

Oxford Surveys in Evolutionary Biology

VOLUME 8

1992



Edited by
DOUGLAS FUTUYMA AND JANIS ANTONOVICS

**OXFORD SURVEYS IN
EVOLUTIONARY BIOLOGY**

Volume 8

This page intentionally left blank

OXFORD SURVEYS IN EVOLUTIONARY BIOLOGY

EDITED BY
DOUGLAS FUTUYMA
JANIS ANTONOVICS

Volume 8

OXFORD UNIVERSITY PRESS
1992

Oxford University Press

Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Kuala Lumpur Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland

and associated companies in
Berlin Ibadan

© Oxford University Press, 1992

Published by Oxford University Press, Inc.
200 Madison Avenue, New York, New York 10016

Oxford is a registered trademark of Oxford University Press

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, without
the prior permission of Oxford University Press.

Library of Congress Cataloging in Publication Data

(LC Card Number 86-641417)

ISBN 0-19-507623-0

9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper

Preface

We are pleased to present, as the opening essay in this volume, Professor Ernst Mayr's response to our suggestion that a personal history of events and ideas that influenced his thinking, and that reflect the development of evolutionary thought for more than half a century, would be of enduring interest. Both the scientific content of Professor Mayr's essay and the anecdotes, interactions, and personalities that figure in it will interest every biologist and historian of evolutionary biology.

There is a conceptual relationship between Mayr's emphasis on the role of genetic coadaptation in speciation, and Sewall Wright's conviction of the reality and importance of gene interactions. Michael Wade, using a statistical framework, contrasts Wright's and Fisher's views on the likely role of gene interactions, and concludes that such interactions warrant more attention by theoreticians and empiricists alike. Moving from genetic to ecological considerations that bear on speciation, the chapter by Mark Taper and Ted Case reviews the evidence for, and theory of, character displacement. Taper and Case develop theory suggesting that the conditions for character displacement may be broader than some previous models have implied. If, as Mayr has proposed, the capacity for coexistence is an important facet of speciation, the issue of character displacement among competitors should be understood in depth.

More detailed examinations of the origins of certain species are presented in the two chapters that follow. Wen-Hsiung Li and Lori Sadler examine evidence from variation in DNA sequences on the history of *Homo sapiens*, addressing in particular the question of whether or not our species has experienced a substantial bottleneck in population size, and whether the characteristics of modern *Homo sapiens* throughout the world are attributable to replacement of Eurasian by African populations or to gene flow from Africa.

Because of their global significance, crop plants such as rice have been probably more extensively studied with respect to genetics and origins than any naturally occurring species. Studies on crop plants, when interpreted or designed in an evolutionary context, can therefore provide detailed and comprehensive information of a kind that is lacking for other species. In rice, there emerges a rich portrait, painted by Hiroko Morishima, Yoshio Sano, and Hiko-Ichi Oka, of speciation as a process that can involve the interplay of adaptation and coadaptation with gene-level phenomena such as meiotic drive. Pollen killers and gamete eliminators add a star-wars-like drama to processes that a few years ago we might never have imagined to be anything other than gradualistic.

Speciation in plants is the theme, also, of James Bever and François Felber's essay on the population genetics of polyploids. Even though polyploidy has long been accepted as a mode of speciation in plants, its population genetic consequences have remained a rather marginal area of investigation. Bever and Felber point out the often unexpected consequences of polyploid inheritance and emphasize that any explicit theory of speciation by polyploidy must take these genetic features into account.

Speciation is the foundation of biological diversity, a macroevolutionary topic addressed by Warren Allmon, who organizes many familiar elements of evolutionary theory into a framework of isolate formation, persistence, and differentiation that may provide a better understanding of several major macroevolutionary and ecological issues. In particular, Allmon addresses the difficult question of why, and under what conditions, improved fitness (adaptation) of individual organisms may increase the species diversity of a clade. The themes of macroevolution and divergence in isolated populations are pursued by V. Louise Roth, who describes the remarkable phenomenon of dwarfism in island populations of elephants, now extinct. By combining evidence from fossils with an analysis of the development, physiology, and ecology of extant elephants, Roth shows that even seemingly immovable morphologies can evolve rapidly under novel circumstances.

Organismal form, function, and diversity are the subjects, too, of the final essays in this volume. In an age of preoccupation with molecular mechanisms, we do well to remember that the richness of the living world lies in the behavior, physiology, and ecology of organisms – and that these are expressed by their morphological features. Marvalee Wake's comprehensive survey of contemporary currents in morphology shows how intimately related this field is to phylogeny, how functional morphological studies shed light on adaptation and constraints, how profound the questions are about the evolution of complexity, and how the tensions between neo-Darwinism and 'structuralism' might be resolved. In a more physiological and ecological vein, A. E. Douglas surveys the biology of that most neglected form of interspecific interaction, symbiosis. From the level of biochemical mechanisms to that of evolution and ecology, symbiosis has had important effects on life's history: Douglas argues that symbiosis with microorganisms has played a key role in eukaryotes' acquisition of diverse and critically important metabolic capacities.

D. J. F.

J. A.

Contents

Preface	v
List of contributors	ix
Controversies in retrospect <i>Ernst Mayr</i>	1
Sewall Wright: gene interaction and the Shifting Balance Theory <i>Michael J. Wade</i>	35
Coevolution among competitors <i>Mark L. Taper and Ted J. Case</i>	63
DNA variation in humans and its implications for human evolution <i>Wen-Hsiung Li and Lori Sadler</i>	111
Evolutionary studies in cultivated rice and its wild relatives <i>Hiroko Morishima, Yoshio Sano and Hiko-Ichi Oka</i>	135
The theoretical population genetics of autopolyploidy <i>James D. Bever and François Felber</i>	185
A causal analysis of stages in allopatric speciation <i>Warren D. Allmon</i>	219
Inferences from allometry and fossils: dwarfing of elephants on islands <i>V. Louise Roth</i>	259
Morphology, the study of form and function, in modern evolutionary biology <i>Marvalee H. Wake</i>	289

viii Contents

Symbiosis in evolution	347
<i>A. E. Douglas</i>	
Index	383
Content lists of previous volumes	393

Contributors

- WARREN D. ALLMON**, Department of Geology, University of South Florida, Tampa, Florida 33620, USA
- JAMES D. BEVER**, Department of Botany, Duke University, Durham, North Carolina 27706, USA
- TED J. CASE**, Department of Biology, University of California at San Diego, La Jolla, California 92093, USA
- A. E. DOUGLAS**, Department of Zoology, University of Oxford, Oxford OX1 3PS, England
- FRANÇOIS FELBER**, Laboratoire de Phanérogamie, Institut de Botanique, Université de Neuchâtel, ch. de Chantemerle 22, CH 2007 Neuchâtel, Switzerland
- WEN-HSIUNG LI**, Center for Demographic and Population Genetics, University of Texas, PO Box 20334, Houston, Texas 77225, USA
- ERNST MAYR**, Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA
- HIROKO MORISHIMA**, National Institute of Genetics, Mishima, Shizuoka-Ken 411, Japan
- HIKO-ICHI OKA**, National Institute of Genetics, Mishima, Shizuoka-Ken 411, Japan
- V. LOUISE ROTH**, Department of Zoology, Duke University, Durham, North Carolina 27706, USA
- LORI SADLER**, Center for Demographic and Population Genetics, University of Texas, PO Box 20334, Houston, Texas 77225, USA
- YOSHIO SANO**, National Institute of Genetics, Mishima, Shizuoka-Ken 411, Japan
- MARK L. TAPER**, Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA
- MICHAEL J. WADE**, Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA
- MARVALEE H. WAKE**, Department of Integrative Biology and Museum of Vertebrate Zoology, University of California, Berkeley, California 94720, USA

This page intentionally left blank

**OXFORD SURVEYS IN
EVOLUTIONARY BIOLOGY**

Volume 8

This page intentionally left blank

Controversies in retrospect

ERNST MAYR

1. INTRODUCTION

No one who has not witnessed it himself can imagine the confusion and dissension that characterized the views of the evolutionists in the pre-Synthesis period. That we ever got out of this confusion, and that the major step happened so rapidly during the so-called Evolutionary Synthesis, seems almost a miracle. In order to place on record how one pre-Synthesis evolutionist became modernized, the editors of *Oxford Surveys in Evolutionary Biology* suggested to me that I should present ‘a somewhat autobiographical account of how and why your views were formed and changed during your career, and how your own experience and ideas reflect the changes in evolutionary biology during this time.’ Indeed, the story of the controversies of the last 60 years is a fascinating one, and a historical report cannot help but contribute to an understanding of our current views.

I have told on a previous occasion (Mayr and Provine 1980, pp. 413–423) how I became a naturalist, and eventually a museum taxonomist. It is convenient, even though it is a vast oversimplification, to say that in the 1920s the evolutionists fell clearly into two camps: the naturalists–systematists were anti-Mendelians; they believed in gradual evolution, and the origin of organic diversity was one of their chief interests. The geneticists still showed their origin from Mendelism as characterized by De Vries, Bateson, and Johannsen. Their major interest was in the phenomenon of mutation, the change of a given (usually closed) population, and in gene physiology. Even though we were now in the 1920s and beginning the 1930s, the division still reflected the division characterized around 1900 by the terms biometrists and Mendelians. Typological thinking was widespread in both camps, reflected not only in a large number of journal articles, but also in macromutation theories of the important books of Goldschmidt (1940), Willis (1922, 1940), and Schindewolf (1936, 1950). Population thinking made only slow advances, particularly among paleontologists and botanists, and could not be found at all among developmental biologists, a situation continuing right up to modern times. I had taken a rather traditional course in genetics when I was a medical student at the University of Greifswald. If I remember correctly it consisted largely in exercises demonstrating the ‘Mendelian Laws’. The emphasis was on mutation and, as was characteristic for

2 Ernst Mayr

Germany, on physiological genetics. I don't think the connection between genetics and evolution was dealt with at all. Ten years later, in 1934, in my second letter to Dobzhansky, I still upheld the old Darwinian idea that there were two kinds of genetic variation. The differences between *Drosophila melanogaster* and *D. simulans*, I said, were surely something entirely different from the conspicuous mutations, like white-eye, yellow body, crumpled wing, etc. And I agreed with Darwin also – or perhaps rather with the old tradition prevalent among the taxonomists – in the belief that continuous variation was far more important in evolution than the conspicuous mutants.

We young evolutionary taxonomists considered geographic variation perhaps the most interesting of all evolutionary phenomena. It not only demonstrated climatic adaptation convincingly but it also provided the evidence for the theory of geographic speciation. It was a source of particular distress for us that the geneticists had no interest in this phenomenon at all. Consequently, when one of them, Theodosius Dobzhansky, published in 1933 a paper on geographical variation in lady beetles, I was wildly enthusiastic and I did something I had never done before in my life, I sent him a fan letter. I said to myself, here is finally a geneticist who talks like a naturalist. In the ensuing correspondence we both deplored how ignorant the taxonomists were of genetics, and how ignorant the geneticists of the exciting findings of the taxonomists. We were fervently wishing for a synthesis. I presume neither of us had the slightest inkling how close we were to the fulfillment of our wish, initiated by the publication in 1937 of Dobzhansky's *Genetics and the Origin of Species*.

I was at the time an Associate Curator at the American Museum of Natural History in New York, working on that museum's unexcelled collections of birds from the South Sea Islands. There was not a genus nor any widespread species that did not contain clear-cut cases of geographic speciation. My interest in geographic speciation had, of course, been primed by Stresemann and Rensch at the Berlin Museum, but I had never before encountered material documenting geographic speciation quite so graphically as these island birds. A final report on these researches is about ready to go to press (Mayr and Diamond 1992).

My colleague James P. Chapin, the explorer of the Congo birds, was the only person with whom I could talk about these problems; other associates there, like F. M. Chapman, R. C. Murphy, and G. K. Noble, had rather mutationist views, as I have described elsewhere (Mayr and Provine 1980). I read voraciously at that period, not only on systematics and evolution, but also on paleontology, anthropology, behavior, ecology, and genetics. The American Museum had a superb library subscribing to over a thousand journals in the mentioned fields.

I was always most anxious to show my splendid material to others, but

had few opportunities to do so. I showed it to Dobzhansky in 1936, in the summer before he gave the Jesup lectures, and my evidence visibly very much impressed him. At about the same time I showed it to Goldschmidt, when he visited New York, but he entirely ignored it (except in a footnote) when writing his *Material Basis of Evolution* (1940) (see below).

The American Museum of Natural History and Columbia University were in the same city (although 38 city blocks distant from each other!), yet the Department of Paleontology (with H. F. Osborn, Gregory, Simpson, and Colbert) was the only museum department to have any connection with Columbia. I am always asked how I established contact with Columbia. Curiously it was not through Dobzhansky or evolution, but through bird plumages. At that period I was very much interested in the causation of sexual dimorphism in birds and of the differences between juvenal and adult plumages. Through Walter Landauer (Connecticut) who studied similar questions in chickens, I was introduced to L. C. Dunn, professor of genetics at Columbia, who encouraged me to attend their genetics seminars and we became life-long friends. My connection with Columbia became much more intimate after Dobzhansky gave the Jesup Lectures in 1936 and joined the Department of Zoology in 1940.

By that time, of course, the evolutionary synthesis had already begun, initiated by Dobzhansky's *Genetics and the Origin of Species* (1st edn, 1937). Dobzhansky and I, both of us great arguers, discussed problems of evolution by the hour, particularly those of species and speciation. Simpson, even though he worked in the same institution as I and we lunched at the same staff table, never had any scientific discussions with me. Everybody finds this rather curious, considering that we were colleagues for 22 years, but it is a fact.

As I became better acquainted with the evolutionary literature, particularly dealing with the genetic aspects, I became distressed with what seemed to me its great one-sidedness. The emphasis was almost exclusively on the change of gene frequencies in a single gene pool, in a single closed population, as if all evolution was adaptational, best represented as a vertical movement, a change in the time dimension. What attracted me so much to Dobzhansky was that he was an exception. He saw clearly that speciation, the origin of organic diversity, was at least as important an evolutionary problem as adaptation. Yet, even he did not have a chapter on speciation in his 1937 book.

With genetics occupying the major attention in that volume, Dobzhansky did not have the space for a full treatment of the problems of species and speciation and their relation to macroevolution. To provide an expanded treatment of this area of evolutionary biology was the objective of my first book, the outcome of my portion of the Jesup Lectures of 1941 (together with Edgar Anderson). I had presented a preliminary

4 Ernst Mayr

account in my paper on ‘Speciation phenomena in birds’ (1940) in which I demonstrated the high significance of geographic speciation. I am sometimes asked why I said so little about genetics in my *Systematics and the Origin of Species* (1942). The reason is that I felt that I had little to add to what Dobzhansky had said and also that I was rather incompetent to deal with the subject. I had not yet read anything of Fisher or Haldane and knew only two papers of Wright (1940a, 1940b). To be sure I cited some others in the bibliography of my 1942 book, but these were cases of ‘ceremonial citation’, I having learned the titles from Dobzhansky (1937, 1941). In matters of genetics I deferred to Dobzhansky, from whom I had learned most of what I knew about evolutionary genetics. One must not forget that I did not work at a university.

It has become a fashionable sport in recent years, particularly by those who had been too young to participate, to denigrate the Synthesis, to deny its achievements, and, in the worst case (Antonovics 1987), even to argue that it had been a bad thing for evolutionary biology. That such claims are quite unjustified has been demonstrated by Futuyma (1988) and several other recent writers (Eldredge 1985; Levinton 1988; Hoffman 1989). The particular achievements of the Synthesis were a far more general acceptance of natural selection on the positive side, and on the negative side a rather decisive defeat of all the non-Darwinian theories such as neo-Lamarckism, orthogenesis (and other orthogenetic theories), and of saltationism. Others of the achievements of the Synthesis, such as the incorporation of the study of the origin of diversity, will be discussed below. It is remarkable how widely the achievements of the Synthesis are misunderstood even today. For instance, in the books edited by Mae Wan Ho (Ho and Saunders 1984; Ho and Fox 1988), the reductionist Fisher–Haldane kind of mathematical genetics is considered to be representative of the Synthesis. Some developmental biologists still attack the Synthesis because of its support for Darwinian variational rather than developmental transformational evolution.

Historians of genetics have tended to misrepresent the actual achievement of the Synthesis. They have described it as the acquisition by the Darwinian naturalists of a knowledge of genetics, or as one of them has characterized the genuine Synthesis, as ‘the quantitative synthesis of Mendelian inheritance and various factors that can change gene frequencies in populations’ (Provine 1988, p. 56). This narrow-minded interpretation is based on the obsolete idea that evolution is a change of gene frequencies, and a strictly vertical movement. It completely overlooks what Eldredge (1989, p. 207) has correctly stressed, that Dobzhansky and Mayr brought the emphasis on discontinuity into the Synthesis, the problem of the origin of diversity, and of the hierarchical levels above the population. It ignores that evolution is not merely a change in gene frequencies but the origin of adaptation and of new species and higher taxa. It was

implied that the Synthesis consisted of the quantitative modeling of nature, but this is simply not true. Certain components of evolution were indeed studied quantitatively, selection most successfully, but Fisher, Haldane and Wright were far too ignorant about species, speciation, and macroevolution to 'model evolution in nature quantitatively.' Indeed, the quantitative formulation of evolution, based on the assumption that evolution is a change in gene frequencies, was in many respects a retarding element in evolutionary thinking until replaced by a formulation that adequately represented both of the two crucial aspects of evolution: adaptation and diversity.

The three genetical aspects of evolution that were firmly adopted during the Synthesis were (1) that inheritance is hard, there is no inheritance of acquired characters nor any effects of use and disuse or direct influence by the environment; (2) that inheritance is particulate, that is, the genetic contributions of the parents do not blend but remain separate, to be differently recombined in future generations; and (3) that most mutations are very small, not at all the macromutations about which the Mendelians had written. None of these insights were new discoveries of the 1920s, indeed some go back to the 1880s. But it was during the Evolutionary Synthesis that they were universally adopted.

As I pointed out recently (Mayr 1988a, pp. 528–550) the positive achievements of the Synthesis are manifold. As Futuyma (1988) has summarized it, the Synthesis 'reestablished the validity of Darwin's ideas and united them with genetics; it established the validity and usefulness of numerous concepts; and it did establish some communication among fields, especially between genetics and systematics and paleontology. It established the framework within which we work today. It was a synthesis.'

