G. (eds) (2001) *The Takahe: fifty years of conservation management and research*. University of Otago Press, Dunedin, New Zealand.

Lee, W-S., Monaghan, P. & Metcalfe, N.B. (2013) Experimental demonstration of the growth rate – lifespan trade-off. *Proceedings of the Royal Society of London, Series B*, 280, 20122370.

Lefebvre, M., Langrell, S.R.H. & Gomez-y-Paloma, S. (2015) Incentives and policies for integrated pest management in Europe: a review. *Agronomy for Sustainable Development*, 35, 27–45.

Leibold, M.A., Holyoak, M., Mouquet, N. *et al.* (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, 7, 601–613.

Leigh, E. (1975) Population fluctuations and community structure. In: *Unifying Concepts in Ecology* (W.H. van Dobben & R.H. Lowe-McConnell, eds), pp. 67–88. Junk, The Hague.

Leigh, E.G. (1981) The average lifetime of a population in a varying environment. *Journal of Theoretical Biology*, 90, 213–239.

Leimu, R. & Fischer, M. 2008. A meta-analysis of local adaptation in plants. *PLoS One*, 3, e4010.

Lekve, K., Ottersen, G., Stenseth, N.C. & Gjøsæter, J. (2002) Length dynamics in juvenile coastal Skagerrak cod: effects of biotic and abiotic processes. *Ecology*, 83, 1676–1688.

Lemieux, A.M. & Clarke, R.V. (2009) The international ban on ivory sales and its effects on elephant poaching in Africa. *British Journal of Criminology*, 49, 451–471.

Lenka, N.K. & Lal, R. (2012) Soil-related constraints to the carbon dioxide fertilization effect. *Critical Reviews in Plant Sciences*, 31, 342–357.

Lennartsson, T., Nilsson, P. & Tuomi, J. (1998) Induction of overcompensation in the field gentian, *Gentianella campestris*. *Ecology*, 79, 1061–1072.

Leroux, S., Larrivee, M., Boucher-Lalande, V. *et al.* (2013) Mechanistic models for the spatial spread of species under climate change. *Ecological Applications*, 23, 815–828.

Lester, P.J., Haywood, J., Archer, M.E. & Shortall, C.R. (2017) The long-term population dynamics of common wasps in their native and invaded range. *Journal of Animal Ecology*, 86, 337–347.

Letnic, M. Crowther, M.S. & Koch, F. (2009) Does a top-predator provide an endangered rodent with refuge from an invasive mesopredator? *Animal Conservation*, 12, 302–312.

Letourneau, D.K. & Dyer, L.A. (1998a) Density patterns of *Piper* ant-plants and associated arthropods: top-predator trophic cascades in a terrestrial system? *Biotropica*, 30, 162–169.

Letourneau, D.K. & Dyer, L.A. (1998b) Experimental test in a lowland tropical forest shows top-down effects through four trophic levels. *Ecology*, 79, 1678–1687.

Leung, T.L.F. & Poulin, R. (2008) Parasitism, commensalism, and mutualism: exploring the many shades of symbioses. *Vie et Milieu – Life and Environment*, 58, 107–115.

Leverich W.J. & Levin, D.A. (1979) Age-specific survivorship and reproduction in *Phlox drummondii*. *American Naturalist*, 113, 881–903.

Levey, D.J. (1988) Tropical wet forest treefall gaps and distributions of understorey birds and plants. *Ecology*, 69, 1076–1089.

Levins, R. (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, 15, 237–240.

Levins, R. (1970) Extinction. In: *Lectures on Mathematical Analysis of Biological Phenomena*, pp. 123–138. Annals of the New York Academy of Sciences, Vol. 231.

Lewis, S.L., Lloyd, J., Sitch, S., Mitchard, E.T.A & Laurance, W.F. (2009) Changing ecology of tropical forests: evidence and drivers. *Annual Review of Ecology Evolution and Systematics*, 40, 529–549.

Lewontin, R.C. & Levins, R. (1989) On the characterization of density and resource availability. *American Naturalist*, 134, 513–524.

Li, L., Weiner, J., Zhou, D., Huang, Y. & Shen, L. (2013) Initial density affects biomass–density and allometric relationships in self-thinning populations of *Fagopyrum esculentum*. *Journal of Ecology*, 101, 475–483.

Li, X., Sorensen, P., Li, F., Petersen, S.O. & Olesen, J.E. (2015) Quantifying biological nitrogen fixation of different catch crops, and residual effects of roots and tops on nitrogen uptake in barley using in-situ ^{15}N labelling. *Plant Soil*, 395, 273–287.

Li, Y., Vina, A., Yang, W. *et al.* (2013) Effects of conservation policies on forest cover change in giant panda habitat regions, China. *Land Use Policy*, 33, 42–53.

Li, Z., Wakao, S., Fischer, B.B. & Niyogi, K.K. (2009) Sensing and responding to excess light. *Annual Review of Plant Biology*, 60, 239–260.

Liang, C. & Pauly, D. (2017) Growth and mortality of exploited fishes in China's coastal seas and their uses for yield-per-recruit analyses. *Journal of Applied Ichthyology*, 33, 746–756.

Lichter, J. (2000) Colonization constraints during primary succession on coastal Lake Michigan sand dunes. *Journal of Ecology*, 88, 825–839.

Lidicker, W.Z. Jr (1975) The role of dispersal in the demography of small mammal populations. In: *Small Mammals: their productivity and population dynamics* (K. Petruscwicz, F.B. Golley & L. Ryszkowski, eds), pp. 103–128. Cambridge University Press, New York.

Lidicker, W.Z. Jr (2010) The Allee effect: its history and future importance. *The Open Journal of Ecology*, 3, 71–82

Liebig, J. (1840) *Organic Chemistry and its Application to Agriculture and Physiology*. Taylor and Walton, London.

Likens, G.E. & Bormann, F.G. (1975) An experimental approach to New England landscapes. In: *Coupling of Land and Water Systems* (A.D. Hasler, ed.), pp. 7–30. Springer-Verlag, New York.

Lillebo, A.I., Flindt, M.R., Pardal, M.A. & Marques, J.C. (1999) The effect of macrofauna, meiofauna and microfauna on the degradation of *Spartina maritima* detritus from a salt marsh area. *Acta Oecologica*, 20, 249–258.

Lim, J.Y & Marshall, C.R. (2017) The true tempo of evolutionary radiation and decline revealed on the Hawaiian archipelago. *Nature*, 543, 710–713.

Lima, M., Ernest, S.K.M., Brown, J.H., Belgrano, A. & Stenseth, N.C. (2008) Chihuahuan desert kangaroo rats: nonlinear effects of population dynamics, competition and rainfall. *Ecology*, 89, 2594–2603.

Lima, M., Keymer, J.E. & Jaksic, F.M. (1999) El Niño–Southern Oscillation-driven rainfall variability and delayed density dependence cause rodent outbreaks in western South America: linking demography and population dynamics. *American Naturalist*, 153, 476–491.

Lindemann, R.L. (1942) The trophic–dynamic aspect of ecology. *Ecology*, 23, 399–418.

Lipowsky, A., Roscher, C., Schumacher, J. & Schmid, B. (2012) Density-independent mortality and increasing plant diversity are associated with differentiation of *Taraxacum officinale* into r- and K-strategists. *PLoS One*, 7, e28121.

Lips, K.R., Brem, F., Brenes, R. *et al.* (2006) Emerging infectious disease and the loss of biodiversity in a Neotropical amphibian community. *Proceedings of the National Academy of*

Sciences of the USA, 103, 3165–3170.

Liu X., Zhang, Y., Han, W. *et al.* (2013) Enhanced nitrogen deposition over China. *Nature*, 494, 459–462.

Lloyd-Smith, J.O., Schreiber, S.J., Kopp, P.E. & Getz, W.M. (2005) Superspreading and the effect of individual variation on disease emergence. *Nature*, 438, 355–359.

Lochmiller, R.L. & Derenberg, C. (2000) Trade-offs in evolutionary immunology: just what is the cost of immunity? *Oikos*, 88, 87–98.

Lofgren, A. & Jerling, L. (2002) Species richness, extinction and immigration rates of vascular plants on islands in the Stockholm Archipelago, Sweden, during a century of ceasing management. *Folia Geobotanica*, 37, 297–308.

Loik, M.E. (2008) The effect of cactus spines on light interception and Photosystem II for three sympatric species of Opuntia from the Mojave Desert. *Physiologia Plantarum*, 134, 87–98.

Loladze, I. (2002) Rising atmospheric CO2 and human nutrition: toward globally imbalanced plant stoichiometry? *Trends in Ecology and Evolution*, 17, 457–461.

Lonsdale, W.M. (1990) The self-thinning rule: dead or alive? *Ecology*, 71, 1373–1388.

Lonsdale, W.M. & Watkinson, A.R. (1983) Light and self-thinning. *New Phytologist*, 90, 431–435.

Loreau, M. (2000) Are communities saturated? On the relationship between α, β and γ diversity. *Ecology Letters*, 3, 73–76.

Loreau. M. (2004) Does functional redundancy exist? *Oikos*, 104, 606–611.

Loreau, M. (2010a) *From Populations to Ecosystems: Theoretical Foundations for a New Ecological Synthesis*. Princeton University Press, Princeton NJ.

Loreau, M. (2010b) Linking biodiversity and ecosystems: towards a unifying ecological theory. *Philosophical Transactions of the Royal Society B*, 365, 49–60.

Loreau, M. & Hector, A. (2001) Partitioning selection and complementarity in biodiversity experiments. *Nature*, 412, 72–76.

Loreau, M., Naeem, S., Inchausti, P. *et al.* (2001) Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science*, 294, 804–808.

Lortie, C.J. & Callaway, R.M. (2006) Re-analysis of meta-analysis: support for the stress-gradient hypothesis. *Journal of Ecology*, 94, 7–16.

Losos, E. (1993) The future of the US Endangered Species Act. *Trends in Ecology and Evolution*, 8, 332–336.

Lotka, A.J. (1932) The growth of mixed populations: two species competing for a common food supply. *Journal of the Washington Academy of Sciences*, 22, 461–469.

Loucks, C.J., Zhi, L., Dinerstein, E., Dajun, W., Dali, F. & Hao, W. (2003) The giant pandas of the Qinling Mountains, China: a case study in designing conservation landscapes for elevational migrants. *Conservation Biology*, 17, 558–565.

Louda, S.M., Arnett, A.E., Rand, T.A. & Russell, F.L. (2003a) Invasiveness of some biological control insects and adequacy of their ecological risk assessment and regulation. *Conservation Biology*, 17, 73–82.

Louda, S.M., Pemberton, R.W., Johnson, M.T. & Follett, P.A. (2003b) Non-target effects – the Achilles' heel of biological control? Retrospective analyses to reduce risk associated with biocontrol introductions. *Annual Review of Entomology*, 48, 365–396.

Loveridge, A.J., Valeix, M., Davidson, Z., Murindagomo, F., Fritz, H. & Macdonald, D.W. (2009) Changes in home range size of African lions in relation to pride size and prey biomass in a semi-arid savanna. *Ecography*, 32, 953–962.

Lovett Doust, L. & Lovett Doust, J. (1982) The battle strategies of plants. *New Scientist*, 95, 81–84.

Lowe, C.D., Minter, E.J., Cameron, D.D. & Brockhurst, M.A. (2016) Shining a light on exploitative host control in a photosynthetic endosymbiosis. *Current Biology*, 26, 207–211.

Lubbers, I.M., van Groenigen, K.J., Brussaard, L. *et al.* (2015) Reduced greenhouse gas mitigation potential of no-tillage soils through earthworm activity. *Scientific Reports*, 5, 13787.

Lubchenco, J., Olson, A.M., Brubaker, L.B. *et al.* (1991) The sustainable biosphere initiative: an ecological research agenda. *Ecology*, 72, 371–412.

Luna-Ramirez, K., Skaljac, M., Grotmann, J., Kirfel, P. & Vilcinskas, A. (2017) Orally delivered scorpion antimicrobial peptides exhibit activity against pea aphid (*Acyrthosiphon pisum*) and its bacterial symbionts. *Toxins*, 9, 261–277.

Luo, T., Li, W. & Zhu, H. (2002) Estimated biomass and productivity of natural vegetation on the Tibetan Plateau. *Ecological Applications*, 12, 980–997.

Luo, Z., Tang, S., Li, C. *et al.* (2012) Environmental effects on vertebrate species richness: testing the energy, environmental stability and habitat heterogeneity hypotheses. *PLoS One*, e35514.

Lusher, A.L., Hernandez-Milian, G., O'Brien, J., Berrow, S., O'Connor, I. & Officer, R. (2015) Microplastic and macroplastic ingestion by a deep diving, oceanic cetacean: the True's beaked whale Mesoplodon mirus. *Environmental Pollution*, 199, 185–191.

Lusher, A.L., O’Donnell, C., Officer, R. & O’Connor, I. (2016) Microplastic interactions with North Atlantic mesopelagic fish. *ICES Journal of Marine Science*, 73, 1214–1225.

Lussenhop, J. (1992) Mechanisms of microarthropod–microbial interactions in soil. *Advances in Ecological Research*, 23, 1–33.

Lustenhouwer, N., Wilschut, R.A., Williams, J.L, van der Putten, W.H. & Levine, J.M. (2017) Rapid evolution of phenology during range expansion with recent climate change. *Global Change Biology*, 24, e534–e544.

Lutz, M.J., Caldeira, K., Dunbar, R.B. & Behrenfield, M.J. (2007) Seasonal rhythms of net primary production and particulate organic carbon flux to depth describe the efficiency of biological pump in the global ocean. *Journal of Geophysical Research*, 112, doi:10.1029/2006JC003706.

Lynch, H.J., Zeigler, S., Wells, L., Ballou, J.D. & Fagan, W.F. (2010) Survivorship patterns in captive mammalian populations: implications for estimating population growth rates. *Ecological Applications*, 20, 2334–2345.

MacArthur, J.W. (1975) Environmental fluctuations and species diversity. In: *Ecology and Evolution of Communities* (M.L. Cody & J.M. Diamond, eds), pp. 74–80. Belknap, Cambridge, MA.

MacArthur, R.H. (1955) Fluctuations of animal populations and a measure of community stability. *Ecology*, 36, 533–536.

MacArthur, R.H. (1962) Some generalized theorems of natural selection. *Proceedings of the National Academy of Science of the USA*, 48, 1893–1897.

MacArthur, R.H. (1972) *Geographical Ecology*. Harper & Row, New York.

MacArthur, R.H. & Levins, R. (1964) Competition, habitat selection and character displacement in a patchy environment. *Proceedings of the National Academy of Sciences of the USA*, 51, 1207–1210.

MacArthur, R.H. & Pianka, E.R. (1966) On optimal use of a patchy environment. *American Naturalist*, 100, 603–609.

MacArthur, R.H. & Wilson, E.O. (1967) *The Theory of Island Biogeography*. Princeton University Press, Princeton, NJ.

Mace, G.M. (2014) Whose conservation? *Science*, 345, 1558–1560.

Mace, G.M, Collar, N.J., Gaston, K.J. *et al.* (2008) Quantification of extinction risk: IUCN's system for classifying threatened species. *Conservation Biology*, 22, 1424–1442.

MacLeod, C.D. (2010) The relationship between body mass and relative investment in testes mass in cetaceans: implications for inferring interspecific variations in the extent of sperm competition. *Marine Mammal Science*, 26, 370–380.

MacLulick, D.A. (1937) Fluctuations in numbers of the varying hare (*Lepus americanus*). *University of Toronto Studies, Biology Series*, 43, 1–136.

Maestre, F.T., Valladares, F. & Reynolds, J.F. (2005) Is the change of plant–plant interactions with abiotic stress predictable? A meta-analysis of field results in arid environments. *Journal of Ecology*, 93, 748–757.

Magnacca, K.N. & Price, D.K. (2015) Rapid adaptive radiation and host plant conservation in the Hawaiian picture wing Drosophila (Diptera: Drosophilidae). *Molecular Phylogenetics and Evolution*, 92, 226–242.

Magnússon, B., Magnússon, S.H., Ólafsson, E. & Sigurdsson, B.D. (2014) Plant colonization, succession and ecosystem development on Surtsey with reference to neighbouring islands. *Biogeosciences Discussions*, 11, 9379–9420.

Maguire, L.A., Seal, U.S. & Brussard, P.F. (1987) Managing critically endangered species: the Sumatran rhino as a case study. In: *Viable Populations for Conservation* (M.E. Soulé, ed.), pp. 141–158. Cambridge University Press, Cambridge, UK.

Mahowald, N., Jickells, T.D., Baker, A.R. *et al.* (2008) Global distribution of atmospheric phosphorus sources, concentrations and deposition rates, and anthropogenic impacts. *Global Biogeochemical Cycles*, 22, GB4026.

Makkonen, M., Berg, M.P., Handa, T. *et al.* (2012) Highly consistent effects of plant litter identity and functional traits on decomposition across a latitudinal gradient. *Ecology Letters*, 15, 1033–1041.

Manne, L.L., Pimm, S.L., Diamond, J.M. & Reed, T.M. (1998) The form of the curves a direct evaluation of MacArthur & Wilson's classic theory. *Journal of Animal Ecology*, 67, 784–794.

Manzoli, D.E., Saravia-Pietropaolo, M.J., Antoniazzi, L.R. *et al.* (2018) Contrasting consequences of different defence strategies in a natural multihost–parasite system. *International Journal for Parasitology*, 48, 445–455.

Marchetti, M.P. & Moyle, P.B. (2001) Effects of flow regime on fish assemblages in a regulated California stream. *Ecological Applications*, 11, 530–539.

Marchinko, K.B., Nishizaki, M.T. & Burns, K.C. (2004) Community-wide character displacement in barnacles: a new perspective for past observations. *Ecology Letters*, 7, 114–120.

Marcos, G.M. & Lancho, J.F.G. (2002) Atmospheric deposition in oligotrophic *Quercus pyrenaica* forests: implications for forest nutrition. *Forest Ecology and Management*, 171, 17–29.

Margalida, A., Sanchez-Zapata, J.A., Blanco, G., Hiraldo, F. & Donazar, J.A. (2014) Diclofenac approval as a threat to Spanish vultures. *Conservation Biology*, 28, 631–632.

Maron, J.L., Combs, J.K. & Louda, S.M. (2002) Convergent demographic effects of insect attack on related thistles in coastal vs continental dunes. *Ecology*, 83, 3382–3392.

Maron, J.L., Marler, M., Klironomos, J.N. & Cleveland, C.C. (2011) Soil fungal pathogens and the relationship between plant diversity and productivity. *Ecology Letters*, 14, 36–41.

Marquis, R.J, Salazar, D., Baer, C., Reinhardt, J., Priest, G. & Barnett, K. (2016) Ode to Ehrlich and Raven or how herbivorous insects might drive plant speciation. *Ecology*, 97, 2939–2951.

Marshall, H.H., Carter, A.J., Ashford, A., Rowcliffe, J.M. & Cowlishaw, G. (2013) How do foragers decide when to leave a patch? A test of alternative models under natural and experimental conditions. *Journal of Animal Ecology*, 82, 894–902.

Marshall, I.D. & Douglas, G.W. (1961) Studies in the epidemiology of infectious myxomatosis of rabbits. *VIII. Further observations on changes in the innate resistance of Australian wild rabbits exposed to myxomatosis. Journal of Hygiene*, 59, 117–122.

Martens, J. & Packert, M. (2007) Ring species. Do they exist in birds? *Zoologischer Anzeiger*, 246 (2007) 315–324.

Martin, J-L. & Thibault, J-C. (1996) Coexistence in Mediterranean warblers: ecological differences or interspecific territoriality? *Journal of Biogeography*, 13, 169–178.

Martin, M.M. (1991) The evolution of cellulose digestion in insects. In: *The Evolutionary Interaction of Animals and Plant* (W.G. Chaloner, J.L. Harper & J.H. Lawton, eds), pp. 105–112. The Royal Society, London; also in *Philosophical Transactions of the Royal Society of London, Series B*, 333, 281–288.

Martin, P.R. & Martin, T.E. (2001) Ecological and fitness consequences of species coexistence: a removal experiment with wood warblers. *Ecology*, 82, 189–206.

Martinez, N.D. (1991) Artefacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecological Monographs*, 61, 367–392.

Martinez, N.D. (1993) Effects of resolution on food web structure. *Oikos*, 66, 403–412.

Martinez-Garza, C. & Howe, H.F. (2003) Restoring tropical diversity: beating the time tax on species loss. *Journal of Applied Ecology*, 40, 423–429.