That I was one of the architects of the Synthesis, together with Dobzhansky and Simpson, has never been questioned. However, there has been considerable confusion as to what my particular contribution was. The title of my book *Systematics and the Origin of Species*, chosen at Dobzhansky's suggestion, in fact describes the contents quite well. Some reviewers have also pointed out perceptively that there were actually two books within the covers of this volume, one being a treatment of the principles and methods of taxonomy at the species level. Indeed, the volume was the founding document of the new systematics, there having been virtually no new systematics in the volume Huxley published 2 years earlier under that name.

It was no accident that Dobzhansky and myself, among the then leading evolutionists, had the most advanced understanding of species and speciation. We were both able to take advantage of national traditions: Dobzhansky those of Russia, I those of Germany. I had received my taxonomic training in Stresemann's school, and I was familiar with the writings of K. Jordan and B. Rensch. I have described previously (Mayr

6 Ernst Mayr

and Provine 1980) what Stresemann's ideas had been and how he and Rensch influenced my thinking (see also Haffer 1991). It was in the German literature that in the preceding 20 years important battles had been fought about Formenkreise, borderline cases, the influence of the environment, the limitations of natural selection, etc. I had brought all of this knowledge with me to America, and combined it with the knowledge of American experimental biology. Evolutionary systematics was equally advanced in Russia, and Dobzhansky had brought that tradition with him. Of the 25 papers Dobzhansky had quoted in his 1933 paper, 19 were European.

Most of American animal taxonomy was very traditional, hardly ever rising above the level of Linnaean alpha taxonomy, particularly in entomology. There were some exceptions. J. Grinnell in California, a student of D. S. Jordan, was a profound student of animal diversity, particularly in its ecological setting (Grinnell 1943; Griesemer 1990), but he never did a major analysis of the problem of speciation. This was undertaken by his student Alden H. Miller (1906–1965) in a pioneering analysis of geographic speciation in the genus *Junco* (Miller 1941), but the presentation of his findings in a strictly taxonomic framework was rather unfortunate so that his excellent analysis was generally overlooked (Mayr 1972). The other exceptions were the students of the mouse genus *Peromyscus*, beginning with Osgood (1909), Sumner (Provine 1979), and L. R. Dice. Dice had gathered over many years impressive material on geographic variation and adaptation in *Peromyscus*, some of it summarized by me in 1942, but Dice himself had not yet prepared a general account prior to the synthesis, and a comprehensive manuscript (which I saw in the 1950s or 60s) was apparently never published. In England, in spite of the earlier pioneering work of Poulton and K. Jordan, taxonomy had become very conservative, as documented by the volume of Robson and Richards (1936), and the situation was worse in the French, Spanish, and Italian-speaking countries. When writing *Systematics and the Origin of Species*, I felt myself primarily to be a systematist, and the questions that I asked were quite typically those that would be asked by a taxonomist.

The claim made by one reviewer, that I reduced systematics to population genetics, completely misconstrues the contents of my volume. References to population genetics occur only in 68 lines in the 11324 lines of this volume. To be sure, I had completely adopted the neo-Darwinian concepts of hard inheritance, but not in any reductionist manner. My presentation was anti-Lamarckian, but not quantitative. My major objective was to solve the major problem of evolution left open by the geneticists, the origin of discontinuities; or to put it in different words, the origin of organic diversity.

Evolutionary studies from the first rumblings of evolutionary thought in the 18th century, almost up to the present time, have been plagued

by a peculiar dilemma. The species of the local naturalist (I refer to it as the non-dimensional species) is well defined and separated from all others by a 'bridgeless gap', as it was called by Goldschmidt. On the other hand, it is quite obvious that species must evolve from each other. If this logic is valid, one should have all sorts of gradations between species, and yet where are they, when species are so sharply separated from each other? The pre-Darwinians, who were primarily local naturalists and were familiar with the sharpness of the gap between local species, solved this conundrum by denying evolution. Darwin, on the other hand, solved the problem by denying the reality of species (even though he was rather schizophrenic about this). One of the major theses, perhaps *the* major thesis of my volume, was that species are indeed sharply separated by gaps in the local situation but that if one abandons the typological species concept and replaces it by a concept of species taxa as aggregates of populations, then the gradual evolution of new species is no longer a puzzle, and can indeed be clearly demonstrated. I devoted chapters 6 and 7 to abundant proof for my conclusions.

The genetic literature at that time was almost exclusively devoted to evolutionary change in time (mostly in small closed laboratory populations). What I did, basing my conclusions on a long tradition of European systematics, was to introduce the horizontal (geographical) dimension, and show that the process of geographic speciation is the method by which a gradual evolution of new species is possible, in spite of the gaps in the non-dimensional situation.

When some historians of the genetic establishment (most recently Provine 1988) claimed that genetics deserves most of the credit for the evolutionary synthesis, they cited as part of their evidence the fact that Dobzhansky was a geneticist. But this ignores the fact that Dobzhansky's 1937 book was totally different from the writings of other evolutionary geneticists, such as Morgan and Muller, and of mathematical geneticists, like Fisher and Haldane. What characterized the novelty of Dobzhansky's volume was precisely that he introduced the problems (and their solution) posed by the taxonomists. His basic inferences were taken from the world of the systematists, rather than from that of the geneticists. Speciation, and other processes in evolution, are not simply a matter of genes but of populations and of species. This is the lesson that Dobzhansky and I brought into the Synthesis. Dobzhansky's treatment of species and speciation was by no means revolutionary; it simply made the best thinking of the Russian and German taxonomists available to the American genetic community. In that area I learned little, if anything, from him, because I had grown up in Stresemann's circle exposed to these very same ideas.

My conclusion in 1942 that populations and species are far more important actors in the process of evolution than individual genes

necessitated a thorough survey of the species problem. I shall come back to this presently, but want to point out now only that I raised two arguments against the previously dominant concept of the typological-morphological species. These arguments were so convincing that most working biologists adopted the biological species concept. The one argument was the frequency of sibling species, that is, of morphologically indistinguishable species, which nevertheless have all the biological characteristics of species (and have been shown in recent analyses to be just as different from each other at the molecular level as morphologically characterized species). The other evidence consisted of individual variation, producing variants that are so strikingly different that everybody had considered them species until their conspecificity was demonstrated by their reproductive behavior. This showed that the traditional species criterion, degree of morphological difference, is quite unreliable.

One of my critics has described me as a person 'who worked with museum specimens,' in other words, a closet naturalist as such a person is often contemptuously referred to. The actual truth is that I have spent an enormous amount of time in the field, not only as a young ornithologist, but also in the two and a half years on my expeditions in New Guinea and the Solomon Islands, and every year since my return. In the Solomon Islands I had encountered several cases where allopatric 'species' had had seemingly identical songs, even though they were quite different in their color pattern. I had also observed that songs and call notes varied geographically in some species while in others they remained constant over wide areas. I had also worked with several sets of sibling species, analyzing the numerous differences in vocalization, niche utilization and other aspects of behaviour in spite of seeming phenotypic identity. To consider me only as a museum worker because I have not done any experimental ecological field research is a most misleading characterization. Let us not forget that much of Darwin's *Origin of Species* was the result of his having 'worked with museum specimens.'

If one wants to characterize the importance of my volume in a single sentence, it is that it dealt with a vast area of evolutionary theory that the geneticists were not competent to fill, a discussion of species and speciation, of the role of geography at the level of populations and species, and of the role of species in macroevolution. It did so in far more depth than other publications of the period. To demonstrate how much of importance can be extracted from systematics was particularly important at a period when most experimental biologists considered taxonomy nothing but a kind of postage stamp collecting. It is curious to what extent that period had forgotten that Darwin's *Origin of Species* had been primarily the result of taxonomic research. David Lack's *Darwin's Finches* (1947) and many other publications of the post-Synthesis

period, as well as Stebbins' *Variation and Evolution in Plants* (1950), likewise demonstrated the importance of taxonomic research for evolution.

2. SPECIES

One of the achievements of my *Systematics* had been to show the invalidity of the typological species concept. I had introduced instead the biological species concept, defined in short as 'species are groups of actually or potentially interbreeding natural populations that are reproductively isolated from other such groups' (1942, p. 120). This definition fits the local situation usually extremely well. At a given locality, at a given time, every species is usually sharply delimited against the other co-existing species. This is what I have referred to as the non-dimensional species concept. But there is a conflict between this clear-cut situation and the observation that when a polytypic species is followed through its entire geographic range one may encounter peripheral populations or peripherally isolated populations that are incipient new species. A polytypic species, that is a geographical aggregate of allopatric populations, does not really fit the biological species definition very well, as the opponents of the biological species concept have pointed out with considerable justification. To be sure, this critique somehow confounds the *definition* of the species *category* (which is the definition of the biological species), and the *delimitation* of a polytypic species *taxon* (Mayr 1969a). Nevertheless, there is an undeniable tension between these two aspects of the word species and from 1942 until the present time, I have never ceased to struggle with this problem.

Nevertheless, to return to the typological species of old, as has been recommended by a number of recent authors (particularly cladists and botanists), would seem counterproductive for a number of reasons. First of all, it would lead to an incredible increase in the number of the to-be-recognized species, as was first demonstrated by the botanist P. Jordan, who described hundreds of morphological species within a single Linnaean species. Secondly, a typological-morphological definition is stultified by the large number of cryptic sibling species, which occur in all major taxa of organisms. Finally, a morphological definition is in conflict with the empirical findings of strikingly different individual variants. This is not the place to discuss all the other recently proposed species definitions. I have analyzed them in recent publications (Mayr 1987, 1988a).

Plants, even higher plants, have many complicating features in their reproduction, and botanists, on the whole, have been particularly vocal among the critics of the biological species concept. To be sure, apomixis, hybridization, and autoploidy indeed create difficulties for the

10 Ernst Mayr

biological species concept. However, if one carefully analyzes a local flora (Mayr 1992) one finds that the number of actual difficulties is at most only about 10 per cent. If one deals with a local situation, and this is the area of interest for the student of ecosystems, and in zoology of the students of behaviour, then the biological species concept is more useful than any of the others that have been proposed in recent years.

3. SPECIATION

To explain how new species originate was clearly the most important objective of my *Systematics and the Origin of Species*. The geneticists, except for Dobzhansky, had very much neglected this problem. As is well known, Darwin had at first supported geographic speciation, but had become uncertain when encountering certain situations in plants and had later admitted massive sympatric speciation (Kottler 1978; Sulloway, 1979). And that was the situation I encountered when I first occupied myself with the problem of speciation.

The number of confusions existing at that time was considerable. Many authors, for instance, made no distinction between phyletic change (as observed by paleontologists) and an actual multiplication of species. Other authors, in the tradition of the original Mendelians, still thought of speciation as a phenomenon involving a single individual giving rise to a new species. My own emphasis, therefore, was consistently on stressing the populational aspects of speciation. In my most extreme formulation, speciation was the acquisition of isolating mechanisms by an isolated population. The fact that in the 50 years since the Synthesis no consensus on speciation has yet been reached suggests that there are multiple possible mechanisms and this would mean that one has to determine which is the prevailing mechanism for a particular group of organisms. Being an ornithologist I stressed geographic speciation, and provided massive evidence for the exclusive occurrence of geographic speciation in birds. Indeed, in my extensive work on variation in birds, I did not encounter a single situation that was not easily explained by geographic speciation. Theoretical considerations played no role in my conclusions and, at that time, I did not give much thought to the genetic aspects.

Speciation, being ordinarily a slow historical process, can not be observed directly but must be inferred. In order to be convincing I piled up a very large number of indirect proofs for the universal occurrence of geographic speciation. I was not the first to do so, since Rensch (1929) had likewise produced weighty evidence for it. But my own account was apparently so convincing that for the next 30 years or so, most geneticists accepted geographic speciation as the prevailing mode of speciation. The

population geneticists of the period had little interest in the problem of speciation, since it fell under the heading of 'change of gene frequencies' exactly like adaptational changes. R. A. Fisher knew that geographic isolation was often involved (1930, p. 139) but he also backed the frequency of sympatric speciation as well as a form of speciation that now would be called parapatric speciation. 'In many cases it may safely be asserted that no geographic isolation at all can be postulated.'

Nor did Wright display much interest in the problem of the multiplication of species. He stated correctly that his landscape model has nothing to do with speciation but with 'the evolution of the species as a whole' (1940a, p. 175). Other authors, particularly Dobzhansky, expanded Wright's model of the potential of the genotype to a description of populations and species but this was definitely not implicit in the original treatment by Wright. It is obvious from all of his original discussions of the adaptive landscape metaphor that the shifting balance theory had nothing to do with a theory of speciation. In the 1940s and 50s, whenever a major book on evolution was published, Wright would write a lengthy review. The only exception were books dealing with speciation, because this subject clearly was not of interest to him.

In view of this evident disinterest of the mathematical population geneticists in the problem of speciation, it was particularly important that some of the architects of the Synthesis who came from systematics, Rensch and Mayr, emphasized the importance of speciation. This was the more necessary since the subject was virtually ignored even in Dobzhansky's 1937 volume in spite of its title. There is no chapter on speciation nor is either speciation or origin of species listed in the index. Yet it is speciation that is responsible for organic diversity, and that is the key to macroevolution, as we shall presently point out. In fact, as at least one recent reviewer (Eldredge) has rightly remarked, it is this emphasis on speciation and the origin of diversity that was the crucial component of the Evolutionary Synthesis, not the acceptance of Mendelian inheritance.

Prior to the Synthesis a classification of speciation processes was widely accepted, which distinguished geographic, genetic, and ecological speciation. I rejected this classification, because, as I pointed out, all geographic speciation is also ecological and genetic. Curiously, to pose an alternative of genetic vs. geographic speciation has still been favored by a number of recent authors. It invalidated much of M. J. D. White's argument in his speciation book (1978). He distinguished between geographical and chromosomal speciation, not realizing that all chromosomal speciation takes place in populations, and affects in the course of time the entire population, and since every population is geographically localized, it is simultaneously a geographical phenomenon. The same misconception is found in the writings of Vorontsov and Lyapunova

(1989), who pose the same alternative between chromosomal and geographic speciation. As I pointed out at the Rome conference, all cases of chromosomal speciation can be translated into geographic speciation (Mayr 1982a). The misconception of the separability of chromosomal and populational phenomena led White to propose his theory of stasipatric speciation. Fortunately this is a model of speciation that permits concrete predictions (oasis-like enclaves of new chromosomal types within the range of species), and since these predictions are not met anywhere, the stasipatric model is refuted (Key 1974, 1981; Mayr 1982a).

Numerous modes of non-geographic speciation have been proposed in the course of time. I recognized in 1942 three such modes: (1) semi-geographic speciation (now called parapatric speciation), (2) instantaneous (sympatric) speciation (such as in the case of polyploidy), and (3) gradual sympatric speciation.

The most controversial of the modes of speciation is undoubtedly sympatric speciation. In order to avoid confusion I defined it 'as the origin of isolating mechanisms within the dispersal area of the offspring of a single deme' (1963, p. 451). Critics have frequently accused me of having denied categorically the possibility of sympatric speciation. This is simply not true. Already in 1942 I devoted an entire chapter of 30 pages to non-geographic speciation. In that chapter I concluded that 'species may develop even in bisexual animals by several independent and basically different processes, which have in common only the final result, the origin of new species' (1942, p. 186). In 1963 I gave a very detailed analysis of the problem of sympatric speciation, with particular emphasis on the difficulties into which this model runs, and I pointed out that no well-substantiated cases of sympatric speciation had so far been published. Furthermore, all the reputed cases could be explained quite readily as examples of geographic speciation. I did not deny the possibility of sympatric speciation, but insisted that 'the burden of proof rests on supporters of this alternative mode of speciation' (1963, p. 480). During the reign of the typological species concept when virtually indistinguishable populations of insects (and other invertebrates) were found sympatrically, but either on different host plants or on different structures of the same plant, they were referred to as biological races and were frequently cited as evidence for incipient sympatric speciation. Actually, as I showed (Mayr 1947, 1948) the greatest percentage of these so-called biological races turned out on more careful examination to be sibling species; and there is no evidence that these had originated by sympatric speciation. I was, by no means, the first to question the frequency of sympatric speciation, and I would like to take this opportunity to call attention to the work of Karl Jordan, who preceded me by more than 50 years, and who had proved for case after case the invalidity of claimed sympatric speciation (Mayr 1955).

The incredible species-richness of the tropics has often been explained by special genetic mechanisms (e.g. higher frequency of mutations). My own analysis indicated that it was far more easily explained, at least as far as I am concerned, by ecological factors, such as the nature of ecological barriers and reduced dispersal drive (Mayr 1969b).

4. ADDING TO MY EDUCATION IN GENETICS

I freely confess that my knowledge of genetics was extremely limited when I wrote *Systematics and the Origin of Species*. But after Dobzhansky came to Columbia University I became increasingly in closer contact with him and other Columbia geneticists. Both Dobzhansky and I were very much interested in isolating mechanisms and we both realized that in many animal groups behavioral barriers were the most important ones. I knew the work of J. Phillips (1915) who showed that two duck species, mallard and pintail, produce perfect fertile hybrids, yet virtually never interbreed in their vast area of sympatry, kept apart by behavioral barriers. For a while I kept three species of estrildid finches at the American Museum, planning to do the kinds of experiments later conducted by Immelmann, but logistic difficulties defeated me. When I lamented that there were such technical difficulties experimenting with birds, Dobzhansky suggested using closely related species of *Drosophila* instead. I readily agreed to becoming a behavioral Drosophilologist and in 1944 our first joint paper on the subject appeared. Others, all dealing with what H. Paterson later called the recognition aspect of species, appeared up to 1950 (Mayr 1950). Dobzhansky proposed that we do most of this work in the summer at Cold Spring Harbor, and so, beginning in 1943 and continuing with a few interruptions to 1952, I spent every summer at Cold Spring Harbor. As a result I came into daily intimate contact with M. Demerec, Barbara McClintock, S. Wright, B. Wallace, A. Mirsky, M. J. D. White, E. Caspari, R. Hotchkiss, the French visitors (Ephrussi, Lwoff, Monod, Latarjet), and the phage group (Delbrück, Luria, J. Watson, etc.). Many of them, particularly Caspari and Delbrück, became close family friends. Some historians have been puzzled how I, a museum ornithologist, could have acquired such a wide familiarity with modern genetics and molecular biology. The summers in Cold Spring Harbor are the answer.

The stimulation for my own theorizing was enormous. I was still bothered throughout this period by the last remnants of the biometrician–Mendelian controversy. Obviously I did not accept blending inheritance but I strongly felt that the strictly atomistic–reductionist approach of R. A. Fisher did not tell me the whole story. Every naturalist, I suppose, has a feeling that a purely reductionist approach fails to explain the holistic aspects of organisms and of evolution.

Some geneticists, in the 1940s and 1950s, came to the same conclusion, even though by a very different route. Dobzhansky had sympathies in this direction, but even more so his students Wallace, Brncic, and Vetukhiv. K. Mather's experiments led to similar conclusions, but M. Lerner became the leader of this school of holistic thinking (Lerner 1954). I felt that he had demonstrated convincingly the importance of non-additive genes and of the genetic homeostasis of the genotype as a whole. In Cold Spring Harbor it was Bruce Wallace's work with irradiated populations that opened my eyes to the existence of a remarkable cohesion of the genotype. He and I spent days and weeks discussing these problems.

I attended in 1953 a small symposium on populations genetics in Pavia (Buzzati-Traverso 1954) where Mather, Lerner, Wallace, and Scossioli presented their findings. Their evidence was emphatically questioned by R. A. Fisher, but on the evening of the second day he said: 'Gentleman, if you are right, then I can no longer calculate it.' This dilemma was, of course, part of the reason for the opposition of the advocates of the near all-sufficiency of additive inheritance. This Pavia meeting strengthened my ideas on the subject, ultimately presented by me in great detail in Chapter 10 (pp. 263–296) of *Animal Species and Evolution* (1963) (an improved version is in Chapter 10 of Mayr 1970).

Most population geneticists at that time ignored these ideas, indeed considered them to be 'vitalistic' or otherwise obnoxious. But here they were, and I had the feeling that they were of considerable importance for many evolutionary phenomena, such as stasis, rates of evolutionary change, and even speciation.

Empirical observations on speciation that I had made for more than 25 years now began to acquire a new significance. I began to realize that there are really two kinds of geographic (allopatric) speciation. According to one, presented at that time in all treatments of speciation, a widespread species is divided by a newly arising geographical or ecological barrier into two isolated portions. These henceforth evolve independently of each other, and if sufficient genetic differences accumulate, they will eventually become two separate species. M. J. D. White (1978) has referred to this mode of geographic speciation as the dumbbell model, and Cracraft (1984) has called it *dichopatric speciation*. This may well be the usual form of speciation on continents, but whenever populations of a species tend to have an insular distribution pattern, even on continents, a different mode of speciation occurs, which I first described in 1954 and designated in Mayr (1982a) as *peripatric speciation*. Two observations had led me to this model. The first one was that I had found in my work on South Sea island birds that the peripherally most isolated members of a polytypic species or superspecies tended to be the most distinct. Peripheral isolation clearly may lead to greater divergence than is ordinarily found among conspecific populations.

I combined this observation with the widely held view that many new populations beyond the periphery of a species are established by a few founders, and indeed very often by a single fertilized female. Such a new population contains only a fraction of the genetic variability of the parent species, and owing to inbreeding, it will experience even greater genetic homogenization (loss of deleterious recessives). I postulated that this would permit the setting up of new epistatic balances offering novel possibilities for selection. In extreme cases a rather drastic genetic restructuring might occur, which I designated with the perhaps unfortunate name *genetic revolutions*, resulting in species that in spite of their very close relationship to the parental species (in terms of generations of separation) might differ rather drastically not only in morphology but also in niche occupation or adaptation. Although at first generally denied, this bottleneck effect is more and more frequently acknowledged by students of population genetics (Newman *et al.* 1985; Bryant *et al.* 1986; Goodnight 1987).

Some critics, who had not read my paper carefully, accused me of having postulated 'an avalanche of mutations' and a deterministic postulate of saltational innovations. In a later paper (1982b) I refuted these misconceptions, and I reiterated, in particular, that most founder populations hardly differ at all from the parental population, but that when a rather divergent new taxon originated somewhere, it was usually the result of peripatric speciation. I had virtually no concrete evidence for the genetic phenomena postulated by me, but ultimately H. Carson and his associates (Carson 1975), when studying speciation in Hawaiian *Drosophila*, found very much what I had postulated.

Curiously, the paper (1954) in which I had developed my new ideas and in which I had particularly called attention to their explanatory value as far as macroevolution is concerned, was totally ignored even though Lewis (1962) had independently reached similar conclusions for plants. Finally Eldredge (1971) confirmed that my model would explain the seeming saltations that are found in the fossil record. He and Gould based on it in 1972 their theory of punctuated equilibrium (speciational evolution). This led to a heated controversy that has not yet terminated. In my opinion, which I have presented in 1988, there is little doubt as to the validity of *speciational evolution*, provided one keeps in mind that all the changes occur in populations and are therefore gradual. There is no conflict whatsoever between Darwinian and speciational evolution and between the rapidity with which it can generate macroevolutionary events and Darwinian gradualism. As subsequent authors have pointed out (Huxley 1982; Rhodes 1983), Darwin was well aware of strong differences in evolutionary rates, and would not have found this model in conflict with any of his ideas.

In 1942 I recognized one additional possible form of speciation, referred

to by me as semigeographic speciation. It is now usually called parapatric speciation, because according to those who support this theory, isolating mechanisms may develop in a contact zone between interbreeding populations, eventually leading to a complete split. I was not very enthusiastic about this possibility in 1942, and all the work I have seen since, particularly the excellent genetic analyses of various such zones of intergradation or hybridization (Barton and Hewitt 1985; Butlin 1985; Otte and Endler 1989), indicate the high improbability of the development of reproductive isolation in a contact zone between the two populations. There are thousands of such hybrid belts in vertebrates, insects, and other animals, most of them leftovers of the expansion of refuge populations after the last glaciation. Some of them are clearly many thousands of years old, such as that between carrion and hooded crow (*Corvus corone* and *C. cornix*), which is estimated to be 8000 years old without any indication of the development of reproductive isolation.

It now seems that in higher animals the only well substantiated form of speciation is that of allopatric speciation, with the dumbbell (dichopatric) model perhaps prevailing on continents and the peripatric speciation model in insular regions. The development of biological races in large artificial monocultures is perhaps the nearest approach to sympatric speciation (Feder 1988).

5. THE ORIGIN OF REPRODUCTIVE ISOLATION

The co-discoverers of natural selection, Darwin and Wallace, disagreed on the role of natural selection in the origin of reproductive isolation between developing species. Wallace was sure that sterility was acquired by natural selection, but Darwin said he could not see how this could be effected (Mayr 1959a; Kottler 1978; Sulloway 1979). At the time of the Synthesis the argument was still undecided. Dobzhansky (1937) and Fisher (1930, p. 145) favored Wallace's viewpoint, while Muller (1940) sided with Darwin. And Wright (1940a, p. 176) thought, like Dobzhansky, that partial isolation could be completed by selection when two incipient species established contact. At first I was undecided on this issue, but as time went on I became more and more impressed by the evidence that Darwin and Muller had gathered, demonstrating the utter improbability that selection could re-enforce isolating mechanisms. The old hybrid zones of which taxonomists have found literally thousands in organisms of the most various taxa are perhaps the best evidence against the ability of natural selection to build up isolating mechanisms.

A main reason for the support of the Wallace–Dobzhansky model was no doubt that its supporters could see no other means by which such an

ad hoc mechanism as the isolating mechanisms are, can be produced. The supporters of the Darwin–Muller explanation were forced to postulate that cross-incompatibility developed fortuitously as an incidental by-product of the divergence of the two isolated populations, particularly as a by-product of new ecological adaptation. This was a legitimate explanation, but it really amounted to whistling in the dark, because it was difficult if not impossible to find any direct evidence for this hypothesis. As far as chromosomal changes are involved, and for certain other isolating mechanisms, this explanation is presumably the true one. However, when it comes to behavioral isolation, which in higher animals is no doubt the foremost source for isolating mechanisms, there was no direct evidence whatsoever. This all changed in recent years owing to the greatly increased interest in sexual selection that has developed since about 1972. This led to the hypothesis that females may develop new preferences in small isolated populations, and that through sexual selection this will lead to the development of new recognition patterns (Kaneshiro 1980; West-Eberhard 1983; Thornhill and Alcock 1983; Dominey 1984; Mayr 1984a). Owing to a change of function, these sexual preference features become species recognition marks when two incipient species first come into contact with each other. What is historically interesting is that, except for Dobzhansky and Muller, the population geneticists seemed to have had little interest in this problem, not surprisingly so considering their lack of interest in the whole problem of the origin of diversity. Wright (1940a) in his long article on speciation in Huxley's *The New Systematics* (1940) nowhere mentions the problem of the origin of isolating mechanisms. Nor have I found this problem dealt with by Haldane.

One of the most surprising aspects of the evolutionary synthesis was the almost total neglect of sexual selection from the 1920s to about 1969. To be sure Fisher (1930) had discussed some of its dynamics but for more than 50 years it failed to become an active research program. The reasons for this neglect have not yet been studied by the historians. The main reason perhaps was that the mathematical geneticists, by adopting the gene rather than the individual as the target of selection, had eliminated the difference between natural and sexual selection. Either kind of selection would result in a change in the frequency of genes in the gene pool of the next generation. However, it totally failed to convey the fundamental difference between selection for ‘fitness’ (traditionally defined survival selection or natural selection) and selection for mere reproductive success. Even though natural selection is the selection of individuals, it almost automatically benefits also the species as a whole, while selfish selection for mere reproductive success very often leads to characteristics or reproductive strategies that are not necessarily

advantageous in the long run, e.g. size of the bulls of elephant seals). The previous neglect of selection for reproductive success has been significantly repaired in the last 20 years.

6. MACROEVOLUTION

Genetics has no way of dealing with higher taxa, since they cannot be crossed with each other. By necessity, therefore, the geneticists who wrote about evolution before or during the Synthesis, were forced to ignore macroevolution. Even Dobzhansky, who, as a taxonomist, had a good deal of understanding of macroevolution, does not deal with it in *Genetics and the Origin of Species* (1937). It was the major contribution of Simpson, Rensch, and, for plants, Stebbins, who have brought macroevolution into the Synthesis. Even though I was primarily a microtaxonomist, that is, a student of the taxonomy of species and infraspecific entities, I nevertheless devoted half a chapter of my *Systematics and the Origin of Species* (1942) to macroevolution.

The basic problem of macroevolution at that time, and to some extent right up to the present day, was to reconcile the seeming conflict between Darwin's gradualism and the observed discontinuities in organic nature and in the fossil record. Simpson (1944) attempted to eliminate the seeming contradiction by introducing the concept of quantum evolution, a sudden jump in a phyletic lineage from one to another of Wright's adaptive peaks. As much as Simpson was opposed to the kind of macromutations that Goldschmidt and Schindewolf were using to explain the bridging of seeming gaps in evolution, basically Simpson's explanation was curiously similar to theirs. Owing to considerable criticism, Simpson in 1953 very much reduced reliance on such quantum evolution, but since he failed to use the process of speciation as part of the explanation of macroevolutionary changes, Simpson really never succeeded in providing an explanation for seeming macroevolutionary saltations. For Simpson, macroevolution was always essentially a vertical, a transformational problem.

I realized quite early that species (and still higher taxa) were the 'units of macroevolution.' In particular, I showed that many genera contained somewhat aberrant species that could have just as well been (and often were) called different genera. More specifically it was geographic speciation that was the crucial step in the evolution of new higher taxa.

For me it became clear in 1951 (published in 1954) that geographic variation and the origin of numerous isolates was ultimately the solution to the conundrum of macroevolution. While the evolutionary rate of a widespread populous species was very slow (indeed, owing to strong stabilizing selection such a species might appear to be completely static),

founder populations, going through a genetic bottleneck, might provide the opportunity for greatly accelerated evolution, and indeed for rather drastic evolutionary shifts. I therefore postulated that

. . . the genetic reorganization of peripherally isolated populations, . . . does permit evolutionary changes that are many times more rapid than the changes within populations that are part of a continuous system. Here then is a mechanism that would permit the rapid emergence of macroevolutionary novelties without any conflict with the observed facts of genetics . . . as soon as such a population has completed its genetic reconstruction and ecological transformation, it is ready to break out of its isolation and invade new areas. Only then will it become widespread and thus likely to be found in the fossil record. (Mayr 1954)

Since selection pressures play a great role in founder populations, owing to their new environment, the suggestion has been made that the greatly accelerated evolution (and speciation) may simply be due to intensified selection without any major restructuring of the genotype. There are a number of reasons why I did not adopt this interpretation. First, there are pronounced ecological shifts within many widespread populous species but without any pronounced phenotypic effect on the ecologically shifted populations. Secondly, the majority of peripheral isolates, although occurring in quite novel environments, show no evidence of any major evolutionary departure, and third, some of the most drastic departures in Northern Melanesian birds such as *Phylloscopus amoenus* (Kulambangra), *Dicaeum tristrami* (San Cristobal), and *Dicrurus megarhynchus* (New Ireland) occur on islands that neither in the physical nor the biotic environment differ more drastically from other neighboring islands, than they from each other. There is no indication that any selection pressure *per se* could account for the innovations.

Eldredge (1971) and Eldredge and Gould (1972) have made my theory the basis of their theory of punctuated equilibria. What was particularly new in their theory was the emphasis on the stasis of most species after they had passed through the early speciational stage. When some of the excesses of punctuationism are removed, the theory of punctuated equilibria, or, as I prefer to call it, of speciational evolution, fits the basic Darwinian principles exceedingly well (Mayr 1988a). It shows that evolution is always populational, that is, gradual, and that, even though stochastic processes (chance) are invariably accompanying the speciational process, nevertheless natural selection is involved continuously, that is, in every single generation. The model of speciational evolution also shows why this process is so unpredictable, and why it allows such a smooth transition from intrapopulational phenomena (gene frequencies) to macroevolutionary innovations and trends.

Having worked with the often dramatic but always strictly gradual changes in geographic speciation, I have always been an inflexible foe of

saltationism. This is one of the reasons why I was so opposed to Goldschmidt's hopeful monster theories. It was also the guiding principle in my paper on evolutionary novelties (Mayr 1960). I showed that the Darwin-Dohrn principle of change of function was the explanation of many seemingly inexplicable macroevolutionary saltations. Whenever one encounters in evolution the sudden appearance of a seemingly new structure, one must always ask oneself whether it is not something that had existed all along but had simply acquired a new function. I had written the paper on evolutionary novelties for the Chicago Darwin celebration. The organizing committee had asked me to lecture on speciation but I was so sick and tired of that subject that I chose a theme that would not have anything to do with speciation. Little did I suspect that 25 years later the change of function principle would make a major contribution to speciation research. If the new theory is correct that behavioral isolating mechanisms can originate by sexual selection, and only secondarily function in species recognition, then this is a classical case of change of function (Mayr 1988b).

7. BEANBAG GENETICS

For the naturalists, from Darwin on, it always was the individual as a whole that was the target of selection. For most geneticists, particularly the mathematical population geneticists, it was the gene. (Wright's views will be discussed presently.) Their definition of evolution as a change in gene frequencies resulted in a preoccupation with the isolated gene, and an almost exclusive preoccupation with additive inheritance (even overdominance is an allelic interaction). I objected to this one-sidedness in a number of preliminary papers (1952, 1956) but these failed to evoke any response. The Darwin Centennial in 1959 provided me with an opportunity to spell out my views in far greater detail. I was asked to give the keynote address at the Darwin Symposium organized by the Cold Spring Harbor Laboratory. I decided that it might be interesting if I, a non-geneticist, would spell out what to an outsider would seem to be some of the deficiencies of current evolutionary genetics. Since they were subsequently frequently misrepresented, I would like to quote some of my comments here verbatim. I said that there was an early population genetics that was largely mathematical 'indicated by the names Fisher, Wright, and Haldane' and a newer evolutionary genetics 'indicated by such names as Sumner, Chetverikov, Timofeeff-Ressovsky, and Dobzhansky' with its roots in population systematics.

The emphasis in early population genetics was on the frequency of genes and on the control of this frequency by mutation, selection, and random events. Each

gene was essentially treated as an independent unit favored or discriminated against by various causal factors. In order to permit mathematical treatment, numerous simplifying assumptions had to be made, such as that of an absolute selective value of a given gene. The great contribution of this period was that it restored the prestige of natural selection. . . Evolutionary change was essentially presented as an input or output of genes, as the adding of certain beans into a beanbag and the withdrawing of others. This period of 'beanbag genetics' was a necessary step in the development of our thinking, yet its shortcomings became obvious as the result of the experimental population geneticists, the animal and plant breeders, and the population systematists, which ushered in a [new] era of evolutionary genetics.

It is easy to see in these words the influence which the thinking of the holistic geneticists (Dobzhansky, Lerner, Wallace, etc.) had had on me.

Sewall Wright, who had been in the audience, congratulated me after my lecture. However, when he saw the printed version, he discovered that I had bracketed him with Fisher and Haldane as a beanbag geneticist. I do not know how many papers he has published since that time, devoted to a rejection of my classification and to a restatement of his belief in the importance of epistatic inheritance. Of course he was right and I was wrong. Wright indeed had emphasized the importance of epistatic interactions all along. In 1940 he had pointed out that 'the assignment of selection coefficients to individual genes does not give as realistic a representation of natural selection as is desirable. It is the organism as a whole that is more or less adaptive in relation to prevailing conditions, not single genes' (1940a, p. 165). 'A gene that produces a more favorable effect than its allele in one combination is likely to be less favorable in others' (1940a, p. 165). Yet, because like Fisher he was unable to calculate the contribution of these epistatic interactions, when it came to the equations and the illustrations of his papers, they were largely beanbag genetics. In his well known paper on isolation by distance, Wright (1943), so far as I can see, deals only with additive inheritance, and all of the 70 or more formulae published in this paper are based on the assumption of pure additive inheritance.