Martinez-Padilla, J., Redpath, S.M, Zeineddine, M. & Mougeot, F. (2014) Insights into population ecology from long-term studies of red grouse *Lagopus lagopus scoticus*. *Journal of Animal Ecology*, 83, 85–98.

Martinson, V.G., Moy, J. & Moran, N.A. (2012) Establishment of characteristic gut bacteria during development of the honeybee worker. *Applied Environmental Microbiology*, 78, 2830–2840.

Martorell, C. & Freckleton, R.P. (2014) Testing the roles of competition, facilitation and stochasticity on community structure in a species-rich assemblage. *Journal of Ecology*, 102, 74–85.

Marzusch, K. (1952) Untersuchungen über di Temperaturabhängigkeit von Lebensprozessen bei Insekten unter besonderer Berücksichtigung winter-schlantender Kartoffelkäfer. *Zeitschrift für vergleicherde Physiologie*, 34, 75–92.

Masek, J.G., Hayes, D.J., Hughes, M.J., Healey, S.P. & Turner, D.P. (2015) The role of remote sensing in process-scaling studies of managed forest ecosystems. *Forest Ecology and Management*, 355, 109–123.

Matter, S.F., Hanski, I. & Gyllenberg, M. (2002) A test of the metapopulation model of the species–area relationship. *Journal of Biogeography*, 29, 977–983.

Matthaei, C.D. & Townsend, C.R. (2000) Long-term effects of local disturbance history on mobile stream invertebrates. *Oecologia*, 125, 119–126.

Matthews, T.J., Cottee-Jones, H.E. & Whittaker, R.J. (2014) Habitat fragmentation and the species–area relationship: a focus on total species richness obscures the impact of habitat loss on habitat specialists. *Biodiversity Research*, 20, 1136–1146.

Matthiopoulos, J., Moss, R., Mougeot, F., Lambin, X. & Redpath, S.M. (2003) Territorial behaviour and population dynamics in red grouse *Lagopus lagopus scoticus*. II. Population models. *Journal of Animal Ecology*, 72, 1083–1096.

Matthysen, E. (2012) Muliticausality of dispersal: a review. In: *Dispersal Ecology and Evolution* (J. Clobert, M. Baguette, T.G. Benton & J.M. Bullock, eds), pp. 3–18. Oxford University Press, Oxford.

Mattila, A.L.K., Duplouy, A., Kirjokangas, M., Lehtonen, R., Rastas, P. & Hanski, I. (2012) High genetic load in an old isolated butterfly population. *Proceedings of the National Academy of Sciences of the USA*, 109, E2496–E2505.

May, R.M. (1972) Will a large complex system be stable? *Nature*, 238, 413–414.

May, R.M. (1973) On relationships among various types of population models. *American Naturalist*, 107, 46–57.

May, R.M. (1975a) Biological populations obeying difference equations: stable points, stable cycles and chaos. *Journal of Theoretical Biology*, 49, 511–524.

May, R.M. (1975b) Patterns of species abundance and diversity. In: *Ecology and Evolution of Communities* (M.L. Cody & J.M. Diamond, eds), pp. 81–120. Belknap, Cambridge, MA.

May, R.M. (1976) Estimating *r*: a pedagogical note. *American Naturalist*, 110, 496–499.

May, R.M. (1978) Host–parasitoid systems in patchy environments: a phenomenological model. *Journal of Animal Ecology*, 47, 833–843.

May, R.M. (1981) Models for two interacting populations. In: *Theoretical Ecology: principles and applications*, 2nd edn (R.M. May, ed.), pp. 78–104. Blackwell Scientific Publications, Oxford.

May, R.M. (2010) Tropical arthropod species, more or less? *Science*, 329, 41–42.

May, R.M. & Anderson, R.M. (1979) Population biology of infectious diseases. *Nature*, 280, 455–461.

May, R.M. & Anderson, R.M. (1983) Epidemiology and genetics in the coevolution of parasites and hosts. *Proceedings of the Royal Society of London, Series B*, 219, 281–313.

Maynard Smith, J. (1972) *On Evolution*. Edinburgh University Press, Edinburgh.

Maynard Smith, J. & Slatkin, M. (1973) The stability of predator– prey systems. *Ecology*, 54, 384–391.

McArt, S.H., Urbanowicz, C., McCoshum, S., Irwin, R.E. & Adler, L.S. (2017) Landscape predictors of pathogen prevalence and range contractions in US bumblebees. *Proceedings of the Royal Society of London, Series B*, 284, 20172181.

McBride, P.D., Cusens, J. & Gillman, L.N. (2014) Revisiting spatial scale in the productivity–species richness relationship: fundamental issues and global change implications. *AoB Plants*, 6, plu057; doi:10.1093/aobpla/plu057.

McCalley, C.K., Woodcroft, B.J., Hodgkins, S.B. *et al.* (2014) Methane dynamics regulated by microbial community response to permafrost thaw. *Nature*, 514, 478–481.

McClain, C.R. (2004) Connecting species richness, abundance and body size in deep-sea gastropods. *Global Ecology and Biogeography*, 13, 327–334.

McCollin, D. (2017) Turnover dynamics of breeding land birds on islands: is Island Biogeographic Theory 'true but trivial' over decadal time-scales? *Diversity*, 9, doi:10.3390/d9010003.

McCormick, A.R., Hoellein, T.J., London, M.G., Hittie, J., Scott, J.W. & Kelly, J.J. (2016) Microplastic in surface waters of urban rivers: concentration, sources, and associated bacterial assemblages. *Ecosphere*, 7, e01556.

McGill, B.J., Etienne, R.S., Gray, J.S. *et al.* (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, 10, 995–1015.

McGill, B.J., Maurer, B.A. & Weiser, M.D. (2006) Empirical evaluation of neutral theory. *Ecology*, 87, 1411–1423.

McGrady-Steed, J., Harris, P.M. & Morin, P.J. (1997) Biodiversity regulates ecosystem predictability. *Nature*, 390, 162–165.

McKane, R.B., Johnson, L.C., Shaver, G.R. *et al.* (2002) Resource-based niches provide a basis for plant species diversity and dominance in arctic tundra. *Nature*, 415, 68–71.

McKay, J.K., Bishop, J.G., Lin, J.-Z., Richards, J.H., Sala, A. & Mitchell-Olds, T. (2001) Local adaptation across a climatic gradient despite small effective population size in the rare sapphire rockcress. *Proceedings of the Royal Society of London, Series B*, 268, 1715–1721.

McKay, J.K., Christian, C.E., Harrison, S. & Rice, K.J. (2005) 'How local is local?' – a review of practical and conceptual issues in the genetics of restoration. *Restoration Ecology*, 13, 432–440.

McKean, K.A. & Nunney, L. (2001) Increased sexual activity reduces male immune function in Drosophila melanogaster. *Proceedings of the National Academy of the USA*, 98, 7904–7909.

McKean, K.A., Yourth, C.P., Lazzaro, B.P. & Clark, A.G. (2008) The evolutionary costs of immunological maintenance and deployment. *BMC Evolutionary Biology*, 8, 76.

McKenzie, N.L. & Rolfe, J.K. (1986) Structure of bat guilds in the Kimberley mangroves, Australia. *Journal of Animal Ecology*, 55, 401–420.

McKinney, M.A., Iverson, S.J., Fisk, A.T. *et al.* (2013) Global change effects on the long-term feeding ecology and contaminant exposures of East Greenland polar bears. *Global Change Biology*, 19, 2360–2372.

McKone, M.J., Kelly, D. & Lee, W.G. (1998) Effect of climate change on mast-seeding species: frequency of mass flowering and escape from specialist insect seed predators. *Global Change Biology*, 4, 591–596.

McNaughton, S.J. (1976) Serengeti migratory wildebeest: facilitation of energy flow by grazing. *Science*, 191, 92–94.

McNickle, G.G. & Cahill, J.F. Jr (2009) Plant root growth and the marginal value theorem. *Proceedings of the National Academy of Sciences of the USA*, 106, 4747–4751.

Mduma, S.A.R., Sinclair, A.R.E. & Hilborn, R. (1999) Food regulates the Serengeti wildebeest: a 40 year record. *Journal of Animal Ecology*, 68, 1101–1122.

MEA (Millennium Ecosystem Assessment) (2005a) *Ecosystems and Human Well-being: Synthesis*. Island Press, Washington, DC.

MEA (Millennium Ecosystem Assessment) (2005b) *Living Beyond Our Means: natural assets and human well-being: biodiversity synthesis*. World Resources Institute, Washington, DC.

Medel, R., Vergara, E., Silva, A. & Kalin-Arroyo, M. (2004) Effects of vector behavior and host resistance on mistletoe aggregation. *Ecology*, 85, 12–126.

Médoc, V., Albert, H. & Spataro, T. (2015) Functional response comparisons among freshwater amphipods: ratio-dependence and higher predation for *Gammarus pulex* compared to the non-natives *Dikerogammarus villosus* and *Echinogammarus berilloni*. *Biological Invasions*, 17, 3625–3637.

Melbourne, B.A., Cornell, H.V., Davies, K.F. *et al.* (2007) Invasion in a heterogeneous world: resistance, coexistence or hostile takeover? *Ecology Letters*, 10, 77–94.

Melbourne, B.A. & Hastings, A. (2008) Extinction risk depends strongly on factors contributing to stochasticity. *Nature*, 454, 100–103.

Menge, B.A., Lubchenco, J., Gaines, S.D. & Ashkenas, L.R. (1986) A test of the Menge–Sutherland model of community organization in a tropical rocky intertidal food web. *Oecologia*, 71, 75–89.

Menge, B.A. & Sutherland, J.P. (1976) Species diversity gradients: synthesis of the roles of predation, competition, and temporal heterogeneity. *American Naturalist*, 110, 351–369.

Menges, E.S. & Dolan, R.W. (1998) Demographic viability of populations of *Silene regia* in midwestern prairies: relationships with fire management, genetic variation, geographic location, population size and isolation. *Journal of Ecology*, 86, 63–78.

Merino, G., Barange, M., Rodwell, L. & Mullon, C. (2011) Modelling the sequential geographical exploitation and potential collapse of marine fisheries through economic globalization, climate change and management alternatives. *Scientia Marina*, 75, 779–790.

Merow, C., Dahlgren, J.P., Metcalf, C.J.E. *et al.* (2014) Advancing population ecology with integral projection models: a practical guide. *Methods in Ecology and Evolution*, 5, 99–110.

Merow, C., Treanor Bois, S., Allen, J.M., Xie, Y. & Silander, J.A. Jr (2016) Climate change both facilitates and inhibits invasive plant ranges in New England. *Proceedings of the National Academy of Science*, 114, E3276–E3284.

Meshram, A.R., Vader, A., Kristiansen, S. & Gabrielsen, T.M. (2017) Microbial eukaryotes in an Arctic under-ice spring bloom north of Svarlbard. *Frontiers in Microbiology*, 8, 1099.

Meyer, A.L.S. & Pie, M.R. (2018) Environmental prevalence and the distribution of species richness across climatic niche space. *Journal of Biogeography*, 45, 2348–2360.

Micheli, F., Peterson, C.H., Mullineaux, L.S. *et al.* (2002) Predation structures communities at deep-sea hydrothermal vents. *Ecological Monographs*, 72, 365–382.

Middlemas, S.J., Barton, T.R., Armstrong, J.D. & Thompson, P.M. (2006) Functional and aggregative responses of harbor seals to changes in salmonid abundance. *Proceedings of the Royal Society of London, Series B*, 273, 193–198.

Middleton, B.A. (2003) Soil seed banks and the potential restoration of forested wetlands after farming. *Journal of Applied Ecology*, 40, 1025–1034.

Milinski, M. & Parker, G.A. (1991) Competition for resources. In: *Behavioural Ecology: an evolutionary approach*, 3rd edn (J.R. Krebs & N.B. Davies, eds), pp. 137–168. Blackwell Scientific Publications, Oxford.

Mills, L.S., Soule, M.E. & Doak, D.F. (1993) The keystonespecies concept in ecology and conservation. *Bioscience*, 43, 219–224.

Milner-Gulland, E.J. & Mace, R. (1998) *Conservation of Biological Resources*. Blackwell Science, Oxford.

Milton, Y. & Kaspari, M. (2007) Bottom-up and top-down regulation of decomposition in a tropical forest. *Oecologia*, 153, 163–172.

Mirceta, S., Signore, A.V., Burns, J.M., Cossins, A.R., Campbell, K.L. & Berenbrink, M. (2013) Evolution of mammalian diving capacity traced by myoglobin net surface charge. *Science*, 340, 1303–1311.

Mistri, M. (2003) Foraging behaviour and mutual interference in the Mediterranean shore crab, *Carcinus aestuarii*, preying upon the immigrant mussel *Musculista senhousia*. *Estuarine, Coastal and Shelf Science*, 56, 155–159.

Moeslund, J.E., Brunbjerg, A.K., Clausen, K.K. *et al.* (2017) Using dark diversity and plant characteristics to guide conservation and restoration. *Journal of Applied Ecology*, doi, 10.1111/1365–2664.12867.

Mojica, J.P., Lee, Y.W., Willis, J.H. & Kelly, J.K. (2012) Spatially and temporally varying selection on intrapopulation quantitative trait loci for a life history trade-off in *Mimulus guttatus*. *Molecular Ecology*, 21, 3718–3728.

Moles, A.T. & Westoby, M. (2006) Seed size and plant strategy across the whole life cycle. *Oikos*, 113, 91–105.

Monastersky, R. (2015) The human age. *Nature*, 519, 144–147.

Monks, A. & Kelly, D. (2006) Testing the resource-matching hypothesis in the mast seeding tree Nothofagus truncata (Fagaceae). *Austral Ecology*, 31, 366–375.

Montagnes, D.J.S., Kimmance, S.A. & Atkinson, D. (2003) Using Q10, can growth rates increase linearly with temperature? *Aquatic Microbial Ecology*, 32, 307–313.

Mooney, H.A. & Gulmon, S.L. (1979) Enironmental and evolutionary constraints on the photosynthetic pathways of higher plants. In: *Topics in Plant Population Biology* (O.T. Solbrig, S. Jain, G.B. Johnson & P.H. Raven, eds), pp. 316–337. Columbia University Press, New York.

Mooney, H.A., Vitousek, P.M. & Matson, P.A. (1987) Exchange of materials between terrestrial ecosystems and the atmosphere, *Science*, 238, 926–932.

Moore, B.D., Andrew, R.L., Külheim, C. & Foley, W.J. (2013) Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *New Phytologist*, 201, 733–750.

Mora, C. & Sale, P.F. (2011) Ongoing global biodiversity loss and the need to move beyond protected areas: a review of the technical and practical shortcomings of protected areas on land and sea. *Marine Ecology Progress Series*, 434, 251–266.

Moran, N.A. (2007) Symbiosis as an adaptive process and source of phenotypic complexity. *Proceedings of the National Academy of Sciences of the USA*, 104(suppl 1), 8627–8633.

Moran, N.A., Munson, M.A., Baumann, P. & Ishikawa, H. (1993) A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proceedings of the Royal Society of London, Series B*, 253, 167–171.

Moreira, X., Abdala-Roberts, L. & Castagneyrol, B. (2018) Interactions between plant defence signalling pathways: evidence from bioassays with insect herbivores and plant pathogens. *Journal of Ecology*, 106, 2353–2364.

Moreno, R.S., Kays, R.W. & Samudio, R. Jr (2006) Competitive release in diets of ocelot (*Leopardus pardalis*) and puma (*Puma concolor*) after jaguar (*Panthera onca*) decline. *Journal of Mammalogy*, 87, 808–816.

Mori, S., Yamaji, K., Ishida, A. *et al.* (2010) Mixed-power scaling of whole-plant respiration from seedlings to giant trees. *Proceedings of the National Academy of Sciences of the USA*, 107, 1447–1451.

Morin, P.J., Lawler, S.P. & Johnson, E.A (1988) Competition between aquatic insects and vertebrates: interaction strength and higher order interactions. *Ecology*, 69, 1401–1409.

Morris, C.E. (2002) Self-thinning lines differ with fertility level. *Ecological Research*, 17, 17–28.

Morris, D.W. & Davidson, D.L. (2000) Optimally foraging mice match patch use with habitat differences in fitness. *Ecology*, 81, 2061–2066.

Morris, J.R., Vandermeer, J. & Perfecto, I. (2015) A keystone ant species provides robust biological control of the coffee berry borer under varying pest densities. *PLoS One*, 10, e0142850.

Morris, R.F. (1959) Single-factor analysis in population dynamics. *Ecology*, 40, 580–588.

Morrison, G. & Strong, D.R. Jr (1981) Spatial variations in egg density and the intensity of parasitism in a neotropical chrysomelid (*Cephaloleia consanguinea*). *Ecological Entomology*, 6, 55–61.

Morrissey, C.A., Mineau, P., Devries, J.H. *et al.* (2015) Neonicotinoid contamination of global surface waters and associated risk to aquatic invertebrates: a review. *Environment International*, 74, 291–303.

Morrow, P.A. & Olfelt, J.P. (2003) Phoenix clones: recovery after longterm defoliation-induced dormancy. *Ecology Letters*, 6, 119–125.

Mortensen, H.K., Dupont, Y.L. & Olesen, J.M. (2008) A snake in paradise: disturbance of plant reproduction following extirpation of bird flower-visitors on Guam. *Biological Conservation*, 141, 2146–2154.

Moss, B. (1989) *Ecology of Fresh Waters: man and medium*, 2nd edn. Blackwell Scientific Publications, Oxford.

Moss, G.D. (1971) The nature of the immune response of the mouse to the bile duct cestode, *Hymenolepis microstoma*. *Parasitology*, 62, 285–294.

Moss, R. & Watson, A. (2001) Population cycles in birds of the grouse family (Tetraonidae). *Advances in Ecological Research*, 32, 53–111.

Motisi, N., Poggi, S., Filipe, J.A.N. *et al.* (2013) Epidemiological analysis of the effects of biofumigation for biological control of root rot in sugar beet. *Plant Pathology*, 62, 69–78.

Motta, E.V.S., Raymann, K. & Moran, N.A. (2018) Glyphosate perturbs the gut microbiota of honey bees. *Proceeding of the National Academy of Sciences of the USA*, 115(41), 10305–10310.

Mougeot, F., Evans, S.A. & Redpath, S.M. (2005) Interactions between population processes in a cyclic species: parasites reduce autumn territorial behaviour of male red grouse. *Oecologia*, 144, 289–298.

Mougeot, F., Redpath, S.M., Leckie, F. & Hudson, P.J. (2003a) The effect of aggressiveness on the population dynamics of a territorial bird. *Nature*, 421, 737–739.

Mougeot, F., Redpath, S.M., Moss, R., Matthiopoulos, J. & Hudson, P.J. (2003b) Territorial behaviour and population dynamics in red grouse, *Lagopus lagopus scoticus. I. Population experiments*. *Journal of Animal Ecology*, 72, 1073–1082.

Mouquet, N. & Loreau, M. (2003) Community patterns in source-sink metacommunities. *American Naturalist*, 162, 544–557.

Mouritsen, K.N. (2002) The parasite-induced surfacing behaviour in the cockle *Austrovenus stutchburyi*: a test of an alternative hypothesis and identification of potential mechanisms. *Parasitology*, 124, 521–528.

Mueller, K.E., Tilman, D., Fornara, D. & Hobbie, S.E. (2013) Root depth distribution and the diversity-productivity relationship in a long-term grassland experiment. *Ecology*, 94, 787–793.

Mulder, C. & Elser, J.J. (2009) Soil acidity, ecological stoichiometry and allometric scaling in grassland food webs. *Global Change Biology*, 15, 2730–2738.

Müller, H.J. (1970) Food distribution, searching success and predator–prey models. In: *The Mathematical Theory of the Dynamics of Biological Populations* (R.W. Hiorns, ed.), pp. 87–101. Academic Press, London.

Muller, M.M., Varama, M., Heinonen, J. & Hallaksela, A. (2002) Influence of insects on the diversity of fungi in decaying spruce wood in managed and natural forests. *Forest Ecology and Management*, 166, 165–181.

Murdie, G. & Hassell, M.P. (1973) Food distribution, searching success and predator–prey models. In: *The Mathematical Theory of the Dynamics of Biological Populations* (R.W. Hiorns, ed.), pp. 87–101. Academic Press, London.

Murdoch, W.W. (1966) Community structure, population control and competition – a critique. *American Naturalist*, 100, 219–226.

Murdoch, W.W., Avery, S. & Smith, M.E.B. (1975) Switching in predatory fish. *Ecology*, 56, 1094–1105.

Murdoch, W.W., Briggs, C.J. & Nisbet, R.M. (2003) *Consumer-Resource Dynamics*. Princeton University Press, Princeton NJ.

Murdoch, W.W. & Stewart-Oaten, A. (1989) Aggregation by parasitoids and predators: effects on equilibrium and stability. *American Naturalist*, 134, 288–310.

Muscarella, R. & Fleming, T.H. (2007) The role of frugivorous bats in tropical forest succession. *Biological Reviews,* 82, 573–590.

Muscatine, L. & Pool, R.R. (1979) Regulation of numbers of intracellular algae. *Proceedings of the Royal Society of London, Series B*, 204, 115–139.

Mwendera, E.J., Saleem, M.A.M & Woldu, Z. (1997) Vegetation response to cattle grazing in the Ethiopian Highlands. *Agriculture, Ecosystems and Environment*, 64, 43–51.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, 403, 853–858.

Naeem, S., Thompson, L.J., Lawler, S.P., Lawton, J.H. & Woodfin, R.M. (1995) Empirical evidence that declining species diversity may alter the performance of terrestrial ecosystems. *Philosophical Transactions of the Royal Society of London B*, 347, 249–262.

Nagy, E.S. & Rice, K.J. (1997) Local adaptation in two subspecies of an annual plant: implications for migration and gene flow. *Evolution*, 51, 1079–1089.

Nalepa, C.A., Bignell, D.E. & Bandi, C. (2001) Detritivory, coprophagy, and the evolution of digestive mutualisms in Dictyoptera. *Insectes Sociaux*, 48, 194–201.

Nathan, R., Getz, W.M., Revilla, E. *et al.* (2008) A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences of the USA*, 105, 19052–19059.

Neff, J.C., Holland, E.A., Dentener, F.J., McDowell, W.H. & Russell, K.M. (2002) The origin, composition and rates of organic nitrogen deposition: a missing piece of the nitrogen cycle? *Biogeochemistry*, 57/58, 99–136.

Neilson, E.H., Goodger, J.Q.D., Woodrow, I.E. & Møller, B.L. (2013) Plant chemical defense: at what cost? *Trends in Plant Science*, 18, 250–258.

Nesbitt, H.K. & Moore, J.W. (2016) Species and population diversity in Pacific salmon fisheries underpin indigenous food security. *Journal of Applied Ecology*, 53, 1489–1499.

Neubert, M.G. & Caswell, H. (2000) Demography and dispersal: calculation and sensitivity analysis of invasion speed for structured populations. *Ecology*, 81, 1613–1628.

Neumann, A., Werner, J. & Rauber, R. (2009) Evaluation of yield–density relationships and optimization of intercrop compositions of field-grown pea–oat intercrops using the replacement series and the response surface design. *Field Crops Research*, 114, 284–294.

Neuner, G. (2014) Frost resistance in alpine woody plants. *Frontiers in Plant Science*, 5, 654. doi: 10.3389/fpls.2014.00654.

Neutel, A-M., Heesterbeek, J.A.P. & de Ruiter, P.C. (2002) Stability in real food webs: weak links in long loops. *Science*, 296, 1120–1123.

Newey, P.S., Robson, S.K.A. & Crozier, R.H. (2010) Weaver ants *Oecophylla smaragdina* encounter nasty neighbors rather than dear enemies. *Ecology*, 91, 2366–2372.

Newton, I. & Rothery, P. (1997) Senescence and reproductive value in sparrowhawks. *Ecology*, 78, 1000–1008.

Nicholson, A.J. (1954) An outline of the dynamics of animal populations. *Australian Journal of Zoology*, 2, 9–65.

Nicholson, A.J. & Bailey, V.A. (1935) The balance of animal populations. *Proceedings of the Zoological Society of London*, 3, 551–598.

Niedziela, C.E. Jr, Nelson, P.V. & Dickey, D.A. (2015) Growth, development, and mineral nutrient accumulation and distribution. *International Journal of Agronomy*, 2015, 341287.

Nilsson, L.A. (1988) The evolution of flowers with deep corolla tubes. *Nature*, 334, 147–149.

NOAA (National Oceanic and Atmospheric Administration) (2013) World Ocean Atlas. https://www.nodc.noaa.gov/OC5/indprod.html (last accessed November 2019).

Noble, J.C. & Slatyer, R.O. (1981) Concepts and models of succession in vascular plant communities subject to recurrent fire. In: *Fire and the Australian Biota* (A.M. Gill, R.H. Graves & I.R. Noble, eds). Australian Academy of Science, Canberra.

Noormets, A., Epron, D., Domec, J.C. *et al.* (2015) Effects of forest management on productivity and carbon sequestration: a review and hypothesis. *Forest Ecology and Management*, 355, 124–140.

Norambuena, H. & Piper, G.L. (2000) Impact of *Apion ulicis* Forster on *Ulex europaeus* L. seed dispersal. *Biological Control*, 17, 267–271.

Northrop, A.C., Brooks, R.K., Ellison, A.M., Gotelli, N.J. & Ballif, B.A. (2017) Environmental proteomics reveals taxonomic and functional changes in an enriched aquatic ecosystem. *Ecosphere*, 8, e01954.

Norton, I.O. & Sclater, J.G. (1979) A model for the evolution of the Indian Ocean and the breakup of Gondwanaland. *Journal of Geophysical Research*, 84, 6803–6830.

Novotny, V. & Basset, Y. (2005) Host specificity of insect herbivores in tropical forests. *Proceedings of the Royal Society of London, Series B*, 272, 1083–1090.

Noy-Meir, I. (1975) Stability of grazing systems: an application of predator–prey graphs. *Journal of Ecology*, 63, 459–483.

Obeid, M., Machin, D. & Harper, J.L. (1967) Influence of density on plant to plant variations in fiber flax, *Linum usitatissimum*. *Crop Science*, 7, 471– 473.

O'Brien, E.M. (2006) Biological relativity to water-energy dynamics. *Journal of Biogeography*, 33, 1868–1888.

O'Donnell, K.M., Messerman, A.F., Barichivich, W.J. *et al.* (2017) Structured decision making as a conservation tool for recovery planning of two endangered salamanders. *Journal for Nature Conservation*, 37, 66–72.

Odum, E.P. & Biever, L.J. (1984) Resource quality, mutualism, and energy partitioning in food chains. *American Naturalist*, 124, 360–376.

Ødum, S. (1965) Germination of ancient seeds; floristical observations and experiments with archaeologically dated soil samples. *Dansk Botanisk Arkiv*, 24, 1–70.

Oedekoven, M.A. & Joern, A. (2000) Plant quality and spider predation affects grasshoppers (Acrididae): food-quality-dependent compensatory mortality. *Ecology*, 81, 66–77.

Oinonen, E. (1967) The correlation between the size of Finnish bracken (*Pteridium aquilinum* (L.) Kuhn) clones and certain periods of site history. *Acta Forestalia Fennica*, 83, 3–96.

Oksanen, L., Fretwell, S., Arruda, J. & Niemela, P. (1981) Exploitation ecosystems in gradients of primary productivity. *American Naturalist*, 118, 240–261.

Oksanen, T.A., Koskela, E. & Mappes, T. (2002) Hormonal manipulation of offspring number: maternal effort and reproductive costs. *Evolution*, 56, 1530–1537.

Oksanen, T., Oksanen, L., Dahlgren, J. & Olofsson, J. (2008) Arctic lemmings, *Lemmus* spp. and *Dicrostonyx* spp.: integrating ecological and evolutionary perspectives. *Evolutionary Ecology Research*, 10, 415–434.

Oli, M.K., Loughry, W.J., Caswell, H., Perez-Heydrich, C., McDonough, C.M. & Truman, R.W. (2017) Dynamics of leprosy in nine-banded armadillos: net reproductive number and effects on host population dynamics. *Ecological Modelling*, 350, 100–108.

Oliver, K.M., Smith, A.H. & Russell, J.A. (2014) Defensive symbiosis in the real world – advancing ecological studies of heritable, protective bacteria in aphids and beyond. *Functional Ecology*, 28, 341–355.

Oliver, R.P. & Hewitt, H.G. (2014) *Fungicides in Crop Protection*, 2nd edn. CABI, Wallingford.

O'Meara, B.C. (2012) Evolutionary inferences from phylogenies: a review of methods. *Annual Review of Ecology, Evolution and Systematics*, 43, 267–285.

Orcutt, B.N., Sylvan, J.B., Knab, N.J. & Edwards, K.J. (2011) Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiology and Molecular Biology Reviews*, 75, 361–422.

Orshan, G. (1963) Seasonal dimorphism of desert and Mediterranean chamaephytes and its significance as a factor in their water economy. In: *The Water Relations of Plants* (A.J. Rutter & F.W. Whitehead, eds), pp. 207–222. Blackwell Scientific Publications, Oxford.

Osawa, A. & Allen, R.B. (1993) Allometric theory explains selfthinning relationships of mountain beech and red pine. *Ecology*, 74, 1020–1032.

Osem, Y., Perevolotsky, A. & Kigel, J. (2002) Grazing effect on diversity of annual plant communities in a semi-arid rangeland: interactions with small-scale spatial and temporal variation in primary productivity. *Journal of Ecology*, 90, 936–946.

Ostfeld, R.S. & Keesing, F. (2000) Biodiversity and disease risk: the case of Lyme disease. *Conservation Biology*, 14, 722–728.

Otten, W., Filipe, J.A.N., Bailey, D.J. & Gilligan, C.A. (2003) Quantification and analysis of transmission rates of soilborne epidemics. *Ecology*, 84, 3232–3239.

Otten, W., Filipe, J.A.N. & Gilligan, C.A. (2005) Damping-off epidemics, contact structure, and disease transmission in mixed-species populations. *Ecology*, 86, 1948–1957.

Ottersen, G., Planque, B., Belgrano, A., Post, E., Reid, P.C. & Stenseth, N.C. (2001) Ecological effects of the North Atlantic Oscillation. *Oecologia*, 128, 1–14.

Ovando, D.A., Deacon, R.T., Lester, S.E. *et al.* (2013) Conservation incentives and collective choices in cooperative fisheries. *Marine Policy*, 37, 132–140.

Ovaskainen, O. & Meerson, B. (2010) Stochastic models of population extinction. *Trends in Ecology and Evolution*, 25, 643–652.

Pacala, S.W. & Crawley, M.J. (1992) Herbivores and plant diversity. *American Naturalist*, 140, 243–260.

Pacala, S.W. & Hassell, M.P. (1991) The persistence of host–parasitoid associations in patchy environments. II. Evaluation of field data. *American Naturalist*, 138, 584–605.

Pace, M.L., Cole, J.J., Carpenter, S.R. & Kitchell, J.F. (1999) Trophic cascades revealed in diverse ecosystems. *Trends in Ecology and Evolution*, 14, 483–488.

Paine, R.T. (1966) Food web complexity and species diversity. *American Naturalist*, 100, 65–75.

Paine, R.T. (1969) A note on trophic complexity and community stability. *The American Naturalist*, 103, 91–93.

Paine, R.T. (1979) Disaster, catastrophe and local persistence of the sea palm *Postelsia palmaeformis*. *Science*, 205, 685–687.

Paine, R.T. (1980) Food webs: linkage, interaction strength, and community infrastructure. *Journal of Animal Ecology*, 49, 667–685.

Palmblad, I.G. (1968) Competition studies on experimental populations of weeds with emphasis on the regulation of population size. *Ecology*, 49, 26–34.

Palmer, T.M., Young, T.P., Stanton, M.L. & Wenk, E. (2000) Short term dynamics of an acacia ant community in Laikipia, Kenya. *Oecologia*, 123, 425–435.

Palmqvist, K. (2000) Carbon economy in lichens. *New Phytologist*, 148, 11–36.

Palter, J. (2015) Storms bring ocean nutrients to light. *Nature*, 525, 460–461.

Papadopulos, A.S.T., Baker, W.J., Crayn, D. *et al.* (2011) Speciation with gene flow on Lord Howe Island. *Proceedings of the National Academy of Sciences of the USA*, 108, 13188–13193.

Park, T. (1948) Experimental studies of interspecific competition. I. Competition between populations of the flour beetle *Tribolium confusum* Duval and *Tribolium castaneum* Herbst. *Ecological Monographs*, 18, 267–307.

Park, T. (1954) Experimental studies of interspecific competiton. II. Temperature, humidity and competition in two species of *Tribolium*. *Physiological Zoology*, 27, 177–238.

Park, T. (1962) Beetles, competition and populations. *Science*, 138, 1369–1375.

Parker, G.A. (1970) The reproductive behaviour and the nature of sexual selection in *Scatophaga stercorraria* L. (Diptera: Scatophagidae) II. The fertilization rate and the spatial and temporal relationships of each sex around the site of mating and oviposition. *Journal of Animal Ecology*, 39, 205–228.

Parker, G.A. (1984) Evolutionarily stable strategies. In: *Behavioral Ecology: an evolutionary approach*, 2nd edn (J.R. Krebs & N.B. Davies, eds), pp. 30–61. Blackwell Scientific Publications, Oxford.

Parker, G.A. & Stuart, R.A. (1976) Animal behaviour as a strategy optimizer: evolution of resource assessment strategies and optimal emigration thresholds. *American Naturalist*, 110, 1055–1076.

Passarge, J., Hol, S., Escher, M. & Huisman, J. (2006) Competition for nutrients and light: stable coexistence, alternative stable states, or competitive exclusion? *Ecological Monographs*, 76, 57–72.

Paterson, S. & Viney, M.E. (2002) Host immune responses are necessary for density dependence in nematode infections. *Parasitology*, 125, 283–292.

Pauly, D. & Christensen, V. (1995) Primary production required to sustain global fisheries. *Nature*, 374, 255–257.

Pauly, D., Christensen, V., Dalsgaard, J., Froese, R. & Torres, F. Jr (1998) Fishing down marine food webs. *Science*, 279, 860–863.

Pavia, H. & Toth, G.B. (2000) Inducible chemical resistance to herbivory in the brown seaweed *Ascophyllum nodosum*. *Ecology*, 81, 3212–3225.

Payne, N.L., Smith, J.A., van der Meulen, D.E. *et al.* (2016) Temperature dependence of fish performance in the wild: links with species biogeography and physiological thermal tolerance. *Functional Ecology*, 30, 903–912.

Peay, K.G., Garbelotto, M. & Bruns, T.D. (2010) Evidence of dispersal limitation in soil microorganisms: isolation reduces species richness on mycorrhizal tree islands. *Ecology* 91, 3631–3640.

Pedersen, A.B., Altizer, S., Poss, M. Cunningham, A.A. & Nunn, C.L. (2005) Patterns of host specificity and transmission among parasites of wild primates. *International Journal for Parasitology*, 35, 647–657.

Pedersen, A.B. & Fenton, A. (2006) Emphasizing the ecology in parasite community ecology. *Trends in Ecology and Evolution*, 22, 133–139.

Pedersen, A.B. & Grieves, T.J. (2008) The interaction of parasites and resources cause crashes in a wild mouse population. *Journal of Animal Ecology*, 77, 370–377.

Pellmyr, A. (2002) Yuccas, Yucca moths, and coevolution: a review. *Annals of the Missouri Botanical Garden*, 90, 35–55.

Pellmyr, O. & Leebens-Mack, J. (1999) Reversal of mutualism as a mechanism for adaptive radiation in yucca moths. *American Naturalist*, 156, S62–S76.

Penn, A.M., Sherwin, W.B., Gordon, G., Lunney, D., Melzer, A. & Lacy, R.C. (2000) Demographic forecasting in koala conservation. *Conservation Biology*, 14, 629–638.

Pennings, S.C. & Callaway, R.M. (2002) Parasitic plants: parallels and contrasts with herbivores. *Oecologia*, 131, 479–489.

Perälä, T. & Kuperinen, A. (2017) Detection of Allee effects in marine fishes: analytical biases generated by data availability and model selection. *Proceedings of the Royal Society of London, Series B*, 284, 20171284.

Pereira, R.J & Wake, D.B. (2015) Ring species as demonstrations of the continuum of species formation. *Molecular Ecology*, 24, 5312–5314.

Perez-Pimienta, J., Lopez-Ortega, M.G. & Sanchez, A. (2017) Recent developments in agave performance as a drought-tolerant biofuel feedstock: agronomics, characterization, and biorefining. *Biofuels, Bioproducts and Biorefining*, 11, 732–748.

Perlman, S.J. & Jaenike, J. (2001) Competitive interactions and persistence of two nematode species that parasitize *Drosophila recens*. *Ecology Letters*, 4, 577–584.

Perry, J.N., Smith, R.H., Woiwod, I.P. & Morse, D.R. (eds) (2000) *Chaos in Real Data*. Kluwer, Dordrecht.

Pertierra, L.R., Aragon, P., Shaw, J.D., Bergstrom, D.M., Teraud, A. & Olalla-Tarraga, M.A. (2017) Global thermal niche models of two European grasses show high invasion risks in Antarctica. *Global Change Biology*, 23, 2863–2873.

Petchey, O.W. & Belgrano, A. (2010) Body-size distributions and size-spectra: universal indicators of ecological status? *Biology Letters*, 6, 434–437.

Peters, R.H. (1983) *The Ecological Implications of Body Size*. Cambridge University Press, Cambridge, UK.

Peterson, A.T. (2003) Predicting the geography of species' invasion via ecological niche modeling. *Quarterly Review of Biology*, 78, 419–433.

Petit, J.R., Jouzel, J., Raynaud, D. *et al.* (1999) Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature*, 399, 429–436.

Petranka, J.W. (1989) Chemical interference competition in tadpoles: does it occur outside laboratory aquaria? *Copeia*, 1989, 921–930.

Petren, K. & Case, T.J. (1996) An experimental demonstration of exploitation competition in an ongoing invasion. *Ecology*, 77, 118–132.

Petren, K., Grant, B.R. & Grant, P.R. (1999) A phylogeny of Darwin's finches based on microsatellite DNA variation. *Proceedings of the Royal Society of London, Series B*, 266, 321–329.

Pettifor, R.A., Perrins, C.M. & McCleery, R.H. (2001) The individual optimization of fitness: variation in reproductive output, including clutch size, mean nestling mass and offspring recruitment, in manipulated broods of great tits *Parus major*. *Journal of Animal Ecology*, 70, 62–79.

Pianka, E.R. (1970) On *r*- and *k*-selection. *American Naturalist*, 104, 592–597.

Pianka, E.R. (1973) The structure of lizard communities. *Annual Review of Ecology and Systematics*, 4, 53–74.

Pickett, S.T.A., Cadenasso, M.L. & Meiners, S.J. (2008) Ever since Clements: from succession to vegetation dynamics and understanding to intervention. *Applied Vegetation Science*, 12, 9–21.

Pickett, S.T.A. & White, P.S. (eds) (1985) *The Ecology of Natural Disturbance as Patch Dynamics*. Academic Press, New York.

Piedade, M.T.F., Junk, W.J. & Long, S.P. (1991) The productivity of the C4 grass *Echinochloa polystachya* on the Amazon floodplain. *Ecology*, 72, 1456–1463.

Pierce, S., Negreiros, D., Cerabolini, B.E.L. *et al.* (2017) A global method for calculating plant CSR ecological strategies applied across biomes world-wide. *Functional Ecology*, 31, 444–457.

Piertney, S.B., Lambin, X., MacColl, A.D.C. *et al.* (2008) Temporal changes in kin structure through a population cycle in a territorial bird, the red grouse *Lagopus lagopus scoticus*. *Molecular Ecology*, 17, 2544–2551.

Pimentel, D., Lach, L., Zuniga, R. & Morrison, D. (2000) Environmental and economic costs of nonindigenous species in the United States. *BioScience*, 50, 53–65.

Pimm, S.L. (1982) *Food Webs*. Chapman & Hall, London.

Pimm, S.L. & Lawton, J.H. (1977) The number of trophic levels in ecological communities. *Nature*, 275, 542–544.

Pinotti, B.T., Pagotto, C.P. & Pardini, R. (2015) Wildlife recovery during tropical forest succession: assessing ecological drivers of community change. *Biotropica*, 47, 765–774.

Piper, J.K. (2007) Does the number of species in the seed mix affect the establishment of four tallgrass prairie species? A seven-year study in Kansas. *Ecological Restoration*, 25, 118–122.