In the winter 1959–60 I spent 3 weeks with my friend J. B. S. Haldane in India, and we discussed at length the question to what extent population genetics was still beanbag genetics. As a result, Haldane published in 1964 a paper entitled 'A defense of beanbag genetics.' As far as I was concerned, the paper really missed my major point. I had nowhere denied (indeed I even had stated this specifically) that as a first approach the evolutionary contributions of additive genes would have to be studied. Hence, Haldane's demonstration of the importance of some additive gene analysis required no defense in my opinion. The point I had made was that beanbag genetics was not enough, and that the time had now come

to study the interaction of genes, because it was the genotype as a whole (and the phenotype produced by it) that was the target of selection.

In my 1959 paper, and in others of my writings, I credited the mathematical geneticists with having established the power of natural selection. I was far less certain what other contributions the mathematical approach had made. It was genuine curiosity rather than sneering skepticism that induced me to ask 'but what precisely has been the contribution of this mathematical school to the evolutionary theory, if I may be permitted to ask such a provocative question' (1959b, p. 2). Naturally, my question was furiously attacked, but I failed to evoke a genuine answer to it. Haldane's defense of beanbag genetics certainly was not it, nor were Wright's various papers. Presumably the purely additive approach did not permit it.

Genetics has changed a great deal in the 30 years since my beanbag paper, and in a number of different ways. Modern computers now permit the calculation of things that Fisher could not calculate. Also, and this is far more important, we have now learned that there are many different kinds of genes, some that are neutral or quasi-neutral, and others that can make a major contribution to evolutionary change. It seems to me, however, that the contributions did not come so much from mathematics as from molecular biology. In the meantime, I continued to hammer away at the cohesive aspects of the genotype, and devoted Chapter 10 of my 1963 and 1970 books to this subject. In Mayr (1975) I claimed 'free variability is found only in a limited portion of the genotype.' Most genes are tied together into balanced complexes that resist change. Actually, I had very little concrete evidence for making this claim. It was merely an inference from observed phenomena in populations and species. About at the same time Carson developed similar ideas (1975).

Charlesworth, Lande, and Slatkin (1982) of the additive camp published a spirited defense of the traditional view of mathematical genetics. Although this paper deals primarily with aspects of macroevolution (and there I entirely agree with the conclusion of the authors), it does attempt to solve all evolutionary problems without special reference to non-additive interactions. However, several geneticists (cited above) have demonstrated in recent years the importance of bottleneck effects, and it seems that the role of epistatic interactions and of the breaking up of the cohesive gene complexes are beginning to be studied in earnest.

8. NATURAL SELECTION

The Mendelians did not much believe in natural selection. The pre-Synthesis naturalists accepted natural selection but were convinced that it was only one of several mechanisms of evolution. Indeed, for some of

them the other factors were so much more important than selection that they hardly mentioned it. Rensch, for instance, one of the architects of the Synthesis, listed four suggested explanations of speciation, mutation, orthogenesis, hybridization, and inheritance of acquired characters, without any mentioning of selection at all (1929, p. 2). As I have described elsewhere (Mayr and Provine 1980) I grew up believing in some selection, but also believing (in the 1920s) that all gradual differences, which included most of those involved with geographic speciation, could not be explained by the mutation-selection theory.

To my great amusement, while this paper was being edited, I discovered in an old college notebook a statement of mine entitled 'Credo of a Lamarckian,' representing my ideas, dated 2 February 1926. It consists of the following statements:

First. The controversy in evolutionary theory today does not concern the question of selection. Selection is also acknowledged by the Lamarckian.

Second. The controversy relates to the cause of variability (which Darwin accepts as given).

Third. Varieties (heritable) originate according to De Vries by random mutations. The mode of life of an organism is determined by the structures thus originating. (This seems to be the view of some evolutionists, particularly of experimental zoologists and geneticists.)

Fourth. The Lamarckian, on the other hand, claims that new variants originate under the influence of the mode of life (Lamarck) or of the environment (Geoffroy).

Fifth. The Lamarckian has the right to interpret the laws and findings of genetics in his sense; this refers both to mutations and Dauermodifikationen.

Sixth. While the experimental biologist states that mutations are undirected (random), the Lamarckian asserts mutations on the way to an adaptation are directed. It would make no sense to believe that the destabilization of the germ plasm (caused by a mutation) would be without influence on subsequent mutations.

- Orthogenesis.

Seventh. Changed conditions of the environment influence the reaction of the soma (body plasm) (modification of the phenotype). If this influence continues for a lengthy period the germ plasm will also be influenced, the modifications become heritable, they become *Dauermodifikationen*, which after return to the normal environment will disappear only after many generations (cumulative aftereffect).

Eighth. There are no proper arguments against Lamarck's claim that organs deteriorate by lack of use. It is in line with the economy of the organism that of the available 'fond' (Hesse) [resources] particularly those organs will be endowed which are very much in use (Roux, *The Struggle of Parts in the Organism*). On the other hand it is reasonable that in the organs which are used most actively corresponding to the degree of use, mutations will occur, the maintenance of which will be controlled by selection.

Ninth. The Lamarckian theory (in its modern version) is not teleological.

It is quite evident that this was written in Berlin, where at that time the Dauermodifikationen of Jollos were held in high esteem. Their refutation (which did not really happen until the 1930s) pulled out the rug from under much Lamarckian speculation.

Like Rensch, however, I learned and was persuaded in the years preceding the Evolutionary Synthesis that indeed selection was the crucial factor explaining adaptation and the modification of populations. However, the mere acceptance of the importance of natural selection did not solve all the problems, as is evident even from reading the current literature. Of the numerous problems relating to selection, I will take up a few that were of particular concern to me. Since I have published a good deal in recent years on various aspects of selection, I will omit all detail, in order to avoid too much duplication (Mayr 1982c, 1985, 1986, 1988a, 1991b). I will, however, present the gist of my views.

8.1 The target of selection

It is very evident that for Darwin it was the individual that was either favored or discriminated against by selection; this has been the almost universal view of the naturalists ever since. However, in due time, the gene became the unit of selection for most geneticists. For some naturalists, assemblages of individuals – ‘groups’ – were also the object of selection, and there was a good deal of rather thoughtless referring to selection for ‘the good of the species.’ At any rate, in a review of the subject, Lewontin (1970) recognized multiple hierarchically arranged units of selection.

My own reaction was that, to begin with, the term unit of selection was ill chosen. In science, in general, a unit designates usually something that can be quantitatively expressed, like the units in electricity or mechanics. In line with this tradition, Haldane introduced the term darwin as the unit of rate of evolutionary change. I have elsewhere stated in detail (Mayr 1986) my various objections to the ambiguity of the term ‘unit of selection.’

8.2 Group selection

I cannot see group selection anywhere except in species that have social facilitation. The problem with group selection is that most authors have not made any discrimination between soft and hard group selection. Only the latter deserves to be called group selection (Mayr 1990a).

Social and cultural groups are, in my view, the only groups that as such can be the target of selection. Since the human species is one of the few species clearly forming cultural groups, group selection does

occur in the human species, as has often been demonstrated in the decimation or even extinction of human ethnic groups by others.

9. NEUTRALITY AND PLURALISM

No revolution is ever completed nor any synthesis. The version of the Darwinian theory accepted during the Synthesis left numerous problems open, and was rather uncertain about the relative importance of various factors. Was there a definite trend as a result of the ensuing fine-grained analysis? Gould (1983) has asserted this in his claim of the 'hardening of the synthesis.' By this he meant that selection was given an ever-greater role in the post-Synthesis period. Dobzhansky and Simpson had at first ascribed a rather important role to chance in the course of evolutionary change. They spoke of non-adaptive factors, and Simpson even of a non-adaptive phase. This elicited considerable criticism, particularly by Fisher and E. B. Ford. As a result Dobzhansky and Simpson rather softened their statements in subsequent publications, now ascribing to selection much of what they had previously ascribed to chance. This is the hardening of which Gould has written. I do not know of a corresponding change in any other author.

One must not forget that these were the decades in which the former Lamarckians, persuaded by the Synthesis, had given up their belief in soft inheritance and had adopted the Darwinian paradigm. Consequently selection played an ever-greater role in the literature. Although I myself had completely accepted natural selection by the time I wrote my *Systematics*, I still thought I could find neutral polymorphism to be more common than suggested by Ford (1940) and I suggested reasons for such neutrality (1942, p. 75). However, I realized within a few years after 1942 that the very fact of certain morph ratios remaining constant over long periods of time or over wide geographical areas was *de facto* proof of the selective control of these ratios. If they had been due to chance these ratios should have fluctuated a great deal over time and space.

At that time I was much more of a selectionist than Dobzhansky and I had long arguments with him about the existence of neutrality. For instance he had assumed at first that the gene arrangements in *Drosophila pseudoobscura* were neutral until his own findings proved otherwise. I contended that the pattern and everything else we knew about their lifestyle indicated great dispersal facilities for *Drosophila*, and I went so far as to suggest that an accidental dispersal of a thousand miles was a possibility (Mayr 1944). This was considered totally absurd by Epling and other *Drosophila* researchers until the Colombia endemic population of *D. pseudoobscura* was discovered, substantiating my hypothesis. Dobzhansky, following the suggestion of Wright, considered the human

blood groups as typical cases of neutral evolution. I, by contrast, was convinced that the observed clines in their distribution and various other aspects of blood group distribution indicated a contribution by selection. After my arrival at Harvard in 1953, with the help of Dr Louis Diamond, the hematologist of the Children's Hospital in Boston, and of my wife, I undertook an analysis of thousands of blood groups and we found indeed a number of pathological conditions where the blood group frequency deviated drastically from that of the Boston population (Mayr *et al.* 1956; Mayr and Diamond 1959). While working on these problems, other teams demonstrated the selective influences on blood group frequencies so much more convincingly (Aird *et al.* 1953; Buchwalter *et al.* 1956) that I did not continue this research.

Yet, I was not a rabid selectionist, as is documented by my statement that 'it should not be assumed that all the differences between populations and species are purely adaptational, and that they owe their existence to their superior selective qualities . . . many combinations of color patterns, etc., are probably largely accidental' (1942, p. 86). At that time I was still rather confused about the selective value of phenotype vs. genotype. For instance, I would say that it was of no selective significance whatsoever whether a red ladybug beetle had three or five black spots on its elytra; while of course the gene causing the difference might have certain pleiotropic physiological effects which definitely were of selective significance. It is quite evident, however, that with the discovery of the selective significance of blood groups, gene arrangements, and the polymorphism of *Cepaea* snails (so beautifully worked out by Cain and Sheppard) one became quite hesitant to designate any variable as accidental or neutral.

All this changed, however, at least as far as the gene level is concerned, when Hubby and Lewontin (1966) and Harris (1966) discovered the enormous variation of enzyme alleles. This surprised me as much as most geneticists. This and other discoveries soon led to the theory of neutral evolution by Kimura (1968, 1983) and King and Jukes (1969). At first this seemed to be a theory clearly in conflict with Darwinism. Eventually, however, it became evident that this seeming conflict was only an artifact of terminology. For many geneticists, the gene is the target of selection and the unit of evolution. Therefore, neutral changes at the gene level are neutral evolution, and such neutral evolution is not the result of natural selection. In reality, however, it is the individual as a whole that is the target of selection, and provided an individual has an overall superior genotype and is selected for this genotype, the number of neutral replacements of alleles that occur in such a genotype, is irrelevant for manifest evolution. Such changes will be taken along as hitchhikers of the superior genotype. Kimura misses this point when he claims that 'advantageous mutations may occur, but the neutral theory assumes that

they are so rare as to be negligible' (*New Scientist* 11 July 1985, p. 47). In fact, they are the truly important changes in evolution.

This is not the place to discuss the unresolved conflict that has arisen over the relative significance of neutral evolution and 'real' evolution. There is no doubt that if one wants to measure branching points in the phylogeny, that is, the chronology of such branching points, one might use the molecular clock, which in turn is most reliable when neutral allele replacements are used. On the other hand, if one is interested in adaptive changes, in the origin of new higher taxa, and in the acquisition of evolutionary novelties, one will concentrate on genetic changes of evolutionary significance. This is a problem that has special relevance for systematics, because the 'general reference systems' of Hennig (1950) rely on branching points, which, to a large extent, may be the result of neutral evolution, while traditional taxonomy is based on degrees of similarity and of phylogenetic divergence (Mayr 1990b, 1991a). It is yet uncertain what kind of a compromise can be reached between these rather different views of classification.

10. THE CURRENT STATE OF THE SYNTHESIS

Until about 1970 the Synthesis was remarkably successful. It was almost universally accepted by evolutionists, except for some sniping by members of the pre-Synthesis establishment. When Goldschmidt published his *Material Basis of Evolution* in 1940, it was quite obsolete in spite of some individual brilliant ideas. The volume rather angered me, because of Goldschmidt's total disregard of the evidence for geographical speciation. He had visited me at the American Museum several times (from 1936 on), and I had demonstrated to him how geographic speciation could lead not only to a new species but to such that were so different that they were ranked as different genera. He ignored all that in his book, except for a short comment in parenthesis '(for conclusions opposed to those drawn here see the numerous ornithological papers by E. Mayr 1924–1939).' He had received from me all the papers on which my 1940 review paper on speciation was based. I suppose he was forced to ignore this evidence, first because he was quite unable to think in terms of populations, and secondly, because if he had accepted my conclusions it would have been a refutation of his own viewpoint.

After about 1970, particularly in connection with the punctuation controversy, more and more often the claim was made that the Synthesis was obsolete, that we needed a new paradigm, and that we were at 'a crossroads' (Greene 1990). The most vocal of the critics were developmental biologists, some paleontologists, and continental European morphologists

(see below). The developmental biologists complained that the Synthesis had ignored them. This was really a rather silly claim, because in the 1940s anybody was welcome to join the Synthesis, but the developmental biologists, almost to the last one, vigorously opposed Darwinism and upheld the ideal of transformational evolution. It was they who refused to join the Synthesis rather than the Synthesis keeping them out. When one analyzes their criticisms one discovers that these authors made no distinction between proximate and evolutionary causations, that, as I have said, they still believed in transformational evolution, that they had no conception of the populational nature of evolution, and almost all of them indulged in poorly concealed teleological thinking. Mae Wan Ho has been prominent among the critics, and her writings (e.g., Ho and Saunders 1984), perhaps better than those of anyone else, illustrate the misconceptions still dominant in the camp of developmental biology (Mayr 1984b). Waddington was the only western developmental biologist who joined the Synthesis. He was uneasy about it, and never fully understood the populational aspects. Also, throughout his life he continued to flirt with soft inheritance; at least this is what I read between the lines of his papers on genetic assimilation. His rather typological thinking is most evident in his approach toward what he called 'theoretical biology' (Waddington 1968–73).

The same is true for morphologists, like Gutmann in Frankfurt (Gutmann and Bonik 1981), or Riedl's school in Vienna (Riedl 1975). These schools have made major advances in the analysis of morphogenesis, but I entirely fail to see where any of their findings are in the slightest conflict with Darwinism, as they claim.

Where they are unfair to the architects of the Synthesis is with their claim that we had said that the Synthesis had been the last capstone of Darwinism, and that all problems were now solved. This claim completely misrepresents the actual thinking of the evolutionists. As Dobzhansky, I, and several other architects of the Synthesis continued to emphasize, the Synthesis left many problems open. To be sure, it had reaffirmed the basic Darwinian principles, but the publications in evolutionary biology from the 1940s to the present show how many doubts had remained. As I had emphasized (Mayr 1967, p. 13) 'in spite of the almost universal acceptance of the synthetic theory of evolution, we are still far from fully understanding almost any of the more specific problems of evolution,' and I discussed this under such headings as natural selection, gene and chromosome, the phenotype of the individual, population, closed and open populations, the species, peripheral populations, the higher categories, and different pathways toward evolutionary success. The accusation by critics that the architects of the Synthesis had claimed that they provided a complete answer to all evolutionary problems is clearly contradicted by the literature.

Furthermore, from the very beginning, there were rather fundamental differences among the architects of the Synthesis with respect to certain evolutionary problems. For instance, I differed from Simpson in my emphasis on populations, on the geographical component of evolution, and on the species definition, and from Dobzhansky in the explanation of the origin of isolating mechanisms, and, on the whole, in a more selectionist approach. Similar differences existed among other leading evolutionists. Yet they all agreed on the basic Darwinian theory, and in particular on quite a lengthy list of assumptions which I have recently summarized (Mayr 1988a).

Futuyma (1988) calls attention to one failing of the Synthesis, that is, a failure to develop a unified field of evolutionary biology. He regrets the current narrow specialization of the journal *Evolution* to what he calls synchronic evolution, and in particular to quantitative evolutionary ecology. I sympathize with his complaint. When I started *Evolution* in 1947 I would have had little difficulty filling the journal with articles on *Drosophila*; but I had a conspicuous dearth of manuscripts dealing with paleontology, systematics, behavior, plants, etc. I had ordered 500 letterheads and used up the entire supply in the first 7 months of my editorship by writing to every conceivable person in the whole world, suggesting topics perhaps suitable for a paper to be published in *Evolution*. Curiously, the current seeming one-sidedness of *Evolution* is in some respects actually a result of the enormous success of evolutionary thinking. In all sorts of fields there are now journals specializing in the evolutionary aspects of that field. This is true for *Paleobiology* in evolutionary paleontology; for two journals in molecular evolutionary biology; for *Systematic Zoology*, and several others. Another reason, however, for the particular current assortment of papers is the contemporary infatuation with mathematical models, which are of course much more easily applied to synchronic than to historical analyses.

Looking back over the 45 years of the existence of *Evolution*, one can see a sequence of fashions, each sooner or later to be displaced by a new fashion. This phenomenon will surely continue. One can speculate what the ensuing fashions will be. The discovery of the great heterogeneity of genes by molecular biology has prepared the ground for an analysis of the evolutionary role of the various kinds of genes. Kimura has amply demonstrated how greatly they may differ from each other in this respect.

With the ever better understanding of the genome, it will soon be possible to determine what part of it plays the major role in speciation events. Our present interpretation of evolution is based on a remarkably small fraction of the enormous multitude of organisms. How greatly will we have to widen the pluralism of our theories, when we better understand evolution in protists, sponges, host-specific aphids, marine algae, lichens, fungi, and other organisms, so far rather neglected by evolutionists?