Piraino, S., Fanelli, G. & Boero, F. (2002) Variability of species' roles in marine communities: change of paradigms for conservation priorities. *Marine Biology*, 140, 1067–1074.

Pita, L., Rix, L., Slaby, B.M., Franke, A. and Hentschel, U. (2018) The sponge holobiont in a changing ocean: from microbes to ecosystems. *Microbiome*, 6, 46.

Pitcher, T.J. & Hart, P.J.B. (1982) *Fisheries Ecology*. Croom Helm, London.

Podoler, H. & Rogers, D.J. (1975) A new method for the identification of key factors from life-table data. *Journal of Animal Ecology*, 44, 85–114.

Polis, G.A., Sears, A.L.W., Huxel, G.R., Strong, D.R. & Maron, J. (2000) When is a trophic cascade a trophic cascade? *Trends in Ecology and Evolution*, 15, 473–475.

Polis, G.A. & Strong, D. (1996) Food web complexity and community dynamics. *American Naturalist*, 147, 813–846.

Pollack, L. & Rubenstein, D.R. (2015) The fitness consequences of kin-biased dispersal in a cooperatively breeding bird. *Biology Letters*, 11, 20150336.

Ponisio, L.C., M'Gonigle, L.K., Mace, K.C., Palomino, J., Valpine, P. & Kremen, C. (2014) Diversification practices reduce organic to conventional yield gap. *Proceedings of the Royal Society of London, Series B*, 282, 20141396.

Poorter, H., Jagodzinski, A.M., Ruiz-Peinado, R. *et al.* (2015) How does biomass distribution change with size and differ among species? An analysis for 1200 plant species from five continents. *New Phytologist*, 208, 736–749.

Pope, S.E., Fahrig, L. & Merriam, H.G. (2000) Landscape complementation and metapopulation effects on leopard frog populations. *Ecology*, 81, 2498–2508.

Population Reference Bureau (2006) http://www.prb.org/Publications/GraphicsBank/PopulationTrends.aspx.

Pörtner, H. (2001) Climate change and temperature-dependent biogeography: oxygen limitation of thermal tolerance in animals. *Naturwissenschaften*, 88, 137–146.

Post, D.M. (2002) The long and the short of food-chain length. *Trends in Ecology and Evolution*, 17, 269–277.

Post, D.M., Pace, M.L. & Hairston, N.G. Jr (2000) Ecosystem size determines food-chain length in lakes. *Nature*, 405, 1047–1049.

Poulin, R. & Keeney, D.B. (2007) Host specificity under molecular and experimental scrutiny. *Trends in Parasitology*, 24, 24–28.

Poulson, M.E. & DeLucia, E.H. (1993) Photosynthetic and structural acclimation to light direction in vertical leaves of *Silphium terebinthaceum*. *Oecologia*, 95, 393–400.

Pounds, A.P., Bustamante, M.R. Coloma, L.A. *et al.* (2006) Widespread amphibian extinctions from epidemic disease driven by global warming. *Nature*, 439, 161–167.

Pourmokhtarian, A., Driscoll, C.T., Campbell, J.L. *et al.* (2017) Modeled ecohydrological responses to climate change at seven small watersheds in the northeastern United States.

Global Change Biology, 23, 840–856.

Power, M.E., Tilman, D., Estes, J.A. *et al.* (1996) Challenges in the quest for keystones. *Bioscience*, 46, 609–620.

Preisser, E.L., Bolnick, D.I. & Benard, M.F. (2005) Scared to death? The effects of intimidation and consumption in predator–prey interactions. *Ecology*, 86, 501–509.

Preston, F.W. (1962) The canonical distribution of commoness and rarity. *Ecology*, 43, 185–215, 410–432.

Preszler, R.W. & Price, P.W. (1993) The influence of *Salix* leaf abscission on leaf-miner survival and life-history. *Ecological Entomology*, 18, 150–154.

Primack, R.B. (1993) *Essentials of Conservation Biology*. Sinauer Associates, Sunderland, MA.

Prior, H. & Johnes, P.J. (2002) Regulation of surface water quality in a cretaceous chalk catchment, UK: an assessment of the relative importance of instream and wetland processes. *Science of the Total Environment*, 282/283, 159–174.

Protected Planet (2019) Explore the world's marine protected areas. https://www.protectedplanet.net/marine (last accessed October 2019).

Proulx, M. & Mazumder, A. (1998) Reversal of grazing impact on plant species richness in nutrient-poor vs nutrient-rich ecosystems. *Ecology*, 79, 2581–2592.

Pulliam, H.R. (1988) Sources, sinks and population regulation. *American Naturalist*, 132, 652–661.

Pullin, A., Frampton, G., Jongman, R. *et al.* (2016) Selecting appropriate methods of knowledge synthesis to inform biodiversity policy. *Biodiversity and Conservation*, 25, 1285–1300.

Purschke, O., Schmid, B.C., Sykes, M.T. *et al.* (2013) Contrasting changes in taxonomic, phylogenetic and functional diversity during a long-term succession: insights into assembly processes. *Journal of Ecology*, 101, 857–866.

Putman, R.J. (1978) Patterns of carbon dioxide evolution from decaying carrion. Decomposition of small carrion in temperate systems. *Oikos*, 31, 47–57.

Pyke, G.H. (1982) Local geographic distributions of bumblebees near Crested Butte, Colorada: competition and community structure. *Ecology*, 63, 555–573.

Qiu, Y. (2015) Iron fertilization by Asian dust influences North Pacific sardine regime shifts. *Progress in Oceanography*, 134, 370–378.

Quinn, J.F. & Harrison, S.P. (1988) Effects of habitat fragmentation and isolation on species richness – evidence from biogeographic patterns. *Oecologia*, 75, 132–140.

Quintero, A., Molero, G., Reynolds, M.P. & Calderini, D.F. (2018) Trade-off between grain weight and grain number in wheat depends on GxE interaction: A case study of an elite CIMMYT panel (CIMCOG). *European Journal of Agronomy*, 92, 17–29.

Raberg, L., Sim, D. & Read, A.F. (2007) Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. *Science*, 318, 812–814.

Radchuk, V., Ims, R.A. & Andreassen, H. (2016) From individuals to population cycles: the role of extrinsic and intrinsic factors in rodent populations. *Ecology*, 97, 720–732.

Raffaelli, D. & Hawkins, S. (1999) *Intertidal Ecology*, 2nd edn. Kluwer, Dordrecht.

Rahbek, C. (1995) The elevational gradient of species richness: a uniform pattern? *Ecography*, 18, 200–205.

Raimondi, P., Jurgens, L.J. & Tinker, M.T. (2015) Evaluating potential conservation conflicts between two listed species: sea otters and black abalone. *Ecology*, 96, 3102–3108.

Rainey, P.B. & Travisano, M. (1998) Adaptive radiation in a heterogeneous environment. *Nature*, 394, 69–72.

Ramírez, F., Afán, I., Davis, L.S. & Chiaradia, A. (2017) Climate impacts on global hot spots of marine biodiversity. *Science Advances*, 3, e1601198.

Randall, M.G.M. (1982) The dynamics of an insect population throughout its altitudinal distribution: *Coleophora alticolella* (Lepidoptera) in northern England. *Journal of Animal Ecology*, 51, 993–1016.

Rasher, D.B. & Hay, M.E. (2014) Competition induces allelopathy but suppresses growth and anti-herbivore defence in a chemically rich seaweed. *Proceedings of the Royal Society of London, Series B*, 281, 20132615.

Raunkiaer, C. (1934) *The Life Forms of Plants*. Oxford University Press, Oxford. (Translated from the original published in Danish, 1907.)

Raven, J.A., Cockell, C.S. & De La Rocha, C.L. (2008) The evolution of inorganic carbon concentrating mechanisms in photosynthesis. *Philosophical Transaction of the Royal Society of London, Series B*, 363, 2641–2650.

Ray, J.C. & Sunquist, M.E. (2001) Trophic relations in a community of African rainforest carnivores. *Oecologia*, 127, 395–408.

Read, A.F. & Harvey, P.H. (1989) Life history differences among the eutherian radiations. *Journal of Zoology*, 219, 329–353.

Redfern, M., Jones, T.H. & Hassell, M.P. (1992) Heterogeneity and density dependence in a field study of a tephritid–parasitoid interaction. *Ecological Entomology*, 17, 255–262.

Redpath, S.M., Mougeot, F., Leckie, F.M., Elston, D.A. & Hudson, P.J. (2006) Testing the role of parasites in driving the cyclic population dynamics of a gamebird. *Ecology Letters*, 9, 410–418.

Reece, C.H. (1985) The role of the chemical industry in improving the effectiveness of agriculture. *Philosophical Transactions of the Royal Society of London, Series B*, 310, 201–213.

Rees, M., Childs, D.Z. & Ellner, S.P. (2014) Building integral projection models: a user's guide. *Journal of Animal Ecology*, 83, 528–545.

Rees, M., Childs, D.Z. & Freckleton, R.P. (2012) Assessing the role of competition and stress: a critique of importance indices and the development of a new approach. *Journal of Ecology*, 100, 577–585.

Rees, M., Condit, R., Crawley, M., Pacala, S. & Tilman, D. (2001) Long-term studies of vegetation dynamics. *Science*, 293, 650–655.

Rees, M. & Ellner, S.P. (2009) Integral projection models for populations in temporally varying environments. *Ecological Monographs*, 79, 575–594.

Regan, H.M., Ben-Haim, Y., Langford, B. *et al.* (2005) Robust decision-making under severe uncertainty for conservation management. *Ecological Applications*, 15, 1471–1477.

Reganold, J.P & Wachter, M. (2015) Organic agriculture in the twenty-first century. *Nature Plants*, 2, 15221.

Regus, J.U., Wendlandt, C.E., Bantay, R.M. *et al.* (2017) Nitrogen deposition decreases the benefits of symbiosis in a native legume. *Plant Soil*, 414, 159–170.

Reich, P.B., Grigal, D.F., Aber, J.D. & Gower, S.T. (1997) Nitrogen mineralization and productivity in 50 hardwood and conifer stands on diverse soils. *Ecology*, 78, 335–347.

Reich, P.B., Hobbie, S.E. Lee, T.D. & Pastore, M.A. (2018) Unexpected reversal of C_3 versus C_4 grass response to elevated CO_2 during a 20-year field experiment. *Science*, 360, 317–320.

Reich, P.B. & Oleksyn, J. (2004) Global patterns of plant leaf N and P in relation to temperature and latitude. *Proceedings of the National Academy of Sciences of the USA*, 101, 11001–11006.

Reunanen, P., Monkkonen, M. & Nikula, A. (2000) Managing boreal forest landscapes for flying squirrels. *Conservation Biology*, 14, 218–227.

Reynolds, J.J.H., Lambin, X., Massey, F.P. *et al.* (2012) Delayed induced silica defences in grasses and their potential for destabilising herbivore population dynamics. *Oecologia*, 170, 445–456.

Rezende, E.L., Albert, E.M., Fortuna, M.A. & Bascompte, J. (2009) Compartments in a marine food web associated with phylogeny, body mass, and habitat structure. *Ecology Letters*, 12, 779–788.

Reznick, D.N. (1982) The impact of predation on life history evolution in Trinidadian guppies: genetic basis of observed life history patterns. *Evolution*, 36, 1236–1250.

Reznick, D.N., Bryga, H. & Endler, J.A. (1990) Experimentally induced life history evolution in a natural population. *Nature*, 346, 357–359.

Rhoades, D.F. & Cates, R.G. (1976) Towards a general theory of plant antiherbivore chemistry. *Advances in Phytochemistry*, 110, 168–213.

Ricciardi, A., Blackburn, T.M., Carlton, J.T. *et al.* (2017) Invasion science: a horizon scan of emerging challenges and opportunities. *Trends in Ecology and Evolution*, 32, 464–474.

Ricciardi, A. & MacIsaac, H.J. (2000) Recent mass invasion of the North American Great Lakes by Ponto-Caspian species. *Trends in Ecology and Evolution*, 15, 62–65.

Richman, A.D., Case, T.J. & Schwaner, T.D. (1988) Natural and unnatural extinction rates of reptiles on islands. *American Naturalist*, 131, 611–630.

Rickards, J., Kelleher, M.J. & Storey, K.B. (1987) Strategies of freeze avoidance in larvae of the goldenrod gall moth *Epiblema scudderiana*: winter profiles of a natural population. *Journal of Insect Physiology*, 33, 581–586.

Ricklefs, R.E. & Lovette, I.J. (1999) The role of island area *per se* and habitat diversity in the species–area relationships of four Lesser Antillean faunal groups. *Journal of Animal Ecology*, 68, 1142–1160.

Ringel, M.S., Hu, H.H. & Anderson, G. (1996) The stability and persistence of mutualisms embedded in community interactions. *Theoretical Population Biology*, 50, 281–297.

Ripple, W.J., Estes, J.A., Schmitz, O.J. *et al.* (2016) What is a trophic cascade? *Trends in Ecology and Evolution*, 31, 842–849.

Ritchie, E.G. & Johnson, C.N. (2009) Predator interactions, mesopredator release and biodiversity conservation. *Ecology Letters*, 12, 982–998.

Robertson, J.H. (1947) Responses of range grasses to different intensities of competition with sagebrush (*Artemisia tridentata* Nutt.). *Ecology*, 28, 1–16.

Robertson, L.A., Lagabrielle, E., Lombard, A.T. *et al.* (2017) Pelagic bioregionalisation using open-access data for better planning of marine protected area networks. *Ocean an Coastal Management*, 148, 214–219.

Robinson, S.P., Downton, W.J.S. & Millhouse, J.A. (1983) Photosynthesis and ion content of leaves and isolated chloroplasts of salt-stressed spinach. *Plant Physiology*, 73, 238–242.

Rockström, R., Steffen, W., Noone, K. *et al.* (2009) A safe operating space for humanity. *Nature*, 461, 472–475.

Rodrigues, A.S.L., Pilgrim, J.D., Lamoreux, J.F., Hoffmann, M. & Brooks, T.M. (2006) The value of the IUCN Red List for conservation. *Trends in Ecology and Evolution*, 21, 71–76.

Roff, D.A. & Fairbairn, D. J. (2007) The evolution of trade offs: where are we? *Journal of Evolutionary Biology*, 20, 433–447.

Rogers, B.M., Jantz, P. & Goetz, S.J. (2017) Vulnerability of eastern US tree species to climate change. *Global Change Biology*, 23, 3302–3322.

Rohani, P., Godfray, H.C.J. & Hassell, M.P. (1994) Aggregation and the dynamics of host–parasitoid systems: a discrete-generation model with within-generation redistribution. *American Naturalist*, 144, 491–509.

Rohr, D.H. (2001) Reproductive trade-offs in the elapid snakes *Austrelap superbus* and *Austrelap ramsayi*. *Canadian Journal of Zoology*, 79, 1030–1037.

Root, R. (1967) The niche exploitation pattern of the blue-grey gnatcatcher. *Ecological Monographs*, 37, 317–350.

Rosenzweig, M.L. (1971) Paradox of enrichment: destabilization of exploitation ecosystems in ecological time. *Science*, 171, 385–387.

Rosenzweig, M.L. & MacArthur, R.H. (1963) Graphical representation and stability conditions of predator–prey interactions. *American Naturalist*, 97, 209–223.

Rosenzweig, M.L. & Sandlin, E.A. (1997) Species diversity and latitudes: listening to area's signal. *Oikos*, 80, 172–176.

Rosindell, J., Hubbell, S.P. & Etienne, R.S. (2011) The Unified Neutral Theory of Biodiversity and Biogeography at age ten. *Trends in Ecology and Evolution*, 26, 340–348.

Rosindell, J., Hubbell, S.P., He, F., Harmon, L.J. & Etienne, R.S. (2012) The case for ecological neutral theory. *Trends in Ecology and Evolution*, 27, 203–208.

Ross, K., Cooper, N., Bidwell, J.R. & Elder, J. (2002) Genetic diversity and metal tolerance of two marine species: a comparison between populations from contaminated and reference sites. *Marine Pollution Bulletin*, 44, 671–679.

Rothaupt, K.O. (1988) Mechanistic resource competition theory applied to laboratory experiments with zooplankton. *Nature*, 333, 660–662.

Roughan, M. & Middleton, J.H. (2002) A comparison of observed upwelling mechanisms off the east coast of Australia. *Continental Shelf Research*, 22, 2551–2572.

Rouphael, A.B. & Inglis, G.J. (2001) 'Take only photographs and leave only footprints'? An experimental study of the impacts of underwater photographers on coral reef dive sites.

Biological Conservation, 100, 281–287.

Rousk, K, Sorensen, P.L., Lett, S. & Michelson, A. (2015) Across-habitat comparison of diazotroph activity in the subarctic. *Microbial Ecology*, 69, 778–787.

Rousseaux, C.S. & Gregg, W.W. (2014) Interannual variation in phytoplankton primary production at a global scale. *Remote Sensing*, 6, 1–19.

Rowan, R., Knowlton, N., Baker, A. & Jara, J. (1997) Landscape ecology of algal symbionts creates variation in episodes of coral bleaching. *Nature*, 388, 265–269.

Rowe, C.L. (2002) Differences in maintenance energy expenditure by two estuarine shrimp (*Palaemonetes pugio* and *P. vulgaris*) that may permit partitioning of habitats by salinity. *Comparative Biochemistry and Physiology Part A*, 132, 341–351.

Rowland, H.M., Mappes, J. Ruxton, G.D. & Speed, M.P. (2010) Mimicry between unequally defended prey can be parasitic: evidence for quasi-Batesian mimicry. *Ecology Letters*, 13, 1494–1502.

Royama, T. (1992) *Analytical Population Dynamics*. Chapman & Hall, London.

Royer, D.L., Berner, R.A., Montanez, I.P., Tabor, N.K. & Beerling, D.J. (2004) CO_2 as a primary driver of Phanerozoic climate. *GSA Today*, 14, 4–10.

Rundel, P.W., Arroyo, M.T.K., Cowling, R.M., Keeley, J.E., Lamont, B.B. and Vargas, P. (2016) Mediterranean biomes: evolution of their vegetation, floras, and climate. *Annual Review of Ecology Evolution and Systematics*, 47, 383–407.

Russell, M.L., Vermeire, L.T., Ganguli, A.C. & Hendrickson, J.R. (2015) Season of fire manipulates bud bank dynamics in northern mixed-grass prairie. *Plant Ecology*, 216, 835–846.

Ruxton, G.D., Sherratt, T.N. & Speed, M.P. (2004) *Avoiding Attack: The Evolutionary Ecology of Crypsis, Warning Signals and Mimicry*. Oxford University Press, Oxford.

Ryder, J.J., Webberley, K.M., Boots, M. & Knell, R.J. (2005) Measuring the transmission dynamics of a sexually transmitted disease. *Proceedings of the National Academy of Sciences of the USA*, 102, 15140–15143.

Saccheri, I. & Hanski, I. (2006) Natural selection and population dynamics. *Trends in Ecology and Evolution*, 21, 341–347.

Saccheri, I., Kuussaari, M., Kankare, M., Vikman, P., W Fortelius, W. & Hanski, I. (1998) Inbreeding and extinction in a butterfly metapopulation. *Nature*, 392, 491–494.

Saccheri, I.K., Rousset, F., Watts, P.C., Brakefield, P.M. & Cook, L.M. (2008) Selection and gene flow on a diminishing cline of melanic peppered moths. *Proceedings of the National Academy of Sciences of the USA*, 105, 16212–16217.

Sackville Hamilton, N.R., Matthew, C. & Lemaire, G. (1995) In defence of the –3/2 boundary line: a re-evaluation of self-thinning concepts and status. *Annals of Botany*, 76, 569–577.

Sadeh, A., Guterman, H., Gersani, M. & Ovadia, O. (2009) Plastic bet-hedging in an amphicarpic annual: an integrated strategy under variable conditions. *Evolutionary Ecology*, 23, 373–388.

Sadras, V.O. (2007) Evolutionary aspects of the trade-off between seed size and number in crops. *Field Crops Research*, 100, 125–138.

Saeki, Y., Tuda, M. & Crowley, P.H. (2014) Allocation tradeoffs and life histories: a conceptual and graphical framework. *Oikos*, 123, 786–793.

Sakai, S. (2002) A review of brood-site pollination mutualism: plants providing breeding sites for their pollinators. *Journal of Plant Research*, 115, 161–168.

Sale, P.F. (1977) Maintenance of high diversity in coral reef fish communities. *American Naturalist*, 111, 337–359.

Sale, P.F. (1979) Recruitment, loss and coexistence in a guild of territorial coral reef fishes. *Oecologia*, 42, 159–177.

Salek, M. (2012) Spontaneous succession on opencast mining sites: implications for bird biodiversity. *Journal of Applied Ecology*, 49, 1417–1425.

Salguero-Gomez, R. (2017) Applications of the fast–slow continuum and reproductive strategy framework of plant life histories. *New Phytologist*, 213, 1618–1624.

Salkeld, D.J., Stapp, P., Tripp, D.W. *et al.* (2016) Ecological traits driving the outbreaks and emergence of zoonotic pathogens. *Bioscience*, 66, 118–129.

Saloniemi, I. (1993) An environmental explanation for the character displacement pattern in *Hydrobia* snails. *Oikos*, 67, 75–80.

Salt, D.E., Smith, R.D. & Raskin, I. (1998) Phytoremediation. *Annual Review of Plant Physiology and Plant Molecular Biology*, 49, 643–668.

Salvatteci, R., Field, D., Gutierrez, D. *et al.* (2018) Multifarious anchovy and sardine regimes in the Humboldt Current System during the last 150 years. *Global Change Biology*, 24, 1055–1068.

Sandercock, B.K., Nilsen, E.B., Brøset, H. & Pedersen, H.C. (2011) Is hunting mortality additive or compensatory to natural mortality? Effects of experimental harvest on the survival and cause-specific mortality of willow ptarmigan. *Journal of Animal Ecology*, 80, 244–258.

Sanders, D., Moser, A., Newton, J. & van Veen, F.J.F (2016) Trophic assimilation efficiency markedly increases at higher trophic levels in four-level host-parasitoid food chain. *Proceedings of the Royal Society of London, Series B*, 283, 20153043.

Sanders, N.J., Moss, J. & Wagner, D. (2003) Patterns of ant species richness along elevational gradients in an arid ecosystem. *Global Ecology and Biogeography*, 12, 93–102.

Sanderson, E.W., Jaiteh, M., Levy, M.A., Redford, K.H., Wannebo, A.V. & Woolmer, G. (2002) The human footprint and the last of the wild. *BioScience*, 52, 891–904.

Sandin, S.A. & Pacala, S.W. (2005) Fish aggregation results in inversely density-dependent predation on continuous coral reefs. *Ecology*, 86, 1520–1530.

Santini, F., Miglietta, M.P. & Faucci, A. (2012) Speciation: where are we now? An introduction to a special issue on speciation. *Evolutionary Biology*, 39, 141–147.

Saraeny Rivera-Salinas, I., Hajian-Forooshani, Z., Jiménez-Soto, E., Cruz-Rodríguez, J.A. & Philpott, S.M. (2018) High intermediary mutualist density provides consistent biological control in a tripartite mutualism. *Biological Control*, 118, 26–31.

Saunois, M., Poulter, B., Peregon, A. *et al.* (2016) The Global Methane Budget 2000–2012. *Earth System Science Data*, DOI:10.5194/essd-8-697-2016.

Savolainen, O., Lascoux, M. & Merila, J. (2013) Ecological genomics of local adaptation. *Nature Reviews, Genetics*, 14, 807–820.

Savolainen, V., Anstett, M.C., Lexer, C. *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature*, 441, 210–213.

Schade-Poole, K. & Möller, G. (2016) Impact and mitigation of nutrient pollution and overland water flow change on the Florida Everglades, USA. *Sustainability*, 8, 940.

Schaffer, W.M. & Kot, M. (1986) Chaos in ecological systems: the coals that Newcastle forgot. *Trends in Ecology and Evolution*, 1, 58–63.

Schall, J.J. (1992) Parasite-mediated competition in *Anolis* lizards. *Oecologia*, 92, 58–64.

Scheffer, M. & Carpenter, S.R. (2003) Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends in Ecology and Evolution*, 18, 648–656.

Scheffer, M., Carpenter, S.R., Lenton, T.M. *et al.* (2012) Anticipating critical transitions. *Science*, 338, 344–348.

Schenck, D. & Bacher, S. (2002) Functional response of a generalist insect predator to one of its prey species in the field. *Journal of Animal Ecology*, 71, 524–531.

Schetter, T.A., Lochmiller, R.L., Leslie, D.M. Jr, Engle, D.M. & Payton, M.E. (1998) Examination of the nitrogen limitation hypothesis in non-cyclic populations of cotton rats *Sigmodon hispidus*. *Journal of Animal Ecology*, 67, 705–721.

Schimel, D. & Baker, D. (2002) The wildfire factor. *Nature*, 420, 29–30.

Schindler, D.W. (2012) The dilemma of controlling cultural eutrophication of lakes. *Proceedings of the Royal Society of London, Series B*, 279, 4322–4333.

Schmaljohann, H. & Both, C. (2017) The limits of modifying migration speed to adjust to climate change. *Nature Climate Change*, 7, doi: 10.1038/NCLIMATE3336.

Schmitt, R.J. (1987) Indirect interactions between prey: apparent competition, predator aggregation, and habitat segregation. *Ecology*, 68, 1887–1897.

Schmitz, O.J., Hamback, P.A. & Beckerman, A.P. (2000) Trophic cascades in terrestrial systems: a review of the effects of carnivore removals on plants. *American Naturalist*, 155, 141–153.

Schnell, J.K., Harris, G.M., Pimm, S.L. & Russell, G.J. (2013) Estimating extinction risk with metapopulation models of large-scale fragmentation. *Conservation Biology*, 27, 520–530.

Schöb, C., Armas, C., Guler, M., Prieto, I. & Pugnaire, F.L. (2013) Variability in functional traits mediates plant interactions along stress gradients. *Journal of Ecology*, 101, 753–762.

Schoener, T.W. (1974) Resource partitioning in ecological communities. *Science*, 185, 27–39.

Schoener, T.W. (1983) Field experiments on interspecific competition. *American Naturalist*, 122, 240–285.

Schoener, T.W. (1989) Food webs from the small to the large. *Ecology*, 70, 1559–1589.

Schoenly, K., Beaver, R.A. & Heumier, T.A. (1991) On the trophic relations of insects: a food-web approach. *American Naturalist*, 137, 597–638.

Schultz, T.R. & Brady, S.G. (2008) Major evolutionary transitions in ant agriculture. *Proceedings of the National Academy of Sciences of the USA*, 105, 5435–5440.

Schuur, E.A.G., McGuire, A.D., Schadel, C. *et al.* (2015) Climate change and the permafrost carbon feedback. *Nature*, 520, 171–179.

Schwartz, O.A., Armitage, K.B. & Van Vuren. D. (1998) A 32-year demography of yellow-bellied marmots (*Marmota flaviventris*). *Journal of Zoology*, 246, 337–346.

Scogings, P.F. (2018) Foliar flavonol concentration in *Sclerocarya birrea* saplings responds to nutrient fertilisation according to growth-differentiation balance hypothesis. *Phytochemistry Letters*, 23, 180–184.

Secretariat of the Stockholm Convention (2010) Stockholm Convention on Persistent Organic Pollutants (POPs) as amended in 2009. http://chm.pops.int/Home/tabid/2121/Default.aspx (last accessed November 2019).

Seitzinger, S.P., Mayorga, E., Bouwman, A.F. *et al.* (2010) Global river nutrient export: a scenario analysis of past and future trends. *Global Biogeochemical Cycles*, 24, doi.org/10.1029/2009GB003587.

Selden, R.L., Batt, R.D., Saba, V.S. & Pinsky, M.L. (2017) Diversity in thermal affinity among key piscivores buffers impacts of ocean warming on predator–prey interactions. *Global Change Biology*, 24, 117–131.

Semere, T. & Froud-Williams, R.J. (2001) The effect of pea cultivar and water stress on root and shoot competition between vegetative plants of maize and pea. *Journal of Applied Ecology*, 38, 137–145.

Sendall, K.M., Lusk, C.H. & Reich, P.B. (2015) Becoming less tolerant with age: sugar maple, shade, and ontogeny. *Oecologia*, 179, 1011–1021.

Sergio, F., Blas, J. & Hiraldo, F. (2009) Predictors of floater status in a long-lived bird: a cross-sectional and longitudinal test of hypotheses. *Journal of Animal Ecology*, 78, 109–118.

Serpetti, N., Baudron, A.R., Burrows, M.T. *et al.* (2017) Impact of ocean warming on sustainable fisheries management informs the Ecosystem Approach to Fisheries. *Scientific Reports*, 7, 13438.

Serra, S.R.Q., Graca, M.A.S., Doledec, S. & Feio, M.J. (2017) Chironomidae traits and life history strategies as indicators of anthropogenic disturbance. *Environmental Monitoring and Assessment*, 189, 326.

Severini, M., Baumgärtner, J. & Limonta, L. (2003) Parameter estimation for distributed delay based population models from laboratory data: egg hatching of *Oulema duftschmidi*. *Ecological Modelling*, 167, 233–246.

Sformo, T., Walters, K., Jeannet, K. *et al.* (2010) Deep supercooling, vitrification and limited survival to −100°C in the Alaskan beetle Cucujus clavipes puniceus (Coleoptera: Cucujidae) larvae. *Journal of Experimental Biology*, 213, 502–509.

Shaffer, M.L. (1981) Minimum population sizes for species conservation. *BioScience*, 31, 131–134.

Shanafelt, D.W. & Loreau, M. (2018) Stability trophic cascades in food chains. *Royal Society Open Science*, 5, 180995.

Shankar Raman, T., Rawat, G.S. & Johnsingh, A.J.T. (1998) Recovery of tropical rainforest avifauna in relation to vegetation succession following shifting cultivation in Mizoram, north-east India. *Journal of Applied Ecology*, 35, 214–231.

Sharma, E., Anand, G. & Kapoor, R. (2017) Terpenoids in plant and arbuscular mycorrhiza-reinforced defence against herbivorous insects. *Annals of Botany*, 119, 791–801.

Shaw, A.K. (2016) Drivers of animal migration and implications in changing environments. *Evolutionary Ecology*, 30, 991–1007.

Shaw, D.J. & Dobson, A.P. (1995) Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. *Parasitology*, 111, S111–S133.

Shea, K. & Chesson, P. (2002) Community ecology theory as a framework for biological invasions. *Trends in Ecology and Evolution*, 17, 170–177.

Shea, K. & Kelly, D. (1998) Estimating biocontrol agent impact with matrix models: *Carduus nutans* in New Zealand. *Ecological Applications*, 8, 824–832.

Shea, K., Kelly, D., Sheppard, A.W. & Woodburn, T.L. (2005) Context-dependent biological control of an invasive thistle. *Ecology*, 86, 3174–3181.

Shebitz, D.J. & Kimmerer, R.W. (2005) Reestablishing roots of a Mohawk community and a culturally significant plant: sweetgrass. *Restoration Ecology*, 13, 257–264.

Shen, G., Pimm, S.L., Feng, C. *et al.* (2015) Climate change challenges the current conservation strategy for the giant panda. *Biological Conservation*, 190, 43–50.

Shepherd, J.G. & Pope, J.G. (2002a) Dynamic pool models I: interpreting the past using virtual population analysis. In: *Handbook of Fish and Fisheries, Vol. II* (P.J.B. Hart & J.D. Reynolds, eds), pp 127–163. Blackwell Science, Oxford.

Shepherd, J.G. & Pope, J.G. (2002b) Dynamic pool models II: short-term and long-term forecasts of catch and biomass. In: *Handbook of Fish and Fisheries, Vol. II* (P.J.B. Hart & J.D. Reynolds, eds), pp 164–188. Blackwell Science, Oxford.

Sherman, R.E., Fahey, T.J. & Battles, J.J. (2000) Small-scale disturbance and regeneration dynamics in a neotropical mangrove forest. *Journal of Ecology*, 88, 165–178.

Sherriff, M.J., Krebs, C.J. & Boonstra, R. (2009) The sensitive hare: sublethal effects of predator stress on reproduction in snowshoe hares. *Journal of Animal Ecology*, 78, 1249–1258.

Sherriff, M.J., Krebs, C.J. & Boonstra, R. (2010) The ghosts of predators past: population cycles and the role of maternal programming under fluctuating predation risk. *Ecology*, 91, 2983–2994.

Shrestha, G. & St Clair, L.L. (2013) Lichens: a promising source of antibiotic and anticancer drugs. *Phytochemical Reviews*, 12, 229–244.

Shultz, J.M., Espinel, Z., Espinola, M. & Rechkemmer, A. (2016) Distinguishing epidemiological features of the 2013–2016 West Africa Ebola virus disease outbreak. *Disaster Health*, 3, 78–88.

Shykoff, J.A. & Bucheli, E. (1995) Pollinator visitation patterns, floral rewards and the probability of transmission of *Micro-botryum violaceum*, a venereal disease of plants. *Journal of Ecology*, 83, 189–198.

Sibly, R.M. & Brown, J.H. (2007) Effects of body size and lifestyle on evolution of mammal life histories. Proceedings of the National Academy of Sciences of the USA, 104, 17707–17712.

Sibly, R.M. & Calow, P. (1983) An integrated approach to life-cycle evolution using selective landscapes. *Journal of Theoretical Biology*, 102, 527–547.

Sibly, R.M. & Hone, J. (2002) Population growth rate and its determinants: an overview. *Philosophical Transactions of the Royal Society of London, Series B*, 357, 1153–1170.

Sibly, R.M. & Smith, R.H. (1998) Identifying key factors using λ-contribution analysis. *Journal of Animal Ecology*, 67, 17–24.

Sih, A. & Christensen, B. (2001) Optimal diet theory: when does it work, and when and why does it fail? *Animal Behaviour*, 61, 379–390.

Sih, A., Crowley, P., McPeek, M., Petranka, J. & Strohmeier, K. (1985) Predation, competition and prey communities: a review of field experiments. *Annual Review of Ecology and Systematics*, 16, 269–311.

Silulwane, N.F., Richardson, A.J., Shillington, F.A. & Mitchell-Innes, B.A. (2001) Identification and classification of vertical chlorophyll patterns in the Benguela upwelling system and Angola-Benguela front using an artificial neural network. *South African Journal of Marine Science*, 23, 37–51.

Silvertown, J.W. (1980) The evolutionary ecology of mast seeding in trees. *Biological Journal of the Linnean Society*, 14, 235–250.

Silvertown, J.W. (1982) *Introduction to Plant Population Ecology*. Longman, London.

Silvertown, J.W., Holtier, S., Johnson, J. & Dale, P. (1992) Cellular automaton models of interspecific competition for space – the effect of pattern on process. *Journal of Ecology*, 80, 527–533.

Simberloff, D.S. (1976) Experimental zoogeography of islands: effects of island size. *Ecology*, 57, 629–648.

Simberloff, D. & Boecklen, W.J. (1981) Santa Rosalia reconsidered: size ratios and competition. *Evolution*, 35, 1206–1228.

Simberloff, D., Dayan, T., Jones, C. & Ogura, G. (2000) Character displacement and release in the small Indian mongoose, *Herpestes javanicus*. *Ecology*, 81, 2086–2099.

Simberloff, D. & Von Holle, B. (1999) Positive interactions of nonindigenous species: invasional meltdown? *Biological Invasions*, 1, 21–32.

Simberloff, D.S. & Wilson, E.O. (1969) Experimental zoogeography of islands: the colonization of empty islands. *Ecology*, 50, 278–296.

Simon, K.S., Townsend, C.R., Biggs, B.J.F., Bowden, W.B. & Frew, R.D. (2004) Habitat-specific nitrogen dynamics in New Zealand streams containing native or invasive fish. *Ecosystems*, 7, 777–792.

Sinervo, B. (1990) The evolution of maternal investment in lizards: an experimental and comparative analysis of egg size and its effects on offspring performance. *Evolution*, 44, 279–294.

Siva-Jothy, M.T., Moret, Y. & Rolff, J. (2005) Insect immunity: an evolutionary ecology perspective. *Advances in Insect Physiology*, 32, 1–48.

Skelton, J., Doak, S., Leonard, M., Creed, R.P. & Brown, B.L. (2015) The rules for symbiont community assembly change along a mutualism–parasitism continuum. *Journal of Animal Ecology*, 85, 843–853.

Smetacek, V., Klaas, C., Strass, V.H. *et al.* (2012) Deep carbon export from a Southern Ocean iron-fertilized diatom bloom. *Nature*, 487, 313–319.

Smilanich, A.M., Fincher, R.M. & Dyer, L.A. (2016) Does plant apparency matter? Thirty years of data provide limited support but reveal clear patterns of the effects of plant chemistry on herbivores. *New Phytologist*, 210, 1044–1057.

Smith, C.C. & Fretwell, S.D. (1974) Optimal balance between size and number of offspring. *American Naturalist*, 108, 499–506.

Smith, C.R., De Leo, F.C., Bernardino, A.F., Sweetman, A.K. & Martinez Arbizu, P. (2008) Abyssal food limitation, ecosystem structure and climate change. *Trends in Ecology and Evolution*, 23, 518–528.

Smith, F.A., Gittleman, J.L. & Brown, J.H. (2014) *Foundations of Macroecology: Classic Papers with Commentaries*. University of Chicago Press, Chicago.

Smith, F.D.M., May, R.M., Pellew, R., Johnson, T.H. & Walter, K.R. (1993) How much do we know about the current extinction rate? *Trends in Ecology and Evolution*, 8, 375–378.

Smith, M.J., Telfer, S., Kallio, E.R. *et al.* (2009) Host–pathogen time series data in wildlife support a transmission function between density and frequency dependence. *Proceedings of the National Academy of Sciences of the USA*, 106, 7905–7909.

Smith, R.S., Shiel, R.S., Bardgett, R.D. *et al.* (2003) Soil microbial community, fertility, vegetation and diversity as targets in the restoration management of a meadow grassland. *Journal of Applied Ecology*, 40, 51–64.

Smith, S.E. & Smith, F.A. (2011) Roles of arbuscular myccorrhizas in plant nutrition and growth: new paradigms from cellular to ecosystem scales. *Annual Reviews of Plant Biology*, 62, 227–250.

Smith, S.J., van Aardenne, J., Klimont, Z., Andres, R.J., Volke, A. & Delgado, A.S. (2011) Anthropogenic sulfur dioxide emissions: 1850–2005. *Atmospheric Chemistry and Physics*, 11, 1101–1116.

Smith, S.V. (1984) Phosphorus vs nitrogen limitation in the marine environment. *Limnology and Oceanography*, 29, 1149–1160.

Smyrnioudis, I.N., Harrington, R., Hall, M., Katis, N. & Clark, S.J. (2001) The effect of temperature on variation in transmission of a BYDV PAV-like isolate by clones of *Rhopalosiphum padi* and *Sitobion avenae*. *European Journal of Plant Pathology*, 107, 167–173.

Snajdr, J., Cajthaml, T., Valaskova, V. *et al.* (2011) Transformation of Quercus petraea litter: successive changes in litter chemistry are reflected in differential enzyme activity and changes in the microbial community composition. *Microbial Ecology*, 75, 291–303.

Snaydon, R.W. (1996) Above-ground and below-ground interactions in intercropping. In: *Dynamics and Nitrogen in Cropping Systems of the Semi-arid Tropics* (O. Ito, C. Johansen, J.J. Adu-Gyamfi, K. Katayama, J.V.D.K. Kumar Rao & T.J. Rego, eds), pp. 73–92. JIRCAS, Kasugai, Japan.

Snyder, W.E. & Ives, A.R. (2009) Population dynamics and species interactions. In: *Integrated Pest Management* (E.B. Radcliffe, W.D. Hutchison & R.E. Cancelado, eds), pp. 62–74. Cambridge University Press, Cambridge.

Solbrig, O.T. & Simpson, B.B. (1974) Components of regulation of a population of dandelions in Michigan. *Journal of Ecology*, 62, 473–486.

Soliveres, S. & Allan, E. (2018) Everything you always wanted to know about intransitive competition but were afraid to ask. *Journal of Ecology*, 106, 807–814.

Soliveres, S., Lehmann, A., Boch, S. *et al.* (2018) Intransitive competition is common across five major taxonomic groups and is driven by productivity, competitive rank and functional traits.

Journal of Ecology, 106, 852–864.

Soliveres, S., Smit, C. & Maestre, F.T. (2015) Moving forward on facilitation research: response to changing environments and effects on the diversity, functioning and evolution of plant communities. *Biological Reviews*, 90, 297–313.

Solomon, M.E. (1949) The natural control of animal populations. *Journal of Animal Ecology*, 18, 1–35.