What is perhaps most remarkable is the sturdiness of the basic Darwinian paradigm. Attempts to refute it have been almost countless over the last 130 years and yet they all have been unsuccessful. However, conceptually, Darwinism is fundamentally so different from all previously proposed theories and ideologies that anyone not actually working in evolutionary biology seems to have a hard time understanding it. Perhaps this is, in part, the fault of the evolutionists themselves for fighting so much over rather trivial differences of opinion instead of re-emphasizing the basic contents and achievements of Darwinism.

REFERENCES

- Aird, L., Bentall, H. H. and Fraser Roberts, J. A. (1953). Relationship between cancer of stomach and the ABO groups. *Brit. Med. J.* **1**, 799–801.
- Antonovics, J. (1987). The evolutionary dys-synthesis: which bottles for which wine? *Amer. Nat.* **129**, 321–31.
- Barton, N. H. and Hewitt, G. M. (1985). Analysis of hybrid zones. *Ann. Rev. Ecol. Syst.* **16**, 113–48.
- Bryant, E. H., McCommas, S. A., et al. (1986). The effect of an experimental bottleneck upon quantitative genetic variation in the housefly. *Genetics* **114**, 1191.
- Buchwalter, J. A., et al. (1956). Natural selection associated with the ABO blood group. *Science* **123**, 840–1.
- Butlin, R. K. (1985). Speciation by reinforcement. In *Orthoptera*, Vol. 1 (ed. J. Gosálvez et al.), pp. 84–113. Fundacion Ramon Areces.
- Buzzati-Traverso, A. A. (ed.) (1954) Symposium on Genetics of Population Structure. *Int. Union Biol. Sci. Serie B*, No. 15. UIBS, Naples.
- Carson, H. L. (1975). The genetics of speciation at the diploid level. *Amer. Nat.* **109**, 83–92.
- Charlesworth, B., Lande, R. and Slatkin, M. (1982). A neo-Darwinian commentary on macroevolution. *Evolution* **36**, 474–98.
- Cracraft, J. (1984). The terminology of allopatric speciation. *Syst. Zool.* **33**, 115–6.
- Dobzhansky, Th. (1933). Geographical variation in lady-beetles. *Amer. Nat.* **67**, 97–126.
- (1937). *Genetics and the origin of species*. (2nd edn, 1941.) Columbia University Press, New York.
- Dominey, W. J. (1984). Effects of sexual selection and life history on speciation. In *Evolution in fish species flocks* (ed. A. A. Echelle and I. Kornfield), pp. 231–50. University of Maine Press, Orono, ME.
- Eldredge, N. (1971). The allopatric model and phylogeny in Paleozoic invertebrates. *Evolution* **25**, 156–67.
- (1985). *Unfinished synthesis*. Oxford University Press, New York.
- (1989). *Macroevolutionary dynamics: Species, niches, and adaptive peaks*. McGraw-Hill, New York.
- and Gould, S. J. (1972). Punctuated equilibria: an alternative to phyletic

- gradualism. In *Models in paleobiology* (ed. T. J. M. Schopf), pp. 305–32. Freeman, Cooper, San Francisco.
- Feder, J. L. (1988). Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. Letter to *Nature* **336**, 61.
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon Press, Oxford.
- Ford, E. B. (1940). Polymorphism and taxonomy. In *The new systematics* (ed. J. S. Huxley), pp. 493–513. Clarendon Press, Oxford.
- Futuyma, D. J. (1988). *Sturm und Drang* and the evolutionary synthesis. *Evolution* **42**, 217–26.
- Goldschmidt, R. B. (1940). *The material basis of evolution*. Yale University Press, New Haven, CT.
- Goodnight, C. J. (1987). On the effect of founder events on epistatic genetic variance. *Evolution* **41**, 80–91.
- Gould, S. J. (1983). The hardening of the modern synthesis. In *Dimensions of Darwinism* (ed. M. Grene), pp. 71–93. Cambridge University Press, Cambridge.
- Grene, M. (1990). Is evolution at a crossroads? *Evol. Biol.* **24**, 51–80.
- Griesemer, J. R. (1990). Modeling in the museum: on the role of remnant models in the work of Joseph Grinnell. *Bio. and Phil.* **5**, 3–36.
- Grinnell, J. (1943). *Joseph Grinnell's philosophy of nature*. University of California Press, Berkeley and Los Angeles.
- Gutmann, W. F. and Bonik, K. (1981). *Eine Kritische Evolutionstheorie. Ein Beitrag zur Überwindung altdarwinistischer Dogmen*. Gerstenberg, Hildesheim.
- Haffer, J. (1992). Artbegriff und Artbegrenzung im Werk des Ornithologen Erwin Stresemann (1889–1972). *Mitt. Zool. Mus. Berlin* **67** (suppl.), in press.
- Haldane, J. B. S. (1964). A defense of beanbag genetics. *Perspect. Biol. Med.* **VII(3)** 343–59.
- Harris, H. (1966). Enzyme polymorphism in man. *Proc. R. Soc. B* **164**, 298–316.
- Hennig, W. (1950). *Grundzüge einer Theorie der phylogenetischen Systematik*. Deutscher Zentralverlag, Berlin.
- Ho, M. W. and Fox, S. W. (eds) (1988). *Evolutionary processes and metaphors*. Wiley, New York.
- Ho, M. W. and Saunders, P. T. (eds) (1984). *Beyond neo-Darwinism: an introduction to the new evolutionary paradigm*. Academic Press, London.
- Hoffman, A. (1989). *Arguments on evolution: a paleontologist's perspective*. Oxford University Press, Oxford and New York.
- Hubby, J. L. and Lewontin, R. C. (1966). The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* **54**, 577–94.
- Huxley, A. (1982). Address of the President. *Proc. Roy. Soc. London A* **379**, IX–XVII.
- Huxley, J. S. (ed.) (1940). *The new systematics*. Clarendon Press, Oxford.
- Kaneshiro, K. Y. (1980). Sexual isolation, speciation and the direction of evolution. *Evolution* **34**, 437–44.
- Key, K. (1974). Speciation in the Australian Morabine grasshoppers in taxonomy and ecology. In *Genetic mechanism of speciation in insects* (ed. M. J. D. White), pp. 43–56. Australia and New Zealand Book Co., Sydney.
- (1981). Species, parapatry and the morabine grasshoppers. *Syst. Zool.* **30**, 425–58.

32 Ernst Mayr

- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* **217**, 624–6.
- (1983). *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.
- King, J. L. and Jukes, T. H. (1969). Non-Darwinian evolution. *Science* **164**, 788–98.
- Kottler, M. (1978). Charles Darwin's biological species concept and theory of geographic speciation. *Amer. Sci.* **35**, 275–97.
- Lack, D. (1947). *Darwin's finches*. Cambridge University Press, Cambridge.
- Lerner, I. M. (1954). *Genetic homeostasis*. Oliver and Boyd, Edinburgh.
- Levinton, J. (1988). *Genetics, paleontology, and macroevolution*. Cambridge University Press, Cambridge.
- Lewis, H. (1962). Catastrophic selection as a factor in speciation. *Evolution* **16**, 257–71.
- Lewontin, R. C. (1970). The units of selection. *Ann. Rev. Ecol. Syst.* **1**, 1–14.
- (1983). The organism as the subject and object of evolution. *Scientia* **118**, 65–82.
- Mayr, E. (1940). Speciation phenomena in birds. *Amer. Nat.* **74**, 249–78.
- (1942). *Systematics and the origin of species*. Columbia University Press, New York.
- (1944). Symposium on age of the distribution pattern of the gene arrangements in *Drosophila pseudoobscura*. Introduction and some evidence in favor of a recent date. *Lloydia* **8**(2), 69–82.
- (1947). Ecological factors in speciation. *Evolution* **1**(4), 263–88.
- (1948). The bearing of the new systematics on genetical problems. The nature of species. *Adv. Genet.* **2**, 205–37.
- (1950). The role of the antennae in the mating behavior of female *Drosophila*. *Evolution* **4**(2), 149–54.
- (1952). Comments on evolutionary literature. *Evolution* **6**(2), 248–52.
- (1954). Change of genetic environment and evolution. In *Evolution as a process* (ed. J. Huxley, A. C. Hardy and E. B. Ford), pp. 157–80. Allen and Unwin, London.
- (1955). Karl Jordan's contribution to current concepts in systematics and evolution. *Trans. Royal Entomol. Soc. London* **107**, 45–66.
- (1956). Integration of genotypes: synthesis. *Cold Spring Harbor Symp. Quant. Biol.* **20** (1955), 327–33.
- (1959a). Isolation as an evolutionary factor. *Proc. Amer. Phil. Soc.* **103**(2), 221–30.
- (1959b). Where are we? *Cold Spring Harbor Symp. Quant. Biol.* **24**, 1–14.
- (1960). The emergence of evolutionary novelties. In *The evolution of life* (ed. S. Tax), pp. 349–80. University of Chicago Press, Chicago.
- (1963). *Animal species and evolution*. Harvard University Press, Cambridge, MA.
- (1969a). *Principles of systematic zoology*, 1st edn. McGraw-Hill, New York.
- (1969b). Bird speciation in the tropics. *Biol. J. Linn. Soc.* **1**, 1–17.
- (1970). *Populations, species and evolution*. Harvard University Press, Cambridge, MA.
- (1972). Alden Holmes Miller. *Biographical Mem., Nat. Acad. Sci.* **43**, 177–214.

- (1975). The unity of the genotype. *Biol. Zentralbl.* **94**, 377–88.
- (1982a). Processes of speciation in animals. In *Mechanisms of speciation* (ed. C. Barigozzi), pp. 1–19. Alan R. Liss, New York.
- (1982b). Speciation and macroevolution. *Evolution* **36**(6), 1119–32.
- (1982c). *The growth of biological thought*. Harvard University Press, Cambridge, MA.
- (1984a). Evolution of fish species flocks: a commentary, introduction. In *Evolution of fish species flocks* (ed. A. A. Echelle and I. Kornfield), pp. 3–11. University of Maine at Orono Press, Orono, ME.
- (1984b). The triumph of evolutionary synthesis. *Times Literary Supplement* (2 November 1984), 1261–2.
- (1985). Natürliche Auslese. *Naturwissenschaften* **72**, 231–6.
- (1986). Review of Elliott Sober, 1984, *The nature of selection: evolutionary theory in philosophical focus*. Cambridge: MIT Press. *Paleobiology* **12**(2), 233–9.
- (1987). The species as category, taxon, and population. In *Histoire du Concept d'espèce dans les Sciences de la Vie*, Ch. 15, pp. 303–20. Fondation Singer-Polignac, Paris.
- (1988a). *Toward a new philosophy of biology*. Harvard University Press, Cambridge, MA.
- (1988b). The why and how of species. *Biology and Philosophy* **3**, 431–41.
- (1990a). Myxoma and group selection. *Biol. Zentralblatt* **109**, 453–7.
- (1990b). Die drei Schulen der Systematik (the three schools of systematics). *Verh. Dtsch. Zool. Ges.* **83**, 263–76.
- (1991). *Principles of systematic zoology*, 2nd edn. McGraw-Hill, New York.
- (1992). A local flora and the biological species concept. *Amer. J. Bot.* **79**(2), 222–38.
- and Diamond, J. (1992). *The birds of Northern Melanesia*. (in MS).
- and Diamond, L. (1959). Comment: (blood-group correlations). *J. Med. Education* **34**, 421.
- and Provine, W. (eds) (1980). *The evolutionary synthesis*. Harvard University Press, Cambridge, MA.
- Diamond, L. K., Levine, R. P. and Mayr, M. (1956). Suspected correlation between blood-group frequency and pituitary adenoma. *Science* **124**, 932–4.
- Miller, A. H. (1941). Speciation in the Avian genus *Junco*. *Univ. Calif. Publ. Zoology* **44**, 173–434.
- Muller, H. J. (1940). Bearings of the *Drosophila* work on systematics. *The new systematics* (ed. J. S. Huxley), pp. 185–268. Clarendon Press, Oxford.
- Newman, C. M., Cohen, J. E. and Kipnis, C. (1985). Neo-Darwinian evolution implies punctuated equilibria. *Nature* **315**, 400–1.
- Osgood, W. H. (1909). Revision of the mice of the American genus *Peromyscus*. *North Amer. Fauna (Bur. Biol. Survey)* **28**, 1–285.
- Otte, D. and Endler, J. E. (1989). *Speciation and its consequences*. Sinauer, Sunderland, MA. (Several papers on hybrid zones, pp. 85–207.)
- Phillips, J. C. (1915). Experimental studies of hybridization among ducks and pheasants. *J. Exptl. Zool.* **18**, 69–143.
- Provine, W. B. (1979). Francis B. Sumner and the evolutionary synthesis. *Stud. Hist. Biol.* **3**, 211–40.

- (1988). Progress in evolution and meaning of life. In *Evolutionary Progress* (ed. M. H. Nitecki), pp. 53–70. Chicago University Press, Chicago.
- Rensch, B. (1929). *Das Prinzip geographischer Rassenkreise und das Problem der Artbildung*. Borntraeger, Berlin.
- Rhodes, F. H. T. (1983). Gradualism, punctuated equilibrium and the *Origin of Species*. *Nature* **305**, 269–72.
- Riedl, R. (1975). *Die Ordnung der Lebendigen*. Paul Parey, Hamburg and Berlin.
- Robson, G. C. and Richards, O. W. (1936). *The variations of animals in nature*. Longmans, Green, London.
- Schindewolf, O. (1936). *Paläontologie, Entwicklungslehre und Genetik*. Bornträger, Berlin.
- (1950). *Grundfragen der Paläontologie*. Schweizerbart, Stuttgart.
- Simpson, G. G. (1944). *Tempo and mode in evolution*. Columbia University Press, New York.
- (1953). *The major features of evolution*. Columbia University Press, New York.
- Stebbins, G. L. (1950). *Variation and evolution in plants*. Columbia University Press, New York.
- Sulloway, F. J. (1979). Geographic isolation in Darwin's thinking: the vicissitudes of a crucial idea. *Stud. Hist. Biol.* **3**, 23–65.
- Thornhill, R. and Alcock, J. (1983). *The evolution of insect mating systems*. Harvard University Press, Cambridge, MA.
- Volontsov, N. N. and Lyapunova, E. A. (1989). Two ways of speciation. In *Evolutionary biology of transient unstable populations* (ed. A. Fondevila), pp. 221–45. Springer, Berlin/Heidelberg.
- Waddington, C. H. (ed.) (1968–1973). *Toward a theoretical biology I–4*. Edinburgh University Press, Edinburgh.
- West-Eberhard, M. J. (1983). Sexual selection, social competition, and speciation. *Quart. Rev. Biol.* **58**, 155–83.
- White, M. J. D. (1978). *Modes of speciation*. W. H. Freeman, San Francisco.
- Willis, J. C. (1922). *Age and area*. Cambridge University Press, Cambridge.
- (1940). *The course of evolution by differentiation*. Cambridge University Press, Cambridge.
- Wright, S. (1940a). The statistical consequences of Mendelian heredity in relation to speciation. *The new systematics* (ed. J. S. Huxley), pp. 161–83. Clarendon Press, Oxford.
- (1940b). Breeding structure in relation to speciation. *Amer. Nat.* **74**, 232–48.
- (1943). Isolation by distance. *Genetics* **28**, 114–38.

Sewall Wright: gene interaction and the Shifting Balance Theory

MICHAEL J. WADE

'Context and interaction are of the essence' (Lewontin 1974: 318.)

1. INTRODUCTION

1.1 The question of epistasis in evolution

The role of gene interactions in adaptive evolution is not well understood despite the frequent use of terms such as 'co-adapted gene complexes' and 'genetic revolutions' in evolutionary discussion. Although most evolutionary biologists and geneticists would acknowledge Sewall Wright's principle of the 'universality of gene interactions,' there is no consensus regarding their role in evolution. Gene interactions and their statistical characterization as epistasis (the between-locus non-additive component of genetic variance) are not central concepts in current adaptation theory. The question might be put as follows: Does natural selection assemble gene combinations one gene at a time by operating on the average 'additive' effects of single genes or does natural selection operate to choose directly among different gene combinations or 'systems of interacting genes?' The Fisherian answer is that selection operates primarily on the average additive effects of single genes but, where loci are very tightly linked (e.g. inversions or supergenes), there can be direct selection among combinations of linked loci. Wright would answer in a different way – namely, that the combination of *local* mass selection (operating on the 'additive' effects of single genes) and random genetic drift create different 'adaptive peaks' (each representing a unique gene combination) in different demes. Interdemic selection among the adaptive peaks occurs by differential dispersion. In this way interdemic selection operates to choose directly among different gene interaction systems or gene combinations. Fisher's answer is more widely known and accepted whereas Wright's extension has been largely ignored.

1.2 Definitions of epistasis

The term ‘epistasis’ is often used synonymously for the phrase ‘gene interaction’ but it has two very different meanings in genetics. In molecular and biochemical genetics, genes whose products function sequentially as substrates or catalysts in a common biochemical pathway are considered to be ‘epistatic’ to one another. Biochemical epistasis occurs when the ‘expression of one gene wipes out the phenotypic effects of another gene’ (Lewin 1990, p. 809). In this usage, one gene is ‘epistatic’ to another if the function of its gene product in a biochemical pathway is *conditional* on the success or failure of the other gene operating at a later step in the same pathway. This qualitative use of the term epistasis to indicate biochemical contingency corresponds to what might be called the epistatic effect, a linear measure of gene interaction. It is different from the meaning given to ‘epistasis’ by quantitative or statistical geneticists in reference to a component of genetic variance (a quadratic as opposed to linear measure). In statistical and quantitative genetics, epistasis is a populational concept describing the relationship between phenotypic variation and genetic variation (Fisher 1918; Cockerham 1954). Epistasis, along with dominance, is the ‘non-additive’ component of the genetic variance for a trait. This component of genetic variation measures the statistical effects of variations in gene combinations between individuals in relation to the total phenotypic variance, between individuals in a population. Epistasis between loci in statistical genetics requires genetic variation at each of at least two loci. Without genetic variation at two loci, the differences between individuals within a population cannot be attributed (in the statistical sense) to variations between individuals in gene combinations.

Genes can be epistatic to one another in the biochemical sense but not contribute to epistasis in the statistical sense. For example, using material from isogenic stocks of the fruit fly, *Drosophila melanogaster*, one can study the accumulation reaction products and thus characterize biochemical epistasis. However, there is no genetic variation within an isogenic stock by definition; all individuals share the same genetic constitution. One cannot partition components of genetic variance so there can be no statistical epistasis. Wright’s principle of ‘universality’ of epistasis at the biochemical level (Wright 1969, pp. 59–60) does not guarantee an important or even a significant role for epistasis in evolution at the populational level. It is the statistical effects of gene interactions on the phenotypic variation that determine how gene combinations will evolve and whether or not selection can operate directly on differences between individuals in gene combinations.

1.3 Epistasis and Wright's Shifting Balance Theory

Wright (1931) proposed his Shifting Balance Theory of evolution as a mechanism for selection to operate directly on the vast field of gene combinations. He postulated a three-phase process with all three phases acting simultaneously.

Phase 1: Random Genetic Drift where the gene frequencies at many loci in a finite subpopulation or deme drift at random about the set of gene frequencies corresponding to a local adaptive peak.

Phase 2: Mass Selection Within Demes which moves a local deme from one fitness peak to another whenever random genetic drift affects local gene frequencies sufficiently for the deme to cross over into the domain of attraction of another adaptive peak.

Phase 3: Interdemic Selection in which a deme at a high fitness peak systematically shifts the position of equilibrium of other demes to its own set of equilibrium gene frequencies by the differential dispersion of migrants away from the deme at the high peak and into neighboring demes at lower fitness peaks.

In Wright's theory, random genetic drift and local individual selection in combination can achieve results unattainable by selection alone. This occurs as a result of different gene combinations arising between local demes in subdivided populations. The differences between demes in *adaptive* gene combinations would be associated with between-deme differences in average fitness. Interdemic selection, occurring by the differential dispersion of migrants out from demes of high local fitness and into demes of lower average fitness, is the process by which a favorable gene combination arising in a single locality could be introduced to other demes and thereby spread throughout a species. Individual selection operating solely in large panmictic populations was not viewed by Wright as a feasible method for selecting among gene combinations owing to recombination. Recombination breaks up gene combinations that are favored by individual selection and in this way it opposes within-deme selection. That is, gene combinations cannot be efficiently transmitted from parents to offspring because, owing to recombination, parents pass on genes and not genotypes (gene combinations) to their offspring. However, Wright proposed that locally fixed combinations causing a local adaptive peak in fitness could be exported to neighboring demes by migration over extended periods of time.

1.4 The Fisherian viewpoint

Although Wright's major premise of the ubiquity of gene interactions and the complexity of the relationship between genes and character is shared by most evolutionary biologists, the importance of his Shifting

Balance Theory for adaptive evolution and selection among gene combinations is not similarly accepted. The alternative view, attributable in large part to R. A. Fisher (1918, 1958), is that the ultimate fate of a gene is determined by its average effect with respect to fitness within large populations and that co-dependent gene complexes arise largely as a by-product of the effects of selection on single genes. The average effect of a gene, G , on fitness is defined as the ‘regression in the actual population of the genotypic measurement [fitness] on the number of G genes’ (Fisher 1941, p. 54). When there are gene interactions and small local breeding groups, the average effect of a gene on fitness can vary from group to group (see below). For this reason, Fisher insisted that:

. . . the population used to determine its value comprises, not merely the whole of a species in any one generation attaining maturity, but is conceived to contain all the genetic combinations possible, with frequencies appropriate to their actual probabilities of occurrence and survival, whatever these may be, and if the average is based upon the statures attained by all these genotypes in all possible environmental circumstances, with frequencies appropriate to the actual probabilities of encountering these circumstances (Fisher 1958, pp. 30–31).

The most important difference between Wright and Fisher in their views on evolution is the degree to which natural populations deviate from this ideal and the evolutionary consequences for single genes of such deviations. This may be the source of the ‘intense controversy’ (Provine 1986, p. 232) between Fisher and Wright over gene interactions and effective population size. Unless effective sizes are small and unless there are gene interactions, natural populations cannot significantly deviate from the Fisherian ideal.

For Wright, population subdivision and gene interactions opened the important possibility of significant variation between local demes in the average effect of a gene on fitness. In the absence of population subdivision, this is not a likely possibility and Fisher’s concept of population size is important to his views on the evolutionary process: ‘. . . I believe that N must usually be the total population on the planet . . .’ (correspondence between Fisher and Wright, quoted in Provine 1986, p. 255). Natural populations in Fisher’s view approach the ideal population fairly closely and he expects the evolutionary fate of a gene to be determined by its species-wide average effect on fitness. With this view, it is very unlikely that random genetic drift could lead to the fixation of a gene throughout the whole of a species, except with extremely weak selection. If a gene’s average effect on fitness is positive, then natural selection will sweep that gene to fixation, and conversely if its average effect is negative. Gene interactions are of little consequence in this picture because in the idealized population, a gene is ‘tried’ by

natural selection in all possible genotypic configurations and in all relevant environments exactly in the expected frequencies of each genotype, each environment, and each gene-environment combination. In contrast, Wright focused not on the N of the whole species but on the *local* effective population number.

More recently, Williams (1966, p. 56) has expressed considerations similar to those of Fisher:

Obviously it is unrealistic to believe that a gene actually exists in its own world with no complications other than abstract selection coefficients and mutation rates. The unity of the genotype and the functional subordination of the individual genes to each other and to their surroundings would seem, at first sight, to invalidate the one-locus model of selection. Actually these considerations do not bear on the basic postulates of the theory. No matter how functionally dependent a gene may be, and no matter how complicated its interactions with other genes and environmental factors, it must always be true that a given gene substitution will have an arithmetic mean effect on fitness in any population. One allele can always be regarded as having a certain selection coefficient relative to another at the same locus at any given point in time. Such coefficients are numbers that can be treated algebraically, and conclusions inferred from one locus can be iterated over all loci. Adaptation can thus be attributed to the effect of selection acting independently at each locus.

Genetically complex adaptations are gradually built up by natural selection operating independently on single genes. The variations in the genetic background or environment that characterize real finite populations are of little long-term consequence because they do not significantly affect the relationship of a gene to fitness.

The definitions of gene effects and gene interactions are equivalent to the statistical concepts of main effects and interactions (Anderson and Kempthorne 1954). Fisher (1918, 1925) introduced both concepts. Within statistics, the limits to the utility of these concepts were challenged by Neyman (1935) (see Traxler 1976) and within evolutionary genetics by Wright (1930, 1931). Both Neyman and Wright used remarkably similar arguments, as I show below.

2. MAIN EFFECTS AND INTERACTIONS IN STATISTICS

2.1 Definitions of main effects and interactions

In biological systems, it is often the case that several different factors are important in producing a particular condition. For example, seed yield of a plot of land might depend on several factors such as (A) the variety of plant sown, (B) the amount of fertilizer used, and (C) the rainfall

Table 1
Simple effects, main effects, simple interactions, and interactions

Simple effects of A:	1. $(1, 0, 0) - (0, 0, 0) = \delta_1$ 2. $(1, 1, 0) - (0, 1, 0) = \delta_2$ 3. $(1, 0, 1) - (0, 0, 1) = \delta_3$ 4. $(1, 1, 1) - (0, 1, 1) = \delta_4$
Main effect of A:	$a = \{\sum \delta_k\} = \{\delta_1 + \delta_2 + \delta_3 + \delta_4\}/4$
Simple interactions of A and B:	1. $\{(1, 1, 0) - (1, 0, 0) - (0, 1, 0) + (0, 0, 0)\}/2 = \{\delta_2 - \delta_1\}/2 = \beta_1$ 2. $\{(1, 1, 1) - (1, 0, 1) - (0, 1, 1) + (0, 0, 1)\}/2 = \{\delta_4 - \delta_3\}/2 = \beta_2$
Interaction of A and B:	$ab = [\sum \beta_k]/2 = \{\delta_2 - \delta_1 + \delta_4 - \delta_3\}/4$

during the growing season. Furthermore, the influence of a particular factor on seed yield might depend on the level of one or more of the other factors. For example, some plant varieties (levels of factor A) might be more sensitive to low levels of rainfall (factor C) than other varieties. There are many reasons, biological as well as economic, that make it desirable to understand the individual contribution of each factor to seed yield as well as the joint effects of the factor combinations. The analysis of variance in conjunction with replicated factorial experimental designs was introduced in order to achieve this understanding of individual and joint effects (Fisher 1937).

The individual contribution of a factor to a condition averaged over all contexts, i.e. over all combinations of the other factors, is defined as the *main effect* of that factor (Neyman 1935; Yates 1935; Fisher 1937). In Tables 1 and 2, Neyman's model is presented. There are three factors, A, B, and C, with two levels of each factor, 1 (present) and 0 (absent). A fully factorial design would consist of eight treatments, each of which could be represented as a set of 0's and 1's. For example, (1, 0, 1) represents the treatment with factors A and C present but B absent and

Table 2
The Neyman model of nature

Factors	A	B	C	AB	AC	BC	ABC	—
Code	100	010	001	110	101	011	111	000
Yield	98	91	91	109	109	91	109	102

(0, 0, 0) represents the mean of the control treatment with no factors present. The main effect of factor A is defined as the average of four treatment mean differences (cf. Table 1), which Neyman (1935) referred to as *simple effects*. A simple effect is the difference observed between two treatment means when the treatments differ by only the presence and absence of the factor in question. The simple effects are summed together and then averaged to calculate the *main effect* of a factor. In this example, four simple effects contribute to the average effect of factor A (cf. Table 1). Each simple effect contributing to the average can be considered a *context-specific* contribution of a factor to a main effect, where 'context' refers to a constant combination of all other factors.

The first order *interaction effect* is defined similarly as an average of 'simple' interactions. It is the contribution of a *pair* of factors to the condition under study averaged over all contexts in which the pair of factors appears. A 'simple' interaction is the observed difference in a factor's simple effect when measured in the presence or absence of some other factor. A simple interaction is a linear combination of four treatment means; it is plus or minus one-half multiplied by (a) the mean of the treatment with both factors, (b) the mean of that treatment with the first but not the second factor, (c) the mean of that treatment with the second but not the first factor, and (d) the treatment mean with neither factor (cf. Table 1). In Neyman's example, two such simple interactions are averaged to calculate the interaction effect of the AB pair of factors. The number of terms averaged to calculate a main effect can be greater than the number of terms averaged to calculate an interaction effect (four versus two, respectively, in the present example).

In detecting a main effect of interaction, we need to calculate its standard error (σ). The variance of simple effects about the average effect for factor A is

$$\sigma_{\delta}^2 = \sigma_{\beta}^2 + \frac{1}{8} \{(\delta_1 - \delta_4)^2 + (\delta_2 - \delta_3)^2\} \quad (1)$$

The variance of simple interactions about the interaction effect of A and B is similarly given by

$$\sigma_{\beta}^2 = \left(\frac{1}{4}\right) (\beta_1 - \beta_2)^2 \quad (2)$$

Despite having the same number of treatment means in the definitions of the main effect of A and its interaction with B (*a* and *ab* in Table 1), the standard error of a main effect can be *smaller* than the standard error of the interaction, making it more difficult to detect interaction effects than main effects. In Neyman's example (Table 2), the main effect of A, *a*, is 12.5 with a standard error of 3.19 but the interaction *ab* is equal to 5.5 with a standard error of 3.89 (Neyman 1935). It is this property of the definitions of main effects and interactions that can complicate the

interpretation of results in replicated, factorial experiments, as Neyman (1935) pointed out (Traxler 1976; Milliken and Johnson 1984).

2.2 The experiment as a questionnaire for Nature

Fisher considered the replicated, factorial design and analysis of variance (ANOVA) as a method for investigating and characterizing nature. The metaphor that he used to describe the statistical procedure was that of an experimenter submitting a questionnaire to Nature in the form of an experimental design (Fisher 1937). By interpreting Nature's answers to the experimental questionnaire using ANOVA, the experimenter may learn about the characteristics of the natural world. When Nature is complex and many factors affect the condition under study, Fisher argued, the experimental study of one factor at a time is an inefficient method of investigating the causal relationships between factors and condition. He introduced the factorial design and the analysis of variance in part to address this inefficiency (Fisher 1937).

Factorial designs are more efficient than the sequential study of single factors in isolation for three reasons. First, we do not know *a priori* whether or not a particular factor produces its effect independent of all other possibly influential factors. The study of one factor at a time in isolation from other factors cannot enlighten us on this point. Secondly, in most biological systems, we *cannot* study one factor at a time. Instead, we study the effects of variations in one factor holding all other factors at a single, constant level or over a random sample of conditions unknown or uncontrollable with respect to the other factors. When factors interact, then single-factor effects estimated in this way can be confounded with the interaction effects *unique* to the special context defined by controlling the other factors constant. When we study a single factor and keep other possibly important factors constant, we are estimating that factor's main effect by measuring only one of its simple effects. When we measure a factor's performance in a single context, it may or may not be representative of the array of relevant biological contexts. Thirdly, if the magnitude of one factor's effect is not independent of all other factors, the nature of the interdependence of factor effects may not be a simple one. The methods proposed by Fisher in the ANOVA are designed to accommodate arbitrarily complex interactions among factor effects.

The primary advantage of factorial experimental designs according to Fisher (1937) was that they permit the *consistency* of the effects of one factor to be explored across a field of possible factor combinations. Indeed, this is his motivation for defining the 'main effect' of a factor as the average of the simple effects (see above). For applied purposes, knowledge of this consistency is of considerable interest because an estimate of a factor's main effect from a factorial design is expected to

have predictive value across a wider range of conditions than if only a single simple effect were measured. For the study of biological systems in nature, however, the *lack of consistency*, i.e. the variability of simple effects about the expected main effect, and the causes of such variability are also of concern. Fisher in his remarks on the additive effects of genes chose to emphasize the *consistency* that an estimate of 'genic main effects' offers for evolutionary prediction much in the same way that he argued for consistency of effects of any factor in a factorial design. In contrast, Wright emphasized the *variability* of a gene's simple effects when measured in different genetic backgrounds in his several discussions of gene interactions. The estimates of a gene's effect on fitness can vary not only in magnitude from deme to deme but also vary in sign. Wright emphasized the biological importance of the variation in the components of genic main effects, owing to gene interactions.

2.3 The problem that interactions cause for interpreting the questionnaire

Whenever there are interactions among factors, the detection and estimation of the main effects of single factors and the interaction effects of groups of factors becomes problematic (Milliken and Johnson 1984). The problem becomes more acute as the size of the experiment, i.e. the number of replicates per treatment, decreases (Neyman 1935). Because the standard error of an estimate of an interaction effect can be greater than the standard error of a main effect estimate, it can be more difficult to bound the estimate of an interaction away from zero than it is to bound the estimate of a main effect away from zero when the two kinds of effects are of equivalent magnitude. There can be a bias in the interpretation of Nature's answers to the experimental questionnaire that favors the discovery of main effects and the overlooking of interactions. One is more likely to make a Type II error (i.e. accept the null hypothesis of no effect when an effect is present) in the case of interactions than one is in the case of main effects because the power of the test for interaction can be weaker. For this reason, we are more likely to discover main effects than we are likely to discover interactions when we look for both using the ANOVA.

2.4 Neyman and Fisher

This problem was first brought to light by Neyman (1935). In reply to a paper by Yates (1935) on complex experiments and their interpretation, Neyman (1935) questioned the validity of inferences based on main effects in an ANOVA, especially when the numbers of replicates were small. Neyman's challenge was founded on the basic definitions of main effects

and interactions, their attendant measurement errors, and the problems of interpretation that can arise as a result of these definitions.

In order to illustrate the problems that could arise in the interpretation of an experiment by ANOVA, Neyman (1935) constructed a hypothetical model of nature (Table 2) and conducted 30 factorial 'experiments' using Monte Carlo simulation on this model of nature. That is, he submitted not one but 30 questionnaires to this model and used the recommended ANOVA methods to analyze nature's answers. He then compared the characterization of nature *inferred* from the ANOVA to the *known* underlying model.

He hypothesized a situation in which a factorial experiment was being employed to determine which one or combination of three fertilizers best increased crop yield. One of three factors, A, produced a true negative effect on yield of 4 per cent when added to the control (i.e. the simple effect of A alone was to decrease yield by 4 per cent). However, the simple interactions, ab , ac , and abc , involving A had a positive 7 per cent increase on yield. (The other main effects, b and c , and their simple interaction, bc , were zero.) Using this model of nature, he conducted Monte Carlo simulations of 30 factorial experiments, each with three replications per treatment, exactly analogous to the 'complex experiment' discussed by Yates (1935). Assuming a fixed rate of random errors of treatment means, he applied the ANOVA to the data from each of the 30 experiments.

In 28 of the 30 'experiments,' a significant effect was found for some factor or interaction of factors at the 0.01 level. In 27 of the 28 experiments with significant results (96.4 per cent of the time), a significant main effect of A was discovered, and in 61 per cent of the questionnaires it was the *only* significant effect. On the other hand, a significant interaction of A and B was discovered in five of the 28 cases (17.9 per cent of the time). These results illustrate the bias toward discovering main effects and overlooking interaction effects described above.

This was not the only problem with nature's answers to the questionnaire of the ANOVA, however. *The main effect of A was estimated to be positive in all of the 28 cases where it was detected when in fact the simple effect of A alone was negative.* The inference regarding the sign of the main effect of A was the opposite of its simple effect when B and C were missing.

Neyman concluded that '... the frequency of cases in which practical conclusions concerning the properties of treatment a would be entirely false is very considerable and is certainly much greater than the level of significance 0.01 ...' (Neyman 1935, p. 241). The cause of the problem was two-fold: (1) the effect that the simple interactions had when averaged to calculate the main effect of A according to definition; and, (2) the inability of the experiment to detect interactions when they were present

owing to the larger standard error of estimates of the interaction effect in 'small' experiments (i.e. experiments with few replications). There are ways to address these problems as outlined in Milliken and Johnson (1984).

An important remaining question is how representative of nature Neyman's example might be. The hypothetical model of Nature explored by Neyman's simulation of small experiments is 'frivolous,' in that the simple interactions of A with B and C are all of *opposite sign* to the simple effect of A alone. Factor A alone is bad for yield although in combinations with other factors it is good for yield. When these several simple effects are averaged, the sign of the main effect is the opposite of the sign of the simple effect. Indeed, the only way for the sign of a factor's main effect to be the opposite of that factor's simple effect is for the simple interactions involving one factor to be of opposite sign to the simple effect of one factor alone. In this respect, Neyman's model of nature is a special case and its results cannot be generalized. It illustrates what could happen and not what will necessarily happen. Below, I will argue not only that the main effects of genes on fitness *can be* of opposite sign to their interaction effects on fitness but also that they *should be*. Thus, Neyman's 'frivolous' special case of nature may be *representative* of the case of gene interactions and fitness: the case on which Wright based his Shifting Balance Theory of Evolution.