Sommer, U. (1990) Phytoplankton nutrient competition – from laboratory to lake. In: *Perspectives on Plant Competition* (J.B. Grace & D. Tilman, eds), pp. 193–213. Academic Press, New York.

Sopow, S.L., Bader, M.K-F. & Brockerhoff, E.G. (2015) Bark beetles attacking conifer seedlings: picking on the weakest or feasting upon the fittest? *Journal of Applied Ecology*, 52, 220–227.

Sorato, E., Gullett, P.R., Griffith, S.C. & Russell, A.F. (2012) Effects of predation risk on foraging behaviour and group size: adaptations in a social cooperative species. *Animal Behaviour*, 84, 823–834.

Sorrell, B.K., Hawes, I., Schwarz, A.-M. & Sutherland, D. (2001) Interspecfic differences in photosynthetic carbon uptake, photosynthate partitioning and extracellular organic carbon release by deepwater characean algae. *Freshwater Biology*, 46, 453–464.

Sousa, M.E. (1979a) Experimental investigation of disturbance and ecological succession in a rocky intertidal algal community. *Ecological Monographs*, 49, 227–254.

Sousa, M.E. (1979b) Disturbance in marine intertidal boulder fields: the nonequilibrium maintenance of species diversity. *Ecology*, 60, 1225–1239.

South, A.B., Rushton, S.P., Kenward, R.E. & Macdonald, D.W. (2002) Modelling vertebrate dispersal and demography in real landscapes: how does uncertainty regarding dispersal behaviour influence predictions of spatial population dynamics. In: *Dispersal Ecology* (J.M. Bullock, R.E. Kenward & R.S. Hails, eds), pp. 327–349. Blackwell Science, Oxford.

Southwood, T.R.E., Hassell, M.P., Reader, P.M. & Rogers, D.J. (1989) The population dynamics of the viburnum whitefly (*Aleurotrachelus jelinekii*). *Journal of Animal Ecology*, 58, 921–942.

Souza, M.L., Solar, R.R. & Fagundes M. (2015) Reproductive strategy of Copaifera langsdorfii (Fabaceae): more seeds or better seeds? *Revista de Biologia Tropical*, 63, 1161–1167.

Sparks, T.C. (2013) Insecticide discovery: an evaluation and analysis. *Pesticide Biochemistry and Physiology*, 107, 8–17.

Spiller, D.A. & Schoener, T.W. (1990) A terrestrial field experiment showing the impact of eliminating predators on foliage damage. *Nature*, 347, 469–472.

Spiller, D.A. & Schoener, T.W. (1994) Effects of a top and intermediate predators in a terrestrial food web. *Ecology*, 75, 182–196.

Spiller, D.A. & Schoener, T.W. (1998) Lizards reduce spider species richness by excluding rare species. *Ecology*, 79, 503–516.

Sprent, J.I. (1979) *The Biology of Nitrogen Fixing Organisms*. McGraw Hill, London.

Stadler, B. & Dixon, A.F.G. (1998) Costs of ant attendance for aphids. *Journal of Animal Ecology*, 67, 454–459.

Stamp, N. (2003) Out of the quagmire of plant defense hypotheses. *Quarterly Review of Biology*, 78, 23–55.

Stanton, M.L., Palmer, T.M. & Young, T.P. (2002) Competition-colonization trade-offs in a guild of African acacia ants. *Ecological Monographs*, 72, 347–363.

Stauffer, B. (2000) Long term climate records from polar ice. *Space Science Reviews*, 94, 321–336.

Stearns, S.C. (1977) The evolution of life history traits. *Annual Review of Ecology and Systematics*, 8, 145–171.

Stearns, S.C. (1983) The impact of size and phylogeny on patterns of covariation in the life history traits of mammals. *Oikos*, 41, 173–187.

Stearns, S.C. (2000) Life history evolution: successes, limitations, and prospects. *Naturwissenschaften*, 87, 476–486.

Steffen, W., Richardson, K., Rockström, J. *et al.* (2015) Planetary boundaries: guiding human development on a changing planet. *Science*, 347, 1259855.

Stein, A., Gerstner, K. & Kreft, H. (2013) Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecology Letters*, 17, 866–880.

Steiner, C.F., Long, Z.T., Krumins, J.A. & Morin, P.J. (2005) Temporal stability of aquatic food webs: partitioning the effects of species diversity, species composition and enrichment. *Ecology Letters*, 8, 819–828.

Steiner, E.R., Criddle, K.R. & Adkison, M.D. (2011) Balancing biological sustainability with the economic needs of Alaska's sockeye salmon fisheries. *North American Journal of Fisheries Management*, 31, 431–444.

Steingrimsson, S.O. & Grant, J.W.A. (1999) Allometry of territorysize and metabolic rate as predictors of self-thinning in young-of-the- year Atlantic salmon. *Journal of Animal Ecology*, 68, 17–26.

Stenseth, N.C., Atshabar, B.B., Begon, M. *et al.* (2008) Plague: past, present and future. *PLoS Medicine*, 5, 9–13.

Stenseth, N.C., Falck, W., Bjørnstad, O.N. & Krebs, C.J. (1997) Population regulation in snowshoe hare and Canadian lynx: asymmetric food web configurations between hare and lynx. *Proceedings of the National Academy of Sciences of the USA*, 94, 5147–5152.

Stephens, A.E.A. & Westoby, M. (2015) Effects of insect attack to stems on plant survival, growth, reproduction and photosynthesis. *Oikos*, 124, 266–273.

Stephens, D.W. & Krebs, J.R. (1986) *Foraging Theory*. Princeton University Press, Princeton, NJ.

Stephens, P.A. & Sutherland, W.J. (1999) Consequences of the Allee effect for behaviour, ecology and conservation. *Trends in Ecology and Evolution*, 14, 401–405.

Sterner, R.W., Bajpai, A. & Adams, T. (1997) The enigma of food chain length: absence of theoretical evidence for dynamic constraints. *Ecology*, 78, 2258–2262.

Sterner, R.W. & Elser, J.J. (2002) *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere*. Princeton University Press, Princeton, NJ.

Stevens, C.E. & Hume, I.D. (1998) Contributions of microbes in vertebrate gastrointestinal tract to production and conservation of nutrients. *Physiological Reviews*, 78, 393–426.

Stone, L. & Roberts, A. (1990) The checkerboard score and species distributions. *Oecologia*, 85, 74–79.

Stolp, H. (1988) *Microbial Ecology*. Cambridge University Press, Cambridge, UK.

Storch, D., Sizling, A.L. & Gaston, K.J. (2003) Geometry of the species area relationship in central European birds: testing the mechanism. *Journal of Animal Ecology*, 72, 509–519.

Storey, K.B. & Storey, J.M. (2012) Insect cold hardiness: metabolic, gene, and protein adaptation. *Canadian Journal of Zoology*, 90, 456–475.

Stouffer, D.B. & Bascompte, J. (2011) Compartmentalization increases food web persistence. *Proceedings of the National Academy of Sciences of the USA*, 108, 3648–3652.

Stowe, L.G. & Teeri, J.A. (1978) The geographic distribution of C4 species of the Dicotyledonae in relation to climate. *American Naturalist*, 112, 609–623.

Strauss, S.Y. & Agrawal, A.A. (1999) The ecology and evolution of plant tolerance to herbivory. *Trends in Ecology and Evolution*, 14, 179–185.

Strauss, S.Y., Irwin, R.E. & Lambrix, V.M. (2004) Optimal defence theory and flower petal colour predict variation in the secondary chemistry of wild radish. *Journal of Ecology*, 92, 132–141.

Strauss, S.Y., Rudgers, J.A., Lau, J.A. & Irwin, R.E. (2002) Direct and ecological costs of resistance to herbivory. *Trends in Ecology and Evolution*, 17, 278–285.

Stromberg, P.C., Toussant, M.J. & Dubey, J.P. (1978) Population biology of *Paragonimus kellicotti* metacercariae in central Ohio. *Parasitology*, 77, 13–18.

Strong, D.R. (1992) Are trophic cascades all wet? Differentiation and donor-control in speciose ecosystems. *Ecology*, 73, 747–754.

Strong, D.R. Jr, Lawton, J.H. & Southwood, T.R.E. (1984) *Insects on Plants: community patterns and mechanisms*. Blackwell Scientific Publications, Oxford.

Stuart, Y.E & Losos, J.B. (2013) Ecological character displacement: glass half full or half empty? *Trends in Ecology and Evolution*, 28, 402–408.

Suarez, N. (2016) Effect of nitrogen on the dynamics of leaf population demography in whole plants of *Wedelia trilobata*. *International Journal of Plant Science*, 177, 694–705.

Suggett, D.J., Warner, M.E. & Leggat, W. (2017) Symbiotic dinoflagellate functional diversity mediates coral survival under ecological crisis. *Trends in Ecology and Evolution*, 32, 735–745.

Summers, K., Speed, M.P., Blount, J.D & Stuckert, A.M.M. (2015) Aposematic signals honest? A review. *Journal of Evolutionary Biology*, 28, 1583–1599.

Sun, L., Cao X., Li, M., Zhan, X., Li, X. & Cui, Z. (2017) Enhanced bioremediation of lead-contaminated soil by Solanum nigrum L. with Mucor circinelloides. *Environmental Science and Pollution Research*, 24, 9681–9689.

Sunday, J.M., Bates, A.E. & Dulvy, N.K. (2011) Global analysis of thermal tolerance and latitude in ectotherms. *Proceedings of the Royal Society of London, Series B*, 278, 1823–1830.

Sunderland, K.D., Hassall, M. & Sutton, S.L. (1976) The population dynamics of *Philoscia muscorum* (Crustacea, Oniscoidea) in a dune grassland ecosystem. *Journal of Animal Ecology*, 45, 487–506.

Sutherland, W.J. (1983) Aggregation and the 'ideal free' distribution. *Journal of Animal Ecology*, 52, 821–828.

Sutherland, W.J. (1998) The importance of behavioural studies in conservation biology. *Animal Behaviour*, 56, 801–809.

Sutherland, W.J., Barnard, P., Broad, S. *et al.* (2010) A 2017 horizon scan of emerging issues for global conservation and biological diversity. *Trends in Ecology and Evolution*, 32, 31–40.

Sutton, M.A., Pitcairn, C.E.R. & Fowler, D. (1993) The exchange of ammonia between the atmosphere and plant communities. *Advances in Ecological Research*, 24, 302–393.

Sutton, S.L. & Collins, N.M. (1991) Insects and tropical forest conservation. In: *The Conservation of Insects and their Habitats* (N.M. Collins & J.A. Thomas, eds), pp. 405–424. Academic Press, London.

Suzuki, R. & Suzuki, S. (2012) Morphological adaptation of a palatable plant to long-term grazing can shift interactions with an unpalatable plant from facilitative to competitive. *Plant Ecology*, 213, 175–183.

Swedish Environmental Protection Agency (2009) Sweden's Commitment under the Baltic Sea Action Plan Socio-economic impact assessments. *Report* 5989.

Swift, M.J., Heal, O.W. & Anderson, J.M. (1979) *Decomposition in Terrestrial Ecosystems*. Blackwell Scientific Publications, Oxford.

Swinnerton, K.J., Groombridge, J.J., Jones, C.G., Burn, R.W. & Mungroo, Y. (2004) Inbreeding depression and founder diversity among captive and free-living populations of the endangered pink pigeon *Columba mayeri*. *Animal Conservation*, 7, 353–364.

Swinton, J., Harwood, J., Grenfell, B.T. & Gilligan, C.A. (1998) Persistence thresholds for phocine distemper virus infection in harbour seal Phoca vitulina metapopulations. *Journal of Animal Ecology*, 67, 54–68.

Swope, S.M. & Parker, I.M. (2010) Widespread seed limitation affects plant density but not population trajectory in the invasive plant *Centaurea solstitialis*. *Oecologia*, 164, 117–128.

Symondson, W.O.C., Sunderland, K.D. & Greenstone, M.H. (2002) Can generalist predators be effective biocontrol agents? *Annual Review of Entomology*, 47, 561–594.

Symonides, E. (1979) The structure and population dynamics of psammophytes on inland dunes. II. Loose-sod populations. *Ekologia Polska*, 27, 191–234.

Tabashnik, B.E., Brévault, T. & Carrière, Y. (2013) Insect resistance to Bt crops: lessons from the first billion acres. *Nature Biotechnology*, 31, 510–521.

Tagliabue, A., Bowie, A.R., Boyd, P.W., Buck, K.N., Johnson, K.S. & Saito. M.A. (2017) The integral role of iron in ocean biogeochemistry. *Nature*, 543, 51–59.

Takahashi, Y., Morita, S., Yoshimura, J. & Watanabe, M. (2011) A geographic cline induced by negative frequency-dependent selection. *BMC Evolutionary Biology*, 11, 256.

Takimoto, G. & Post, D.M. (2013) Environmental determinants of food-chain length: a meta-analysis. *Ecological Research*, 28, 675–681.

Takimoto, G., Spiller, D.A. & Post, D.M. (2008) Ecosystem size, but not disturbance, determines food-chain length on islands of the Bahamas. *Ecology*, 89, 3001–3007.

Takimoto, G. & Suzuki, K. (2017) Global stability of obligate mutualism in community modules with facultative mutualists. *Oikos*, 125, 535–540.

Tamiru, A., Khan, Z.R. & Bruce, T.J.A. (2015) New directions for improving crop resistance to insects by breeding for egg induced defence. *Current Opinion in Insect Science*, 9, 51–55.

Tamm, C.O. (1956) Further observations on the survival and flowering of some perennial herbs. *Oikos*, 7, 274–292.

Tanaka, M.O. & Magalhaes, C.A. (2002) Edge effects and succession dynamics in *Brachidontes* mussel beds. *Marine Ecology Progress Series*, 237, 151–158.

Taniguchi, H., Nakano, S. & Tokeshi, M. (2003) Habitat complexity, patch size and the abundance of epiphytic invertebrates on plants. *Freshwater Biology*, 48, 718–728.

Taniguchi, Y. & Nakano, S. (2000) Condition-specific competition: implications for the altitudinal distribution of stream fishes. *Ecology*, 81, 2027–2039.

Tansley, A.G. (1939) *The British Islands and their Vegetation*. Cambridge University Press, Cambridge, UK.

Tant, C.J., Rosemond, A.D., Helton, A.M. *et al.* (2015) Nutrient enrichment alters the magnitude and timing of fungal, bacterial, and detritivore contributions to litter breakdown. *Freshwater Science*, 34, 1259–1271.

Tarwater, C.E. & Arcese, P. (2017) Age and years to death disparately influence reproductive allocation in a short-lived bird. *Ecology*, 98, 2248–2254.

Taylor, A. & Burns, K. (2015) Plant composition patterns inside an endemic birds' nest fern (Asplenium goudeyi) on Lord Howe Island: effects of fern size, fern isolation and plant dispersal abilities. *Journal of Tropical Ecology*, 31, 413–421.

Taylor, A.D. (1993) Heterogeneity in host–parasitoid interactions: the CV2 rule. *Trends in Ecology and Evolution*, 8, 400–405.

Taylor, I. (1994) *Barn Owls. Predator–Prey Relationships and Conservation*. Cambridge University Press, Cambridge, UK.

Taylor, L.R. (1987) Objective and experiment in long-term research. In: *Long-Term Studies in Ecology* (G.E. Likens, ed.), pp. 20–70. Springer-Verlag, New York.

Taylor, M.L., Gwinnett, C., Robinson, L.F. & Woodall, L.C. (2016) Plastic microfiber ingestion by deep-sea organisms. *Scientific Reports*, 6, 33997.

Taylor, R.B., Sotka, E. & Hay, M.E. (2002) Tissue-specific induction of herbivore resistance: seaweed response to amphipod grazing. *Oecologia*, 132, 68–76.

Tchabovsky, A.V., Savinetskaya, L.E., Surkova, E.N., Ovchinnikova, N.L. & Kshnyasev, I.A. (2016) Delayed threshold response of a rodent population. *Oecologia*, 182, 1075–1082.

TEEB (2010) *The Economics of Ecosystems and Biodiversity: Mainstreaming the Economics of Nature: A Synthesis of the Approach, Conclusions and Recommendations of TEEB*. UN Environment, Geneva, Switzerland.

Teirney, L.D. (2003) *Fiordland Marine Conservation Strategy: Te Kaupapa Atawhai o Te Moana o Atawhenua*. Guardians of Fiordland's Fisheries and Marine Environment Inc., Te Anau, New Zealand.

Telfer, S. Lambin, X., Birtles, R. *et al.* (2010) Species interactions in a parasite community drive infection risk in a wildlife population. *Science*, 330, 243–246.

Téllez-Valdés, O. & Dávila-Aranda, P. (2003) Protected areas and climate change: a case study of the cacti in the Tehuacán-Cuicatlán Biosphere Reserve, Mexico. *Conservation Biology*, 17, 846–853.

Temeles, E.J. 1994. The role of neighbours in territorial systems: when are they 'dear enemies'? *Animal Behaviour*, 47, 339–350.

Temming, A., Goetz, S., Mergardt, N. & Ehrich, S. (2004) Predation of whiting and haddock on sandeel: aggregative response, competition and diel periodicity. *Journal of Fish Biology*, 64, 1351–1372.

terHorst, C.P. (2011) Experimental evolution of protozoan traits in response to interspecific competition. *Journal of Evolutionary Biology*, 24, 36–46.

Thaler, J.S., Humphrey, P.T. & Whiteman, N.K. (2012) Evolution of jasmonate and salicylate signal crosstalk. *Trends in Plant Science*, 17, 260–270.

Thijs, S., Langill, T. & Vangronsveld, J. (2017) The bacterial and fungal microbiota of hyperaccumulator plants: small organisms, large influence. *Advances in Botanical Research*, 83, 43–86.

Thibault, K.M., White, E.P., Hurlbert, A.H. & Ernest, S.K.M. (2011) Multimodality in the individual size distributions of bird communities. *Global Ecology and Biogeography*, 20, 145–153.

Thomas, C.D. (1990) What do real population dynamics tell us about minimum viable population sizes? *Conservation Biology*, 4, 324–327.

Thomas, C.D. & Harrison, S. (1992) Spatial dynamics of a patchily distributed butterfly species. *Journal of Applied Ecology*, 61, 437–446.

Thomas, C.D., Thomas, J.A. & Warren, M.S. (1992) Distributions of occupied and vacant butterfly habitats in fragmented landscapes. *Oecologia*, 92, 563–567.

Thomas, H. (2013) Senescence, ageing and death of the whole plant. *New Phytologist*, 197, 696–711.

Thomas, J.A., Simcox, D.J. & Clarke, R.T. (2009) Successful conservation of a threatened *Maculinea* butterfly. *Science*, 325(5936), 80–83.

Thompson, D.J. (1975) Towards a predator–prey model incorporating age-structure: the effects of predator and prey size on the predation of *Daphnia magna* by *Ischnura elegans*. *Journal of Animal Ecology*, 44, 907–916.

Thompson, J.N. (1988) Evolutionary ecology of the relationship between oviposition preference and performance of offspring in phytophagous insects. *Entomologia Experimentia et Applicata*, 47, 3–14.

Thompson, R.M., Mouritsen, K.N. & Poulin, R. (2005) Importance of parasites and their life cycle characteristics in determining the structure of a large marine food web. *Journal of Animal Ecology*, 74, 77–85.

Thompson, R.M. & Townsend, C.R. (2000) Is resolution the solution?: the effect of taxonomic resolution on the calculated properties of three stream food webs. *Freshwater Biology*, 44, 413–422.

Thompson, R.M. & Townsend, C.R. (2006) A truce with neutral theory: local deterministic factors, species traits and neutrality determine patterns of diversity in streams. *Journal of Animal Ecology*, 75, 476–484.

Thrall, P.H., Barrett, L.G., Dodds, P.N. & Burdon, J.J. (2016) Epidemiological and evolutionary outcomes of gene-for-gene and matching allele models. *Frontiers in Plant Science*, 6, 1084.

Thrall, P.H. & Burdon, J.J. (2003) Evolution of virulence in a plant host-pathogen metapopulation. *Science*, 299, 1735–1727.

Tillman, P.G. (1996) Functional response of *Microplitis croceipes* and *Cardiochiles nigriceps* (Hymenoptera: Braconidae) to variation in density of tobacco budworm (Lepidoptera: Noctuidae). *Environmental Entomology*, 25, 524–528.

Tilman, D. (1982) *Resource Competition and Community Structure*. Princeton University Press, Princeton, NJ.