3. GENE INTERACTIONS IN ADAPTIVE EVOLUTION

3.1 Wright and Fisher

The arguments of Wright and Fisher regarding gene interactions and adaptive evolution parallel those of Neyman and Fisher regarding factor interactions. Wright did not challenge Fisher's *definitions* of a gene's average effect or average excess but rather questioned the *utility* of these concepts for evolutionary biology. He discussed the inadequacy of such definitions for describing gene action and evolution on many occasions.

In recognizing the *multiplicity of genic effects* or pleiotropy as a universal property of living systems (Wright 1969, pp. 59–60), he distinguished the effect of a gene on fitness from the effects of that gene on other phenotypic traits. By 'fitness', Wright meant the relative numbers of progeny contributed to the next generation by a particular genotypic or phenotypic class. Wright also considered traits for which maximum fitness was associated with an intermediate value of a quantitative phenotype. In his comments on the 'inadequacy of the simple additive concept of gene effect' (Wright 1969, p. 419), he concluded that '... all genes that approach additivity in their effects on quantitatively varying characters

will be favorable in some combinations and unfavorable in others in terms of natural selective value (fitness) and, thus, exhibit interaction effects of the most extreme sort in the latter respect' (Wright 1969, p. 420). Thus, Wright identifies fitness as a unique trait with respect to the predominance of genetic interaction effects. He stresses the point that genic effects may be additive with respect to many characters but that fitness is a fundamentally different kind of character with strong interaction effects. This point is also made by Falconer (1981, p. 394): 'Abundant evidence proves that virtually all metric characters are genetically variable in populations that are more or less in equilibrium, including characters that affect fitness. There must therefore be genetic variance of fitness. But, since selection for fitness produces no response, there can be no additive genetic variance for fitness; so all the genetic variance for fitness must be non-additive, i.e., variance due to dominance and epistatic interactions.' Genotype-by-environment interactions could further complicate the evaluation of fitness.

Wright (1969, p. 420) elaborated on the multiplicity of the effects of single genes on the phenotype (universal pleiotropy): 'This again insures that natural selective value (fitness) is a function of the system of genes as a whole rather than something that can be assigned individual genes.' Here, Wright questions the proposition that a multivariate analysis of gene effects can result in a strictly additive partitioning of phenotypic variation among genetic factors. A partitioning adequate for describing the effects of a gene on a single character becomes inadequate when characters are combined into a whole system, a single unitary phenotype.

In his review of *The Genetical Theory of Natural Selection* (Fisher 1930), Wright (1930) discussed this difference between himself and Fisher with respect to gene effects. 'He [Fisher] assumes that each gene is assigned a constant value, measuring its contribution to the character of the individual (here fitness) in such a way that the sums of the contributions of all genes will equal as closely as possible the actual values of measures of the character in the individuals of the population' (Wright 1930, p. 353). (The phrase 'additive gene effects' means that 'the sums of the contributions of all genes will equal' the value of the character.) However, Wright (1930, p. 353) continues, 'Genes favorable in one combination, are, for example, extremely likely to be unfavorable in another.' That is, the additive concept of genic effect for Wright is not adequate for characterizing the relationship between a gene and fitness because the sign of a gene's contribution to fitness is not constant, but changes, depending on the genetic background.

The position taken by Fisher (1937, p. 108) is the opposite of that taken by Wright: 'In studying the properties of a system of interaction factors it has been shown (Fisher 1918) that departures from the simple additive law of interaction will usually have effects somewhat similar to

non-heritable modifications. We may therefore be confident that, even if a strictly additive interaction is not exactly realized, the mass effects of segregation in a large number of factors will closely simulate those of simple cumulative systems.' Fisher equated gene interaction effects with 'non-heritable' variation and advocated the additive model of a 'simple cumulative system.' The disparity of opinion on the importance of gene interactions is so extreme as to merit emphasis here. Fisher, on the one hand, dismisses gene interactions as comparable in their effect on the evolutionary process to the transient influence of 'non-heritable factors' (a view common to some of the current models of the evolution of continuous characters, e.g. Lande 1988). Wright, on the other hand, advocated the primacy of interaction effects in evolution and developed his Shifting Balance Theory as the most effective mechanism for selecting directly on gene interactions to insure creative adaptive evolution.

3.2 The parallels between Neyman and Wright regarding interactions

Just as Neyman and Fisher disagreed on the usefulness of the ANOVA as a method for estimating and interpreting main effects when interactions were present, Wright and Fisher also disagreed about the role of gene interactions in experimental genetics and evolutionary theory. There are several interesting parallels between the examples used by Neyman and Wright in their separate arguments with Fisher emphasizing the importance of interaction effects. The first is that the definition of additive genic effects and genic interactions depends on the application of ANOVA to specific breeding designs in experimental statistical genetics. As pointed out by Anderson and Kempthorne (1954, p. 897), 'The factorial gene model . . . is an adaptation of the factorial model used in the design of experiments.' From the statistical perspective, Neyman and Wright are discussing quite similar phenomena in their respective disagreements with Fisher. The argument of Neyman (1935), although not the example, is general and concerns the interpretation of the results of any factorial experimental design in terms of main effects and interactions (Milliken and Johnson 1984). The argument of Wright (1969, ch. 15) is necessarily more specific and concerns the application of such designs to the biological problem of the study of gene action.

The second parallel is that both Neyman and Wright selected *the same example* when interested in emphasizing the potentially important role of interactions. Neyman (1935) illustrated the problems of interpretation that arise in the ANOVA by using the example (Table 2, discussed above) in which the effects of a factor changed sign according to the other factors present. He showed how the average main effect of a factor could be found to be of opposite sign to one simple effect and thus lead to 'very wrong' inferences or conclusions. Similarly, Wright (1930, 1931,

1969) argued that the effect of a gene on fitness was dependent on the entire genetic system in which it was embedded and that this was important in large populations with small subdivisions. In generating gene interactions for fitness from additive effects on a quantitative character, Wright considered maximum fitness to be associated with an intermediate optimum value of the character. The phrase most often used by Wright was that, in such small subpopulations or demes, a gene's effect on fitness might be 'favorable in one genetic background and unfavorable in another.' The sign of a genic main effect (or any average main effect) can be changed by varying the constellation of other interacting factors *only if at least some of the interaction effects are of opposite sign to the gene's simple effect.* Without such a conflict in sign between simple main and interaction effects, the sign of the average genic effect must remain constant. Thus, both Neyman and Wright appealed to the same model with conflicting direct and interaction effects, when emphasizing the importance of interactions. The example used by Neyman was designed to illustrate the potential frequency of 'very wrong conclusions' from the ANOVA for the specific case when Nature was 'frivolous.' However, Wright argued that, with regard to fitness, this model of gene action is expected to be a very general one. That is, genic main effects and interactions on fitness are expected to be similar to those illustrated in Neyman's example. Neyman's specific example is *representative* of the relationship between genes and fitness discussed by Wright.

3.3 When Nature does small experiments

Lastly, Neyman argued that the problems of interpretation become most extreme in the analysis of the results of *small* experiments. As I have shown, he illustrated this point with a Monte Carlo simulation of 30 small experiments applied to the same idealized model of Nature. In his example, the repeated empirical conclusion from small experiments was that factor *A* had a beneficial main effect on yield when in fact its simple direct effect was to decrease yield. In an analogous manner, Wright argued that the evolutionary effects of gene interactions would be most important in large populations with *small* subpopulations. In such small populations, random genetic drift could create variation between populations in the system of gene interactions. Differently put, Wright believed that Nature does indeed do 'small experiments' in the ongoing process of adaptive evolution in subdivided populations.

Furthermore, just as Neyman illustrated how the interactions among factors can determine our inference regarding the nature of main effects, Wright argued that it is the interactions among genetic factors that determines their effect on fitness and ultimately their evolutionary fate. That is, owing to finite population size, it is a gene's interactions with

other genes that determine its relationship to fitness. A population subdivided into more or less isolated breeding groups is analogous to several small experiments being conducted on the same genetic system. The evolutionary fate of a gene in one deme within the subdivided population will be determined by the system of interacting genes characteristic of that small ‘experiment.’ Because the multiplicity of genotypes is extremely large in relation to population size, the genetic results of natural selection operating in local demes (Nature’s small experiments) can be diverse if gene interactions for fitness are important.

3.4 Falconer’s fitness component argument applied to gene interactions

In this section, I apply the arguments of Falconer (1981) regarding interactions among fitness components to the interactions expected among genes affecting fitness. I argue that genic main effects and interactions are *expected* to be of opposite sign, for those cases where selection maintains genetic variation and where genes interact to affect fitness. That is, genic main effects and interactions will usually have properties similar to Neyman’s (1935) model of Nature.

‘The “character” that natural selection selects for is *fitness*’ (Falconer 1981, p. 301) and the distributions of all other characters change in direct proportion to their correlation with fitness (Robertson 1966; Wade and Arnold 1980; Lande and Arnold 1983; Arnold and Wade 1984; Wade 1987). If two correlated characters are both subjected to directional selection, as would be any two components of fitness, then the genetic correlation between these two characters is expected to become negative at equilibrium. This happens because pleiotropic genes with positive effects on both characters experience strong directional selection toward fixation. Those pleiotropic genes with negative effects on both characters will experience strong directional selection toward loss from the population. Those genes, however, with positive effects on one fitness component and negative effects on another will experience much weaker selection and remain at intermediate frequencies for a longer period of time. It is these genes that will remain segregating in the equilibrium population and contribute to the genetic covariance among components of fitness. It is these genes that cause the genetic covariance between fitness components to be negative (Falconer 1981, p. 300).

This same argument can be extended to gene interactions. If alleles at two different loci each contribute positively to fitness and their interaction also has a positive effect on fitness, then these alleles will experience strong directional selection toward fixation. This will occur no matter what the effective population size because, even in small populations, the main effects of such alleles on fitness remain positive. When simple

main effects and simple interactions are of the *same* sign (positive in this case), sampling cannot change the sign of the estimated main effect. Similarly, consider the fate of alleles at two loci with negative main effects on fitness and negative interactions with respect to fitness. These alleles will experience strong directional selection toward loss from the population, again independent of population size. (Random genetic drift and selection will still operate according to classic population genetic theory and the outcome of selection will be less determinate in small populations. However, the direction of selection on such alleles cannot change sign for the reasons discussed above.)

In the population near equilibrium, only those alleles at the two loci whose main effects are of opposite sign to their interaction effects on fitness will remain segregating in the population (or be fixed more slowly). For the same reason that we expect fitness components to be negatively genetically correlated, we also expect that the genes segregating at intermediate frequencies in the population will be those that exhibit main and interaction effects on fitness of conflicting sign. Thus, the special case that Neyman used to illustrate the problems of interpretation expected in ANOVA when there are interaction effects is also the expected case with respect to gene action and fitness. For this reason and because the number of interacting factors is large, I argue that it is likely that small populations would differ sufficiently in genetic background that the effect of a gene on fitness in one deme would be of opposite sign to its effect on fitness in some other deme (Fig. 1). Each deme may be viewed as a small sample of some of the myriad of simple interactions possible (Table 56, p. 283, Lewontin 1974) and a gene's effect on fitness can be estimated within the limits of the local deme. Indeed, this is how we expect local adaptation by natural selection to operate. It remains an open empirical question, however, what proportion of new mutations have beneficial (or detrimental) main and interaction effects on fitness. The immediate fate of a new mutation must be closely tied to its 'simple interaction' effect in the genome in which it arises.

4. PREDICTIONS FOR EXPERIMENTALLY INVESTIGATING THE EFFECTS OF EPISTASIS ON ADAPTIVE EVOLUTION

The above arguments can be interpreted as making some testable predictions in natural or laboratory populations. There are three suggestions for empirical research that can be derived. (1) We should expect that the additive genetic variance for fitness could increase when equilibrium populations are inbred or subdivided. (2) We should expect that the average additive effect on fitness of single genes will vary when measured in different genetic backgrounds in different local demes. (3)

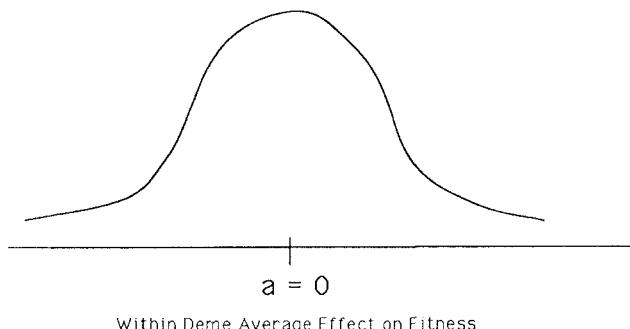


Fig. 1. A schematic illustration of the average effect on fitness and its variance among subpopulations. The abscissa is the value of the average additive effect, a , of an allele and the ordinate is the frequency of subpopulations or demes with different local values of a owing to differences among demes in a genetic background. In a large population at equilibrium, we expect the distribution to be centered at or very near zero, i.e. at equilibrium a gene should have little or no direct effect on fitness, positive or negative. A distribution about zero is created by subdivision of the population into more or less isolated demes so that the effect of the gene on fitness in some demes is positive but in other demes it is negative. The variance of this distribution should be a function of the degree of population subdivision, the extent of gene interactions for fitness, and the multilocus genotype frequencies. See text for further discussion.

Following from (2), we should expect that the genetic basis of the response to selection in small populations might vary from deme to deme even for uniformly imposed selection. I discuss each of these predictions in more detail below.

4.1 The change in additive genetic variance with population subdivision

The first prediction is not unique to epistatic systems. When rare deleterious recessives are made homozygous by inbreeding, an increase in the additive genetic variance for fitness can occur (Robertson 1952; Phillips and Wade 1992). A similar prediction arises from the following considerations for epistasis. Consider a large population of a haploid organism in which N loci affecting fitness are segregating for alternative alleles. I choose a haploid population in order to emphasize the effects of gene interactions between loci rather than allelic interactions or dominance effects within loci, which have received considerably more theoretical attention. There are N main effects and $N(N - 1)/2$ pairwise interaction effects. Both kinds of effects contribute to the total genetic variance. In a large population at mutation-selection equilibrium, we would further expect the additive variance for fitness attributable to main

effects to be nearly zero and owing to deleterious mutations according to Fisher's Fundamental Theorem.

If this population were subdivided to a value of Wright's inbreeding coefficient, $F = 0.50$, then half of the genes would be fixed and half of the genes would remain segregating in the average deme. Demes would differ from one another in the identity of the genes fixed, lost, and segregating. The fixation of $N/2$ genes implies that the numbers of main effects contributing to the genetic variance *within demes* is halved. However, fixation at half the loci also changes the within-deme constellation of simple interactions between the fixed loci and the $N/2$ loci still segregating. At each of these segregating loci, the local or within-deme average effect may be changed. For example, consider a building block model in which the phenotype of an individual is equal to the sum of the separate linear and quadratic contributions of only two component loci, X and Y, each with many alleles. We can represent a haploid two-locus genotype generally as X_iY_j , where the subscripts indicate different alleles at X and Y. Now, using the heuristic of the building block approach, the absolute fitness of a genotype, W, is defined as

$$W(X_iY_j) = 1 + a(X_i) + b(Y_j) + c(X_iY_j). \quad (3)$$

The genetic variance in fitness is then given as

$$\sigma_W^2 = a^2(\sigma_X^2) + b^2(\sigma_Y^2) + c^2(\sigma_{XY}^2). \quad (4)$$

Suppose, in one deme, that the allele at the Y locus is fixed, Y_c , without changing allele frequencies at the X locus. Then the fitness of a genotype in this deme is given by

$$W(X_iY_c) = (1 + bY_c) + (a + cY_c)(X_i) \quad (5)$$

and the genetic variance in fitness within this deme is

$$\sigma_W^2 = (a + cY_c)^2(\sigma_X^2). \quad (6)$$

The coefficient determining the contribution of the X locus to the genetic variance in fitness in eqn (6) can exceed that in eqn (4) depending on the relative values of a , c , and Y_c . That is, interaction effects in a large outbred population can contribute to main effects in a subpopulation derived from it by random genetic drift or inbreeding. Goodnight (1988) has referred to this phenomenon as the 'conversion' of epistatic genetic variance into additive genetic variance by random genetic drift. For this reason, we might expect the additive genetic variance for fitness to increase with inbreeding and population subdivision.

The magnitude of this effect depends on a number of different elements, including the number of loci, N , the relative strength of the linear and

interaction effects (here, a and c), and the gene frequencies. (In my example, I changed only the frequency at the Y locus and not those at the X locus to illustrate the effect. See Goodnight (1988) for a much more general diploid treatment.) If gene interactions for fitness are as important as suggested by Wright, then we might expect to see a measurable effect of inbreeding or random genetic drift on the additive genetic variance for fitness. Dominance variation can have a similar effect (Robertson 1952; Phillips and Wade 1992). The theoretical criteria for efficiently distinguishing the separate effects of dominance and epistasis on the additive genetic variance with inbreeding have not yet been elucidated (Goodnight, personal communication) but dominance effects on the within-deme variance must decrease with inbreeding beyond F of 0.50. For inbreeding with F values near 1, the among-deme genetic variance will be twice the original within-population genic variance plus four times its additive-by-additive epistatic variance. At complete fixation, the dominance variance does not contribute to the among subpopulation genetic variance. Thus, large increases in the genetic variance for fitness among demes with extreme inbreeding might reasonably be attributed to epistatic rather than to dominance effects.