Tilman, D. (1986) Resources, competition and the dynamics of plant communities. In: *Plant Ecology* (M.J. Crawley, ed.), pp. 51–74. Blackwell Scientific Publications, Oxford.

Tilman, D. (1987) Secondary succession and the pattern of plant dominance along experimental nitrogen gradients. *Ecological Monographs*, 57, 189–214.

Tilman, D. (1988) *Plant Strategies and the Dynamics and Structure of Plant Communities*. Princeton University Press, Princeton, NJ.

Tilman, D. (1990) Mechanisms of plant competition for nutrients: the elements of a predictive theory of competition. In: *Perspectives on Plant Competition* (J.B. Grace & D. Tilman, eds), pp. 117–141. Academic Press, New York.

Tilman, D. (1996) Biodiversity: population versus ecosystem stability. *Ecology*, 77, 350–363.

Tilman, D. (1999) The ecological consequences of changes in biodiversity: a search for general principles. *Ecology*, 80, 1455–1474.

Tilman, D. (2007) Resource competition and plant traits: a response to Craine *et al.* 2005. *Journal of Ecology*, 95, 231–234.

Tilman, D., Fargione, J., Wolff, B. *et al.* (2001a) Forecasting agriculturally driven global environmental change. *Science*, 292, 281–284.

Tilman, D., Isbell, F. & Cowles, J.M. (2014) Biodiversity and ecosystem functioning. *Annual Review of Ecology, Evolution and Systematics*, 45, 471–493.

Tilman, D., Mattson, M. & Langer, S. (1981) Competition and nutrient kinetics along a temperature gradient: an experimental test of a mechanistic approach to niche theory. *Limnology and Oceanography*, 26, 1020–1033.

Tilman, D., Reich, P.B. & Isbell, F. (2012) Biodiversity impacts ecosystem productivity as much as resources, disturbance, or herbivory. *Proceedings of the National Academy of Sciences of the USA*, 109, 10394–10397.

Tilman, D., Reich, P.B. & Knops, J.M.H. (2006) Biodiversity and ecosystem stability in a decade-long grassland experiment. *Nature*, 441, 629–632.

Tilman, D., Reich, P.B., Knops, J., Wedin, D., Meilke, T. & Lehman, C. (2001b) Diversity and productivity in a long-term grassland experiment. *Science*, 294, 843–845.

Tilman, D. & Wedin, D. (1991a) Plant traits and resource reduction for five grasses growing on a nitrogen gradient. *Ecology*, 72, 685–700.

Tilman, D. & Wedin, D. (1991b) Dynamics of nitrogen competition between successional grasses. *Ecology*, 72, 1038–1049.

Timmermann, A., Oberhuber, J., Bacher, A., Esch, M., Latif, M. & Roeckner, E. (1999) Increased El Niño frequency in a climate model forced by future greenhouse warming. *Nature*, 398, 694–697.

Tinbergen, L. (1960) The natural control of insects in pinewoods. 1: factors influencing the intensity of predation by songbirds. *Archives Néerlandaises de Zoologie*, 13, 266–336.

Tittensor, D.P., Rex, M.A., Stuart, C.T., McClain, C.R. & Smith, C.R. (2011) Species-energy relationships in deep-sea molluscs. *Biology Letters*, 7, 718–722.

Tjallingii, W.F. & Hogen Esch, Th. (1993) Fine structure of aphid stylet routes in plant tissues in correlation with EPG signals. *Physiological Entomology*, 18, 317–328.

Tokeshi, M. (2009) *Species Coexistence: ecological and evolutionary perspectives*. Blackwell Science, Oxford.

Tompkins, D.M. & Begon, M. (1999) Parasites can regulate wildlife populations. *Parasitology Today*, 15, 311–313.

Tompkins, D.M., Dunn, A.M., Smith, M.J. & Telfer, S. (2011) Wildlife diseases: from individuals to ecosystems. *Journal of Animal Ecology*, 80, 19–38.

Tompkins, D.M., White, A.R. & Boots, M. (2003) Ecological replacement of native red squirrels by invasive greys driven by disease. *Ecology Letters*, 6, 189–196.

Towne, E.G. (2000) Prairie vegetation and soil nutrient responses to ungulate carcasses. *Oecologia*, 122, 232–239.

Townsend, C.R. (2003) Individual, population, community and ecosystem consequences of a fish invader in New Zealand streams. *Conservation Biology*, 17, 38–47.

Townsend, C.R. (2008) *Ecological applications: toward a sustainable world*. Blackwell Publishing, Oxford.

Townsend, C.R. & Hildrew, A.G. (1980) Foraging in a patchy environment by a predatory net-spinning caddis larva: a test of optimal foraging theory. *Oecologia*, 47, 219–221.

Townsend, C.R. & Hildrew, A.G. (1994) Species traits in relation to a habitat templet for river systems. *Freshwater Biology*, 31, 265–275.

Townsend, C.R., Hildrew, A.G. & Francis, J.E. (1983) Community structure in some southern English streams: the influence of physiochemical factors. *Freshwater Biology*, 13, 521–544.

Townsend, C.R., Scarsbrook, M.R. & Dolédec, S. (1997) The intermediate disturbance hypothesis, refugia and biodiversity in streams. *Limnology and Oceanography*, 42, 938–949.

Townsend, C.R., Winfield, I.J., Peirson, G. & Cryer, M. (1986) The response of young roach *Rutilus rutilus* to seasonal changes in abundance of microcrustacean prey: a field demonstration of switching. *Oikos*, 46, 372–378.

Tranchida, M.C., Centeno, N.D. & Cabello, M.N. (2014) Soil fungi: their potential use as a forensic tool. *Journal of Forensic Science*, 59, 785–789.

Tregenza, T. (1995) Building on the ideal free distribution. *Advances in Ecological Research*, 26, 253–307.

Trewick, S.A. & Worthy, T.H. (2001) Origins and prehistoric ecology of takahe based on morphometric, molecular, and fossil data. In: *The Takahe: fifty years of conservation* *management and research* (W.G. Lee & I.G. Jamieson, eds), pp. 31–48. University of Otago Press, Dunedin, New Zealand.

Tribouillois, H., Dürr, C., Demilly, D., Wagner, M-H. & Justes, E. (2016) Determination of germination response to temperature and water potential for a wide range of cover crop species and related functional groups. *PLoS One*, 11, e0161185.

Tripet, F. & Richner, H. (1999) Density dependent processes in the population dynamics of a bird ectoparasite *Ceratophyllus gallinae*. *Ecology*, 80, 1267–1277.

Trochet, A., Courtois, E.A., Stevens, V.M. *et al.* (2016) Evolution of sex-biased dispersal. *Quarterly Review of Biology*, 91, 297–320.

Turchin, P. (2003) *Complex Population Dynamics*. Princeton University Press, Princeton, NJ.

Turchin, P. & Batzli, G.O. (2001) Availability of food and the population dynamics of arvicoline rodents. *Ecology*, 82, 1521–1534.

Turchin, P. & Hanski, I. (1997) An empirically based model for latitudinal gradient in vole population dynamics. *American Naturalist*, 149, 842–874.

Turchin, P. & Hanski, I. (2001) Contrasting alternative hypotheses about rodent cycles by translating them into parameterized models. *Ecology Letters*, 4, 267–276.

Turesson, G. (1922a) The species and variety as ecological units. *Hereditas*, 3, 100–113.

Turesson, G. (1922b) The genotypical response of the plant species to the habitat. *Hereditas*, 3, 211–350.

Turkington, R. & Harper, J.L. (1979) The growth, distribution and neighbour relationships of *Trifolium repens* in a permanent pasture. IV. Fine scale biotic differentiation. *Journal of Ecology*, 67, 245–254.

Turkington, R. & Mehrhoff, L.A. (1990) The role of competition in structuring pasture communities. In: *Perspectives on Plant Competition* (J.B. Grace & D. Tilman, eds), pp. 307–340. Academic Press, New York.

Uchida, Y. & Inoue, M. (2010) Fish species richness in spring-fed ponds: effects of habitat size versus isolation in temporally variable environments. *Freshwater Biology*, 55, 983–994.

Uchmanski, J. (1985) Differentiation and frequency distributions of body weights in plants and animals. *Philosophical Transactions of the Royal Society of London, Series B*, 310, 1–75.

Umbanhowar, J., Maron, J. & Harrison, S. (2003) Density dependent foraging behaviors in a parasitoid lead to density dependent parasitism of its host. *Oecologia*, 137, 123–130.

UNEP (United Nations Environment Program) (2014) *Assessing Global Land Use: balancing consumption with sustainable supply*. United Nations Environment Programme, Paris, France.

UNEP (United Nations Environment Program) (2017a) *Handbook for the Montreal Protocol on Substances that Deplete the Ozone Layer*, 11th edn. UNEP, Nairobi, Kenya.

UNEP (United Nations Environment Program) (2017b) *Minemata Convention on Mercury: text and annexes*. UNEP, Nairobi, Kenya.

United Nations (2011) *World Population Prospects: The 2010 Revision*. United Nations, Department of Economic and Social Affairs, Population Division, New York.

United States Geological Survey (2017) Nonindegenous acuatic species. https://nas.er.usgs.gov/viewer/omap.aspx?SpeciesID=5 (last accessed November 2019).

Urabe, J., Suzuki, T., Nishita, T. & Makino, W. (2013) Immediate ecological impacts of the 2011 tohoku earthquake tsunami on intertidal flat communities. *PLoS One*, 8, e62779.

Usio, N. & Townsend, C.R. (2002) Functional significance of crayfish in stream food webs: roles of omnivory, substrate heterogeneity and sex. *Oikos*, 98, 512–522.

Usio, N. & Townsend, C.R. (2004) Roles of crayfish: consequences of predation and bioturbation for stream invertebrates. *Ecology*, 85, 807–822.

Vacchiano, G., Derose, R.J., Shaw, J.D., Svoboda, M. & Motta, R. (2013) A density management diagram for Norway spruce in the temperate European montane region. *European Journal of Forest Research*, 132, 535–549.

Valiente-Banuet, A., Rumebe, A.V., Verdu, M. & Callaway, R.M. (2006) Modern quaternary plant lineages promote diversity through facilitation of ancient tertiary lineages. *Proceedings of the National Academy of Sciences of the USA*, 103, 16812–16817.

Valladares, V.F. & Pearcy, R.W. (1998) The functional ecology of shoot architecture in sun and shade plants of *Heteromeles arbutifolia* M. Roem., a Californian chaparral shrub. *Oecologia*, 114, 1–10.

Van Arnam, E.B., Currie, C.R. & Clardy, J. (2018) Defense contracts: molecular protection in insect-microbe symbioses. *Chemical Society Reviews*, 47, 1638–1651.

Van Breeman, N. (2002) Natural organic tendency. *Nature*, 415, 381–382.

Van Buskirk, J. & Ostfeld, R.S. (1995) Controlling Lyme disease by modifying the density and species composition of tick hosts. *Ecological Applications*, 5, 1133–1140.

van de Koppel, J., Rietkerk, M. & Weissing, F.J. (1997) Catastrophic vegetation shifts and soil degredation in terrestrial grazing systems. *Trends in Ecology and Evolution*, 12, 352–356.

Van der Heijden, M.G.A., Martin, F.M., Selosse, M-A. & Sanders, I.R. (2015) Mycorrhizal ecology and evolution: the past, the present and the future. *New Phytologist*, 205, 1406–1423.

Van der Werf, E. (1992) Lack's clutch size hypothesis: an examination of the evidence using meta-analysis. *Ecology*, 73, 1699–1705.

van Groenigen, J.W., Lubbers, I.M., Vos, H.M.J., Brown, G.G., De Deyn, G.B. & van Groenigen, K.J. (2014) Earthworms increase plant production: a meta-analysis. *Scientific Reports*, 4, 6365.

van Lenteren, J.C. (2012) The state of commercial augmentative biological control: plenty of natural enemies, but a frustrating lack of uptake. *BioControl*, 57, 1–20.

Van Moorter, B., Rolandsen, C.M., Basille, M. & Gaillard, J–M. (2016) Movement is the glue connecting home ranges and habitat selection. *Journal of Animal Ecology*, 85, 21–31.

Van Oosten, H.H., Van Turnhout, C., Hallmann, C.A. *et al.* (2015) Site-specific dynamics in remnant populations of Northern Wheatears *Oenanthe oenanthe* in the Netherlands. *Ibis*, 157, 91–102.

Van Oppen, M.J.H., Gates, R.D., Blackall, L.L. *et al.* (2017) Shifting paradigms in restoration of the world's coral reefs. *Global Change Biology*, 23, 3437–3448.

van Rijn, I., Buba, Y., DeLong, J., Kiflawi, M. & Belmaker, J. (2017) Large but uneven reduction in fish size across species in relation to changing sea temperatures. *Global Change Biology*, 23, 3667–3674.

van Riper, C., van Riper, S.G., Goff, M.L. & Laird, M. (1986) The epizootiology and ecological significance of malaria in Hawaiian land birds. *Ecological Monographs*, 56, 327–344.

van Strien, N.J., Manullang, B., Isnan, W. *et al.* (2008) Dicerorhinus sumatrensis. The IUCN Red List of Threatened Species 2008, e.T6553A12787457.

Vandenkoornhuyse, P., Ridgway, K.P., Watson, I.J., Fitter, A.H. & Young, J.P.W. (2003) Co-existing grass species have distinctive arbuscular mycorrhizal communities. *Molecular Ecology*, 12, 3085–3095.

Vander Wall, S.B. (2010) How plants manipulate the scatter-hoarding behaviour of seed-dispersing animals. *Philosophical Transactions of the Royal Society of London B*, 365, 989–997.

Vander Wall, S.B., Barga, S.C. & Seaman, A.E. (2017) The geographic distribution of seed-dispersal mutualisms in North America. *Evolution and Ecology*, 31, 725–740.

van't Hof, A.E., Campagne, P., Rigden, D.J. *et al.* (2016) The industrial melanism mutation in British peppered moths is a transposable element. *Nature*, 534(7605), 102–105.

Vaquer-Sunyer, R. & Duarte, C.M. (2008) Thresholds of hypoxia for marine biodiversity. *Proceedings of the National Academy of Sciences of the USA*, 17, 1788–1797.

Varley, G.C. (1947) The natural control of population balance in the knapweed gall-fly (*Urophora jaceana*). *Journal of Animal Ecology*, 16, 139–187.

Varley, G.C. & Gradwell, G.R. (1968) Population models for the winter moth. *Symposium of the Royal Entomological Society of London*, 9, 132–142.

Varley, G.C. & Gradwell, G.R. (1970) Recent advances in insect population dynamics. *Annual Review of Entomology*, 15, 1–24.

Vázquez, G.J.A. & Givnish, T.J. (1998) Altitudinal gradients in tropical forest composition, structure, and diversity in the Sierra de Manantlán. *Journal of Ecology*, 86, 999–1020.

Veen, G.F., Freschet, G.T., Ordonez, A. & Wardle, D.A. (2015) Litter quality and environmental controls of home-field advantage effects on litter decomposition. *Oikos*, 124, 187–195.

Velando, A., Drummond, H. & Torres, R. (2006) Senescent birds redouble reproductive effort when ill: confirmation of the terminal investment hypothesis. *Proceedings of the Royal Society of London, Series B*, 273, 1443–1448.

Vellend, M. (2010) Conceptual synthesis in community ecology. *The Quarterly Review of Biology*, 85, 183–206.

Venterink, H.O., Wassen, M.J., Verkroost, A.W.M. & de Ruiter, P.C. (2003) Species richness–productivity patterns differ between N-, P- and K-limited wetlands. *Ecology*, 84, 2191–2199.

Verhoeven, J.T.A., Arheimer, B., Yin, C. & Hefting, M.M. (2006) Regional and global concerns over wetlands and water quality. *Trends in Ecology and Evolution*, 21, 96–103.

Verhulst, J., Montana, C., Mandujano, M.C. & Franco, M. (2008) Demographic mechanisms in the coexistence of two closely related perennials in a fluctuating environment. *Oecologia*, 156, 95–105.

Vet, R., Artz, R.S., Carou, S. *et al.* (2014) A global assessment of precipitation chemistry and deposition of sulfur, nitrogen, sea salt, base cations, organic acids, acidity and pH, and phosphorus. *Atmospheric Environment*, 93, 3–100.

Viana, M., Mancy, R, Biek, R. *et al.* (2014) Assembling evidence for identifying reservoirs of infection. *Trends in Ecology and Evolution*, 29, 270–279.

Villa, F., Tunesi, L. & Agardy, T. (2002) Zoning marine protected areas through spatial multiple-criteria analysis: the case of the Asinara Island National Marine Reserve of Italy. *Conservation Biology*, 16, 515–526.

Virolainen, K.M., Suomi, T., Suhonen, J. & Kuitunen, M. (1998) Conservation of vascular plants in single large and several small mires: species richness/rarity and taxonomic diversity. *Journal of Applied Ecology*, 35, 700–707.

Vitousek, P.M. (2004) *Nutrient Cycling and Limitation: Hawaii as a Model System*. Princeton University Press, Princeton, NJ.

Vitousek, P.M. & Matson, P.A. (2009) Nutrient cycling and biogeochemistry. In: *Princeton Guide to Ecology* (S.A. Levin, S.R. Carpenter, H.C.J. *et al.*, eds). Princeton University Press, Princeton, NJ.

Vivas-Martinez, S., Basanez, M.-G., Botto, C. *et al.* (2000) Amazonian onchocerciasis: parasitological profiles by host-age, sex, and endemicity in southern Venezuela. *Parasitology*, 121, 513–525.

Vogt, R.J., Romanuk, T.N. & Kolasa, J. (2006) Species richness-variability relationships in multi-trophic aquatic microcosms. *Oikos*, 113, 55–66.

Vohnık, M., Sadowsky, J.J., Kohout, P. *et al.* (2012) Novel root-fungus symbiosis in Ericaceae: sheathed ericoid mycorrhiza formed by a hitherto undescribed basidiomycete with affinities to Trechisporales. *PLoS One*, 7, e39524.

Volk, M., Niklaus, P.A. & Korner, C. (2000) Soil moisture effects determine CO_2 responses of grassland species. *Oecologia*, 125, 380–388.

Volterra, V. (1926) Variations and fluctuations of the numbers of individuals in animal species living together. (Reprinted in 1931. In: R.N. Chapman, *Animal Ecology*. McGraw Hill, New York.)

Voriskova, J. & Baldrian, P. (2013) Fungal community on decomposing leaf litter undergoes rapid successional change. *The ISME Journal*, 7, 477–486.

Vucetich, J.A., Peterson, R.O. & Schaefer, C.L. (2002) The effect of prey and predator densities on wolf predation. *Ecology*, 83, 3003–3013.

Waage, J.K. & Greathead, D.J. (1988) Biological control: challenges and opportunities. *Philosophical Transactions of the Royal Society of London, Series B*, 318, 111–128.

Wake, D.B. & Vredenburg, V.T. (2008) Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proceedings of the National Academy of Sciences of the USA*, 105, 11466–11473.

Wall, D.H., Bradford, M.A., St John, M.G. *et al.* (2008) Global decomposition experiment shows soil animal impacts on decomposition are climate dependent. *Global Change Biology*, 14, 2661–2677.

Wallis, D.W. & Hodgson, J.A. (2018) Condatis; software to assist the planning of habitat restoration. Version 1.0. http://wordpress.condatis.org.uk (last accessed November 2019).

Wallis, G.P., Waters, J.M., Upton, P. & Craw, D. (2016) Transverse alpine speciation driven by glaciation. *Trends in Ecology and Evolution*, 31, 916–926.

Walls, J.T. III, Caciagli, P., Tooker, J.F., Russo, J.M., Rajotte, E.G. & Rosa, C. (2016) Modeling the decision process for barley yellow dwarf management. *Computers and Electronic in Agriculture*, 127, 775–786.

Walsh, M.R. & Reznick, D.N. (2008) Interactions between the direct and indirect effects of predators determine life history evolution in a killifish. *Proceedings of the National Academy of Sciences of the USA*, 105, 594–599.

Walter, J., Britton, R.A. & Roos, S. (2010) Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm. *Proceedings of the National Academy of Sciences of the USA*, 108, 4645–4652.

Wang, C-J., Wan, J-Z., Qu, H. & Zhang, Z-X. (2017) Modelling plant invasion pathways in protected areas under climate change: implication for invasion management. *Web Ecology*, 17, 69–77.

Wang, D., Yang, S., Tang, F. & Zhu, H. (2012) Symbiosis specificity in the legume – rhizobial mutualism. *Cellular Microbiology*, 14, 334–342.

Wang, J. & Epstein, H.E. (2013) Estimating carbon source-sink transition during secondary succession in a Virginia valley. *Plant Soil*, 362, 135–147.

Wang, R., Goll, D., Balkanski, Y. *et al.* (2017) Global forest carbon uptake due to nitrogen and phosphorus deposition from 1850 to 2100. *Global Change Biology*, 23, 4854–4872.

Wang, Y.P., Law, R.M. & Pak, B. (2010) A global model of carbon, nitrogen and phosphorus cycles for the terrestrial biosphere. *Biogeosciences*, 7, 2261–2282.

Ward, C.L., McCann, K.S. & Rooney, N. (2015) HSS revisited: multi-channel processes mediate trophic control across a productivity gradient. *Ecology Letters*, 18, 1190–1197.

Ward, J.V. (1988) Riverine–wetland interactions. In: *Freshwater Wetlands and Wildlife* (R.R. Sharitz & J.W. Gibbon, eds), pp. 385–400. Office of Science and Technology Information, US Department of Energy, Oak Ridge, TN.

Wardrop, P., Shimeta, J., Nugegoda, D. *et al.* (2016) Chemical pollutants sorbed to ingested microbeads from personal care products accumulate in fish. *Environmental Science and Technology*, 50, 4037–4044.

Ware, D.M. (2000) *Fisheries Oceanography: an integrative approach to fisheries ecology and management*. Blackwell Science, Oxford.

Waring, R.H. & Schlesinger, W.H. (1985) *Forest Ecosystems: concepts and management*. Academic Press, Orlando, FL.

Warren, D.R., Kraft, C.E., Josephson, D.C. & Driscoll, C.T. (2017) Acid rain recovery may help to mitigate the impacts of climate change on thermally sensitive fish in lakes across eastern North America. *Global Change Biology*, 23, 2149–2153.

Waser, N.M. & Price, M.V. (1994) Crossing distance effects in *Delphinium nelsonii*: outbreeding and inbreeding depression in progeny fitness. *Evolution*, 48, 842–852.

Wassmuth, B.E., Stoll, P., Tscharntke, T. & Thies, C. (2009) Spatial aggregation facilitates coexistence and diversity of wild plant species in field margins. *Perspectives in Plant Ecology, Evolution and Systematics*, 11, 127–135.

Watanabe, H. & Tokuda, G. (2010) Cellulolytic systems in insects. *Annual Review of Entomology*, 55, 609–632.

Watanabe, Y.Y., Ito, M. & Takahashi, A. (2014) Testing optimal foraging theory in a penguin–krill system. *Proceedings of the Royal Society of London, Series B*, 281, 20132376.

Watkinson, A.R. (1984) Yield–density relationships: the influence of resource availability on growth and self-thinning in populations of *Vulpia fasciculata*. *Annals of Botany*, 53, 469–482.

Watkinson, A.R., Freckleton, R.P., Robinson, R.A. & Sutherland, W.J. (2000) Predictions of biodiversity response to genetically modified herbicide-tolerant crops. *Science*, 289, 1554–1557.

Watson, A. & Moss, R. (1980) Advances in our understanding of the population dynamics of red grouse from a recent fluctuation in numbers. *Areda*, 68, 103–111.

Watt, A.S. (1947) Pattern and process in the plant community. *Journal of Ecology*, 35, 1–22.

Watts, M.E., Ball, I.R., Stewart, R.S. *et al.* (2009) Marxan with zones: software for optimal conservation based land- and sea-use zoning. *Environmental Modelling and Software*, 24, 1513–1521.

Way, M.J. & Cammell, M. (1970) Aggregation behaviour in relation to food utilization by aphids. In: *Animal Populations in Relation to their Food Resource* (A. Watson, ed.), pp. 229–247. Blackwell Scientific Publications, Oxford.

Weathers, K.C., Caldenasso, M.L. & Pickett, S.T.A. (2001) Forest edges as nutrient and pollutant concentrators: potential synergisms between fragmentation, forest canopies and the atmosphere. *Conservation Biology*, 15, 1506–1514.

Webb, R.E., Leslie, D.M. Jr, Lochmiller, R.L. & Master, R.E. (2005) Impact of food supplementation and methionine on high densities of cotton rats: support of the amino-acid-quality hypothesis? *Journal of Mammalogy*, 86, 46–55.

Webb, W.L., Lauenroth, W.K., Szarek, S.R. & Kinerson, R.S. (1983) Primary production and abiotic controls in forests, grasslands and desert ecosystems in the United States. *Ecology*, 64, 134–151.

Webster, M.S. & Almany, G.R. (2002) Positive indirect effects in a coral reef fish community. *Ecology Letters*, 5, 549–557.

Wegener, A. (1915) *Entstehung der Kontinenter und Ozeaner*. Samml. Viewig, Braunschweig. English translation (1924) *The Origins of Continents and Oceans*. Translated by J.G.A. Skerl. Methuen, London.

Weiner, J. (1986) How competition for light and nutrients affects size variability in *Ipomoea tricolor* populations. *Ecology*, 67, 1425–1427.

Weiner, J. (1990) Asymmetric competition in plant populations. *Trends in Ecology and Evolution*, 5, 360–364.

Weiner, J. & Freckleton, R.P. (2010) Constant final yield. *Annual Review of Ecology, Evolution and Systematics*, 41, 173–192.

Weladji, R.B., Holand, O., Gaillard, J-M. *et al.* (2010) Age-specific changes in different components of reproductive output in female reindeer: terminal allocation or senescence? *Oecologia*, 162, 261–271.

Weller, D.E. (1987) A reevaluation of the −3/2 power rule of plant self-thinning. *Ecological Monographs*, 57, 23–43.

Weller, D.E. (1990) Will the real self-thinning rule please stand up? A reply to Osawa and Sugita. *Ecology*, 71, 1204–1207.

Werner, E.E., Gilliam, J.F., Hall, D.J. & Mittlebach, G.G. (1983a) An experimental test of the effects of predation risk on habitat use in fish. *Ecology*, 64, 1540–1550.

Werner, E.E., Mittlebach, G.G., Hall, D.J. & Gilliam, J.F. (1983b) Experimental tests of optimal habitat use in fish: the role of relative habitat profitability. *Ecology*, 64, 1525–1539.

Werner, E.E. & Peacor, S.D. (2003) A review of trait-mediated interactions in ecological communities. *Ecology*, 84, 1083–1100.

Werner, H.H. & Hall, D.J. (1974) Optimal foraging and the size selection of prey by the bluegill sunfish *Lepomis macrohirus*. *Ecology*, 55, 1042–1052.

Wesson, G. & Wareing, P.F. (1969) The induction of light sensitivity in weed seeds by burial. *Journal of Experimental Biology*, 20, 413–425.

West, G.B., Brown, J.H. & Enquist, B.J. (1997) A general model for the origin of allometric scaling laws in biology. *Science*, 276, 122–126.

Westhoek, A., Field, E., Rehling, F. *et al.* (2017) Policing the legume-Rhizobium symbiosis: a critical test of partner choice. *Scientific Reports*, 7, 1419.

Westphal, M.I., Pickett, M., Getz, W.M. & Possingham, H.P. (2003) The use of stochastic dynamic programming in optimal landscape reconstruction for metapopulations. *Ecological Applications*, 13, 543–555.

Wharton, D.A. (2002) *Life at the Limits: organisms in extreme environments*. Cambridge University Press, Cambridge, UK.

Wheeler, B.D. & Shaw, S.C. (1991) Above-ground crop mass and species richness of the principal types of rich-fen vegetation of lowland England and Wales. *Journal of Ecology*, 79, 285–301.

Whipps, J.M., Hand, P., Pink, D. & Bending, G.D. (2008) Phyllosphere microbiology with special reference to diversity and plant genotype. *Journal of Applied Microbiology*, 105, 1744–1755.

White, E.P., Ernest, S.K.M, Kerkhoff, A.J. & Enquist, B.J. (2007) Relationships between body size and abundance in ecology. *Trends in Ecology and Evolution*, 22, 323–330.

White, J. (1980) Demographic factors in populations of plants. In: *Demography and Evolution in Plant Populations* (O.T. Solbrig, ed.), pp. 21–48. Blackwell Scientific Publications, Oxford.

White, T.C.R. (2008) The role of food, weather and climate in limiting the abundance of animals. *Biological Reviews*, 83, 227–248.

Whiteman, G., Hope, C. & Wadhams, P. (2013) Vast costs of Arctic change. *Nature*, 499, 401–403.

Whitfield, S.M., Lips, K.R. & Donnelly, M.A. (2016) Amphibian decline and conservation in Central America. *Copeia*, 104, 351–379.

Whittaker, R.H. (1953) A consideration of climax theory: the climax as a population and pattern, *Ecological Monographs*, 23, 41–78.

Whittaker, R.H. (1956) Vegetation of the Great Smoky Mountains. *Ecological Monographs*, 23, 41–78.

Whittaker, R.H. (1975) *Communities and Ecosystems*, 2nd edn. Macmillan, London.

Whittaker, R.J. (2010) Meta-analyses and mega-mistakes: calling time on meta-analysis of the species richness-productivity relationship. *Ecology*, 91, 2522–2533.

Whittaker, R.J., Triantis, K.A. & Ladle, R.J. (2008) A general dynamic theory of oceanic island biogeography. *Journal of Biogeography*, 35, 977–994.

Whittaker, R.J., Willis, K.J. & Field, R. (2003) Climatic–energetic explanations of diversity: a macroscopic perspective. In: *Macroecology: concepts and consequences* (T.M. Blackburn & K.J. Gaston, eds), pp. 107–129. Blackwell Publishing, Oxford.

WHO (World Health Organization) (2007) *Insecticide-treated Mosquito Nets: a WHO position statement*. World Health Organization, Geneva.

WHO (World Heath Organization) (2012) *Global Plan for Insecticide Resistance Management in Malaria Vectors*. World Health Organization, Geneva.

WHO (World Health Organization) (2013) *Obesity and Overweight, fact sheet #311*. United Nations, Geneva. http://www.who.int/nutrition/topics/3_foodconsumption/en/ (last accessed November 2019).

WHO (World Health Organization) Ebola Response Team (2014) Ebola virus disease in West Africa – the first 9 months of the epidemic and forward projections. *New England Journal of Medicine*, 371, 1481–1495.

Wiebes, J.T. (1979) Coevolution of figs and their insect pollinators. *Annual Review of Ecology and Systematics*, 10, 1–12.

Wiens, J.J. & Donoghue, M.J. (2004) Historical biogeography, ecology and species richness. *Trends in Ecology and Evolution*, 19, 639–644.

Wilcove, D.S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. (1998) Quantifying threats to imperiled species in the United States. *Bioscience*, 48, 607–615.

Williams, G.C. (1966) *Adaptation and Natural Selection*. Princeton University Press, Princeton, NJ.

Williams, K.S., Smith, K.G. & Stephen, F.M. (1993) Emergence of 13-year periodical cicadas (Cicacidae: *Magicicada*): phenology, mortality and predator satiation. *Ecology*, 74, 1143–1152.

Williams, T.A., Foster, P.G., Cox, C.J. & Embley, T.M. (2013) An archaeal origin of eukaryotes supports only two primary domains of life. *Nature*, 504, 231–236.

Williams, W.D. (1988) Limnological imbalances: an antipodean viewpoint. *Freshwater Biology*, 20, 407–420.

Williamson, M.H. (1981) *Island Populations*. Oxford University Press, Oxford.

Williamson, M. (1989) The MacArthur-Wilson theory today: true but trivial. *Journal of Biogeography*, 16, 3–4.

Williamson, M. (1999) Invasions. *Ecography*, 22, 5–12.

Wilson, D.S. (1992) Complex interactions in metacommunities, with implications for biodiversity and higher levels of selection. *Ecology*, 73, 1984–2000.

Wilson, E.O. (1961) The nature of the taxon cycle in the Melanesian ant fauna. *American Naturalist*, 95, 169–193.

Wilson, J.B. & Agnew, A.D. (1992) Positive-feedback switches in plant communities. *Advances in Ecological Research*, 23, 263–333.

Wilson, J.B., Spijkerman, E. & Huisman, J. (2007) Is there really insufficient support for Tilman's R* concept? A comment on Miller *et al. American Naturalist*, 169, 700–706.

Wilson, K., Bjørnstad, O.N., Dobson, A.P. *et al.* (2002) Heterogeneities in macroparasite infections: patterns and processes. In: *The Ecology of Wildlife Diseases* (P.J. Hudson, A. Rizzoli, B.T. Grenfell, H. Heesterbeek & A.P. Dobson, eds), pp. 6–44. Oxford University Press, Oxford.

Wisniewski, M., Nassuth, A., Teulières, C. *et al.* (2014) Genomics of cold hardiness in woody plants. *Critical Reviews in Plant Sciences*, 33, 92–124.

Wittmann, A.C. & Pörtner, H.O. (2013) Sensitivities of extant animal taxa to ocean acidification. *Nature Climate Change*, 3, 995–1001.

Wood, C.L. & Lafferty, K.D. (2013) Biodiversity and disease: a synthesis of ecological perspectives on Lyme disease transmission. *Trends in Ecology and Evolution*, 48, 239–247.

Woodward, F.I. & Lomas, M.R. (2004) Vegetation dynamics: simulating responses to climatic change. *Biological Reviews of the Cambridge Philosophical Society*, 79, 643–670.

Woodwell, G.M., Whittaker, R.H. & Houghton, R.A. (1975) Nutrient concentrations in plants in the Brookhaven oak pine forest. *Ecology*, 56, 318–322.

Woolley, S.N.C., Tittensor, D.P., Dunstan, P.K. *et al.* (2016) Deep-sea diversity patterns are shaped by energy availability. *Nature*, 533, 393–396.

Wootton, J.T. (1992) Indirect effects, prey susceptibility, and habitat selection: impacts of birds on limpets and algae. *Ecology*, 73, 981–991.

Worm, B. & Duffy, J.E. (2003) Biodiversity, productivity and stability in real food webs. *Trends in Ecology and Evolution*, 18, 628–632.

Worm, B. & Paine, R.T. (2016) Humans as hyperkeystone species. *Trends in Ecology and Evolution*, 31, 600–607.

Worrall, J.W., Anagnost, S.E. & Zabel, R.A. (1997) Comparison of wood decay among diverse lignicolous fungi. *Mycologia*, 89, 199–219.

Worthington, E.B. (ed.) (1975) *Evolution of I.B.P.* Cambridge University Press, Cambridge, UK.

Wright, S.L., Rowe, D., Thompson, R.C. & Galloway, T.S. (2013) Microplastic ingestion decreases energy reserves in marine worms. *Current Biology*, 23, R1031–R1033.

Wu, X. & Sun, S. (2010) The roles of beetles and flies in yak dung removal in an alpine meadow of eastern Qinghai-Tibetan Plateau. *Ecoscience*, 17, 146–155.

Wynne-Edwards, V.C. (1962) *Animal Dispersion in Relation to Social Behaviour*. Oliver & Boyd, Edinburgh.

Yako, L.A., Mather, M.E. & Juanes, F. (2002) Mechanisms for migration of anadromous herring: an ecological basis for effective conservation. *Ecological Applications*, 12, 521–534.

Yang, X. & Post, W.M. (2011) Phosphorus transformations as a function of pedogenesis: a synthesis of soil phosphorus data using Hedley fractionation method. *Biogeosciences*, 8, 2907–2916.

Yao, I., Shibao, H. & Akimoto, S. (2000) Costs and benefits of ant attendance to the drepanosiphid aphid *Tuberculatus quercicola*. *Oikos,* 89, 3–10.

Yellowlees, D., Rees, T.A.V. & Leggat, W. (2008) Metabolic interactions between algal symbionts and invertebrate hosts. *Plant, Cell and Environment*, 31, 679–694.

Yoda, K., Kira, T., Ogawa, H. & Hozumi, K. (1963) Self thinning in overcrowded pure stands under cultivated and natural conditions. *Journal of Biology, Osaka City University*, 14, 107–129.

Yodzis, P. (1986) Competiton, mortality and community structure. In: *Community Ecology* (J. Diamond & T.J. Case, eds), pp. 480–491. Harper & Row, New York.

Yoshida, T., Jones, L.E., Ellner, S.P., Fussmann, G.F. & Hairston, N.G. Jr (2003) Rapid evolution drives ecological dynamics in a predator–prey system. *Nature*, 424, 303–306.

Young, H.S., Maxwell, S.M., Conners, M.G. & Shaffer, S.A. (2015) Pelagic marine protected areas protect foraging habitat for multiple breeding seabirds in the central Pacific. *Biological Conservation*, 181, 226–235.

Young, S.L. (2017) A systematic review of the literature reveals trends and gaps in integrated pest management studies conducted in the United States. *Pest Management Science*, 73, 1553–1558.

Young, T.P. (1981) A general model of comparative fecundity for semelparous and iteroparous life histories. *American Naturalist*, 118, 27–36.

Yu, L.M. (1995) Elicitins from *Phytophthora* and basic resistance in tobacco. *Proceedings of the National Academy of Science of the USA*, 92, 4088–4094.

Yu, S.J. (2014) *The Toxicology and Biochemistry of Insecticides*, 2nd edn. CRC Press, Boca Raton, FL.

Zalasiewicz, J., Waters, C.N., Summerhayes, C.P. *et al.* (2017) The Working Group on the Anthropocene: summary of evidence and interim recommendations. *Anthropocene*, 19, 55–60.

Zavala, J.A., Paankar, A.G., Gase, K. & Baldwin, I.T. (2004) Constitutive and inducible trypsin proteinase inhibitor production incurs large fitness costs in *Nicotinia attenuata*. *Proceedings of the National Academy of Sciences of the USA*, 101, 1607–1612.

Zeldes, B.M., Lee, L.L., Straub, C.T. *et al.* (2017) Physiological, metabolic and biotechnological features of extremely thermophilic microorganisms. *Wiley Interdisciplinary Reviews – Systems Biology and Medicine*, 9, UNSP e1377.

Zenuto, R.R. (2010) Dear enemy relationships in the subterranean rodent *Ctenomys talarum*: the role of memory of familiar odours. *Animal Behaviour*, 79, 1247–1255.

Zhang, J., Marshall, K.E., Westwood, J.T., Clark, M.S. & Sinclair, B.J. (2011) Divergent transcriptomic responses to repeated and single cold exposures in *Drosophila melanogaster*. *Journal of Experimental Biology*, 214, 4021–4029.

Zheng, D., Prince, S. & Wright, R. (2003) Terrestrial net primary production estimates in 0.5 degree grid cells from field observations – a contribution to global biogeochemical modelling. *Global Change Biology*, 9, 46–64.

Zhu, J. & Fry, J.D. (2015) Preference of ethanol in feeding and oviposition in temperate and tropical populations of *Drosophila melanogaster*. *Entomologia Experimentalis et Applicata*, 155, 64–70.

Zhu, X.G., Long, S.P. & Ort, D.R. (2010) Improving photosynthetic efficiency for greater yield. *Annual Review of Plant Biology*, 61, 235–261.

Ziemba, R.E. & Collins, J.P. (1999) Development of size structure in tiger salamanders: the role of intraspecific interference. *Oecologia*, 120, 524–529.

Zimmer, M. & Topp, W. (2002) The role of coprophagy in nutrient release from feces of phytophagous insects. *Soil Biology and Biochemistry*, 34, 1093–1099.

Ziska, L.H., Pettis, J.S., Edwards, J. *et al.* (2016) Rising atmospheric CO_2 is reducing the protein concentration of a floral pollen source essential for North American bees. *Proceedings of the Royal Society of London, Series B*, 283, 20160414.

Züst, T., Joseph, B., Shimizu, K.K., Kliebenstein, D.J. & Turnbull, L.A. (2011) Using knockout mutants to reveal the growth costs of defensive traits. *Proceedings of the Royal Society of London, Series B*, 278, 2598–2603.

Zvereva, E.L. & Kozlov, M.V. (2012) Sources of variation in plant responses to belowground insect herbivory: a meta-analysis. *Oecologia*, 169, 441–452.

Zvereva, E.L., Lanta, V. & Kozlov, M.V. (2010) Effects of sap-feeding insect herbivores on growth and reproduction of woody plants: a meta-analysis of experimental studies. *Oecologia*, 163, 949–960.

Organism Index

Page numbers with ‘f’ indicate figures. Page numbers with ‘t’ indicate tables.

Subject Index

Page numbers with ‘f’ indicate figures. Page numbers with ‘t’ indicate tables.