We investigated the effects of population subdivision with laboratory populations of flour beetles, *Tribolium spp.* (Wade 1977, 1982, 1985, 1990, 1991; McCauley and Wade 1980, 1981; Wade and McCauley 1980, 1984; Wade and Goodnight 1991). We created experimental arrays of demes, each with a different population structure, using different values of N and different amounts and patterns of migration (m). In almost all population structures, we found large heritable differences in fitness arising among demes within a period of 10 to 15 generations. The population structures we studied correspond to values of F across the range of 0.02 and 0.33 (breeding adults per deme: $6 < N < 96$; migration rates: $0.00 < M < 0.25$). This rapid differentiation among demes occurred despite finding little or no additive genetic variance for fitness (Wade 1985) and, for the range of F values studied, a large increase in among-deme variance is not expected on the basis of dominance effects (Crow and Kimura 1970, pp. 339–345). However, approximately $4F(1 - F)$ of the original epistatic variance for fitness is converted into additive genetic variance within demes (Whitlock, personal communication).

The most relevant experimental study is that of Wade and Goodnight (1991) in which we created a laboratory model of Wright's process of interdemic dispersion. In our study, each subdivided population consisted of 50 demes and each deme was founded with 20 breeding adults. After a 62-day period of population growth, Wright's process of interdemic selection was imposed on the entire array of demes. Migrants were chosen from the most productive demes, those with the highest values of fitness, and distributed into the least productive demes. Our protocol weighted each deme by its fitness *relative to the average demic fitness across the*

array: the largest demes contributed the most migrants and the lowest demes received the most migrants. Each experimental array had its own paired control array of 50 demes which experienced the same amount of *random* migration. We observed a significant increase in mean fitness over time in all three population structures ($0.025 < F < 0.050$) in which we imposed Wright's process.

The empirical evidence necessary to separate dominance effects from epistatic effects in these experiments is not decisive, but it tends to implicate epistasis for fitness rather than dominance. For small values of Wright's inbreeding coefficient F and variance in fitness owing only to recessive deleterious genes in low frequency in the stock populations, we would not expect to observe a large increase in among-deme genetic variance for fitness or a large response to artificial interdemic selection (Wade and McCauley 1980, 1984; Wade 1982, 1990, 1991). For an F of 0.10, less than 3 per cent of the total genetic variance is expected to be among demes. In some studies, we were able to detect a correlation between population mean fitness and expected heterozygosity (McCauley and Wade 1981; Wade and McCauley 1984), even for weak levels of random genetic drift ($0.02 < F_{ST} < 0.10$). If this were due to an increase in frequency of rare deleterious recessives with inbreeding, then we would expect to observe an increase in mean deme fitness on outbreeding. However, direct and extensive experimental tests comparing inbred and outbred subpopulations at different densities did not support this expectation for the low values of F (McCauley and Wade 1980; Wade and McCauley 1980). In summary, the rapid genetic differentiation among demes for fitness for low levels of inbreeding ($0.02 < F < 0.10$) and the lack of a direct response of mean fitness to outbreeding argue against dominance variance for fitness (owing to recessive deleterious alleles) as the sole cause of the genetic effects we observed in subdivided laboratory populations of flour beetles. Direct measures of the change in the within-deme additive genetic variance for fitness with inbreeding are underway and may assist in discriminating the effects of dominance from epistatic genetic variance.

A second line of empirical evidence suggesting that epistasis for fitness is important in these experiments comes from the Shifting Balance Experiment (Wade and Goodnight 1991). We found that interdemic selection at every second generation produced nearly a two-fold larger increase in demic fitness than selection every generation or every third generation. In an additive world, one would expect the response to selection to be proportional to the intensity of selection (the selection differential). We did not find this. It is as though a store were 10 blocks away and we arrived at the store sooner by walking every other block than by running every block.

4.2 Change in the average additive effect on fitness with inbreeding and random genetic drift

Wright argued that an allele might have a positive effect on fitness in one genetic background but a negative effect on fitness in a different genetic background (see above). This is especially true for an allele that contributes additively to a continuously distributed trait for which an intermediate value has the highest or optimum fitness. If gene interactions for fitness are often or usually of opposite sign to genic simple effects, then we ought to be able to detect among-deme variations in the average additive effect on fitness of single genes. The degree of population subdivision necessary to detect such variance in average additive effects *among demes* will depend on several different factors. I will use a simple two-locus model from Crow and Kimura (1970, p. 79) to illustrate this dependence.

The model presented in Table 3 (from Crow and Kimura 1970, p. 79) includes additive effects (A and B), dominance effects (D_A and D_B) and epistatic interactions (I , J , K , and L) between two diallelic loci. The average effect, a_1 , of the A_1 allele is defined as the linear regression of genotypic value on the number of A_1 alleles in the genotype. Let p_{A1} and p_{A2} be the frequencies of the A_1 and A_2 alleles at the A locus and, similarly, v_{B1} and v_{B2} , be the frequencies of the B_1 and B_2 alleles at the B locus. The a_1 is given by

$$a_1 = A + D_A(p_{A2} - p_{A1}) + I(v_{B1} - v_{B2}) + J2v_{B1}v_{B2}(p_{A2} - p_{A1}) \\ + K2v_{B1}v_{B2} + L(v_{B1} - v_{B2})(p_{A2} - p_{A1}).$$

Table 3
Two-locus model of genotypic values with dominance and epistasis (from Crow and Kimura 1970)

Genotype at the A locus			
	A_1A_1	A_1A_2	A_2A_2
B_1B_1	$A+B+I$	D_A+B+L	$-A+B+I$
B_1B_2	$A+D_B+K$	D_A+D_B+J	$-A+D_B-K$
B_2B_2	$A-B-I$	D_A-B-L	$-A-B+I$

A , B = additive effect of the A_1 and B_1 alleles, respectively.

D_A , D_B = dominance effects at A and B loci, respectively.

I = additive-by-additive interaction.

J = dominance-by-dominance interaction.

K = additive(A)-by-dominance(B) interaction.

L = additive(B)-by-dominance(A) interaction.

When there is no dominance ($D_A = D_B = 0$) and no epistasis ($I = J = K = L = 0$), then a_1 is simply equal to A and independent of gene frequency. To see the effects of epistasis in an array of subpopulations derived from a larger randomly mating population, set $D_A = D_B = 0$. In those demes where B_1 is fixed ($v_{B1} = 1.0$) but gene frequencies are unchanged at the A locus, we have

$$a_1 = A + I + L(p_{A2} - p_{A1}) . \quad (7)$$

But, in those demes where the alternative allele, B_2 , is fixed, we have

$$a_1 = A - I - L(p_{A2} - p_{A1}) . \quad (8)$$

Clearly, the value a_1 can change in magnitude and sign depending on gene frequencies at both loci as well as on the relative magnitudes of the additive (A) and epistatic (I and L) effects. In a strictly additive world, the average effect could not change sign.

Figure 1 illustrates the possible consequences of population subdivision on the average additive effect of a gene on fitness. In a large, randomly mating population, a_1 is expected to be very near zero for a gene contributing to fitness. In a subdivided population, averaged over the entire array of demes, the value of a_1 is also very near zero but it may deviate from this expectation in either the positive or negative direction in different demes (Dempster 1963). Figure 1 is schematic and illustrates what could happen. The specific distribution for any particular gene across an array of demes in a subdivided population is unknown. Several important empirical issues remain unanswered, including: (1) is there significant variation in the average additive effect on fitness of a specific allele in different genetic backgrounds?; (2) can random genetic drift lead to differences among demes in genetic background sufficient to change the average effect of a gene in the manner suggested by Wright?; (3) to what degree and for how long must an array of subdivided populations be isolated before significant variance among demes in the average additive effect of a gene on fitness can be measured?; (4) how does the variance about the average additive effect of a gene on fitness (Fig. 1) change with time and degree of population subdivision?; and (5) do natural populations exhibit a sufficient degree of population subdivision for these effects to be evolutionarily important? The picture in Fig. 1 is a static one and for any real array of subpopulations the variance of a should change with time and continued subdivision if the arguments of Wright are correct.

4.3 The heterogeneous genetic effect of homogeneous selection

The third prediction concerns the expectation of a heterogeneous genetic response among demes to homogeneous directional selection. If artificial selection for increased body size were imposed on a series of replicate small populations, derived independently by inbreeding from a common outbred stock, the considerations discussed above imply that the response to selection in different demes could involve different gene interaction systems (Dempster 1963). Even if the separate demes responded to selection to the same degree phenotypically, the genetic basis of the response might differ among lineages because the same gene could contribute to the local demic response in opposite ways given sufficient differences among demes in genetic background. Again, the important empirical issues are to what degree must the lineages be inbred before such effects can be detected and, if detected, do the levels of inbreeding correspond to those that might occur in natural populations owing to random genetic drift? This is a possibility and we need an empirical evaluation of its probability in natural populations.

Several experimental studies have been conducted to investigate whether the limits to individual selection can be enhanced or diminished by a combination of selection within and among demes. Wright believed that this kind of selection might permit the selection of epistatic gene combinations. The majority of these studies, however, involve artificial selection on primarily additive traits (e.g. Katz and Young 1975; Madalena and Robertson 1975; Rathie and Barker 1968) where among-line selection is always expected to reduce the total response owing to the loss of alleles in lineages extinguished during selection. Only the study by Katz and Young (1975) found population subdivision to enhance significantly the selection response and they did not report an investigation of the genetic basis of the among-lineage genetic variation. Furthermore, a significant fraction of the total response to selection in both the other studies was due to a small number of genes of large effect, initially at low frequency in the base population. In general, strong artificial selection on an additive trait is unlikely to lead to significant among-line variance in gene interactions for the selected trait but may lead to among-line variance in fitness owing to linked loci.

4.4 Epistasis, adaptation, and the Shifting Balance Theory

It is important to emphasize that the three predictions made above, even if confirmed experimentally, do not necessarily guarantee the existence of multiple adaptive peaks or the efficiency of their export by differential migration among demes, the latter phases of Wright's Shifting Balance Theory. Their confirmation would suggest that the genetic basis of local

adaptations is heterogeneous among demes, even for uniformly imposed selection, and that the process of local adaptation is itself inherently variable in genetic outcome owing to gene interactions. However, Wright's theory not only requires the existence of such interactions and their differentiation among demes but also that they must produce significant differences among demes in local mean fitness. Furthermore, the interaction systems underlying local adaptations and local mean fitness must be spread to neighboring demes with different interaction systems and at lower fitness peaks by interdemic selection (differential migration). That is, variation among demes in systems of interacting genes must lead to variation among demes in mean fitness, and this, in turn, must lead to the differential dispersion of migrants away from high peaks and into demes at lower peaks. The latter two effects do not follow necessarily from the existence of demic variation in interaction systems.

In fact, the conditions for genetic differentiation among demes may be contrary to those necessary for the efficient export of adaptive peaks. The genetic differentiation of local demes requires restricted migration but the efficient export of adaptive peaks appears to require somewhat more extensive migration. Recent theoretical research (Crow *et al.* 1990) indicates that perhaps these conditions are not as restrictive or contradictory as they appear (Nei 1987). If the domain of attraction of a high fitness peak is large, then small amounts of interdemic migration may be sufficient to initiate a 'peak shift,' the genetic transformation of a local deme from the gene interaction system characteristic of a low fitness peak to the interaction system characteristic of a higher peak. A peak shift is achieved by the joint action of interdemic dispersal *and* individual selection within the receiving deme and not solely by interdemic dispersal. Crow *et al.* (1990) investigated what they referred to as the 'critical migration rate,' the minimum rate of interdemic migration necessary to effect a peak shift. This can be viewed as a situation in which a peak shift occurs primarily as a result of individual selection subsequent to migration rather than to the interdemic selection process itself. (Clearly, *both* are required for a peak shift.) They found, first, that a peak shift would occur after a number of generations of differential migration *even if further migration was halted completely*. This clearly demonstrates that selection within demes can be responsible for much of the shift from one interaction system to another and that the peak shift process does not depend solely on interdemic selection by differential dispersion. All else being equal, the broader the domain of attraction of the higher fitness peak, the more the shift will depend on selection within demes rather than among demes. Secondly, they found that peak shifts could occur even when the rate of differential migration from the high peak to the low peak was an order of magnitude less than the reverse migration. Again, this demonstrates the potentially important role of

individual selection within demes for the peak shift and indicates that the first and third phases of Wright's process need not be in conflict. The small migration rates necessary for genetic differentiation may also be sufficient for exporting adaptive fitness peaks. (Remember that, for Wright, mass selection within demes is necessary for a 'fitness' peak.)

In a recent study of Wright's Shifting Balance Theory, using experimental arrays of the flour beetle, *T. castaneum*, Wade and Goodnight (1991) demonstrated that significant increases in mean fitness can occur with interdemic migration rates less than 1 individual per deme per generation with an effective deme size of approximately 20 breeding adults. This is a level of migration that is often associated with significant genetic differentiation of demes but not with significant effects on the evolution of mean fitness. Additional experiments indicate that there is a non-additive genetic basis for this response (Wade, in preparation).

The relationship between epistasis, adaptation, and Wright's theory depends on a number of additional genetic and ecological features of natural populations. However, despite these caveats, the role of gene interactions in adaptive evolution and the origin and spread of adaptations involving the coordinated action of many genes would be better understood if the kinds of empirical data discussed above could be obtained. Furthermore, the empirical predictions made here appear to be related directly to the evolutionary inferences regarding gene interactions drawn by Wright.

5. CONCLUSIONS

Neyman in statistics and Wright in evolutionary genetics both challenged Fisher on the role of interactions in the interpretation of Nature and they did so using arguments of remarkable similarity. By investigating the issues in these disputes, some general implications can be drawn and some novel experimental predictions for evaluating the role of gene interactions can be made. The experimental investigation of these predictions may permit a more complete understanding of the evolutionary importance of gene interactions and the robustness of current theory that by and large ignores them. The ultimate test of Wright's Shifting Balance Theory depends on empirical demonstration of the heterogeneity of genetic response to local selective pressures, the existence, density, and size of adaptive peaks, and the competence of interdemic selection by differential migration for exporting a favorable gene combination from one deme to the next. Although Wright's Shifting Balance Theory has been referred to as a 'cornerstone of modern evolutionary thought' (Mettler *et al.* 1988) and 'the dominant theory of evolution in the 20th century' (Collias 1991), a number of its requirements and predictions await empirical investigation.

ACKNOWLEDGMENTS

I would like to acknowledge gratefully the thoughtful and critical comments of Drs Jim Crow, Jerry Coyne, Patrick Phillips, John Willis, Felix Breden, Jim Griesemer, Joe Travis, Dave McCauley, Charles Goodnight, Lori Stevens, Chris Boake, Alan Templeton, Robert Boyd, Thomas Park, Michael Whitlock, Doug Futuyma, and Alexander Gimelfarb. Drs Coyne and Crow, in particular, generously provided thoughtful and extensive comments on the manuscript and helped me greatly to correct and clarify my ideas. Many of the conceptual aspects of this work started to develop during and after a 6-week visit to the University of Edinburgh in 1984, which was supported by a PHS grant K04 HD00431, NIH grant GM22523, and a Burroughs-Wellcome travel grant. I am grateful for the open and challenging intellectual environment provided by Drs Alan Robertson, Bill Hill, Douglas Falconer, Trudy MacKay, Linda Partridge, and Jerry Wilkinson during that visit.

REFERENCES

- Anderson, V. L. and Kempthorne, O. (1954). A model for the study of quantitative inheritance. *Genetics* **39**, 833–98.
- Arnold, S. J. and Wade, M. J. (1984). On the measurement of natural and sexual selection: theory. *Evolution* **38**, 709–19.
- Cockerham, C. C. (1954). An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**, 859–82.
- Collias, N. E. (1991). The role of American zoologists and behavioural ecologists in the development of animal sociology. *Anim. Behav.* **41**, 613–31.
- Crow, J. F., Engels, W. R. and Denniston, C. (1990). Phase three of Wright's Shifting-Balance Theory. *Evolution* **44**, 233–47.
- Crow, J. H. and Kimura, M. (1970). *An introduction to population genetics theory*. Harper & Row, New York.
- Dempster, E. R. (1963). Concepts and definitions in relation to selection schemes. In *Statistical genetics and plant breeding* (ed. W. D. Hanson and H. F. Robinson), pp. 34–43. Natl. Acad. Sci.-Natl. Res. Council. Pub. No. 982, Washington, DC.
- Falconer, D. S. (1981). *Introduction to quantitative genetics*, 2nd edn. Longman, New York.
- Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edin.* **3**, 399–433.
- (1925). *Statistical methods for research workers*. Oliver & Boyd, Edinburgh.
- (1930, 1958). *The genetical theory of natural selection*. Dover Publications, Inc., Toronto.
- (1937). *The design of experiments*. Oliver & Boyd, Edinburgh.
- (1941). Average excess and average effect of a gene substitution. *Ann. Eugen.* **11**, 53–63.

- Futuyma, D. J. (1986). *Evolutionary biology*, 2nd edn. Sinauer Associates, Inc., Sunderland, MA.
- Goodnight, C. J. (1988). Epistasis and the effect of founder events on the additive genetic variance. *Evolution*, **42**, 399–403.
- Katz, A. J. and Young, S. S. Y. (1975). Selection for high adult body weight in *Drosophila* populations with different structures. *Genetics* **81**, 163–75.
- Lande, R. (1988). Quantitative genetics and evolutionary theory. In *Proceedings of the second international conference on quantitative genetics* (ed. B. S. Weir, E. J. Eisen, M. M. Goodman, G. Namkoong), pp. 71–84. Sinauer Associates, Inc., Sunderland, MA.
- and Arnold S. J. (1983). The measurement of selection on correlated characters. *Evolution* **36**, 1210–26.
- Lewin, B. (1990). *Genes IV*. Oxford University Press, New York.
- Lewontin, R. C. (1974). *The genetic basis of evolutionary change*. Columbia University Press, New York.
- Madalena, F. E. and Robertson, A. (1975). Population structure in artificial selection: studies with *Drosophila melanogaster*. *Genet. Res., Camb.* **24**, 113–26.
- McCauley, D. E. and Wade, M. J. (1980). Group selection: the genetic and demographic basis for the phenotypic differentiation of small populations of *Tribolium castaneum*. *Evolution* **34**, 813–20.
- (1981). The populational effects of inbreeding in *Tribolium*. *Heredity* **46**, 59–67.
- Mettler, L. E., Gregg, T. G. and Schaffer, H. E. (1988). *Population genetics and evolution*, 2nd edn. Prentice Hall, California.
- Milliken, G. A. and Johnson, D. E. (1984). *Analysis of messy data*. Van Nostrand Reinhold Co., New York.
- Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press, New York.
- Neyman, J. (1935). Discussion of a paper, 'Complex experiments,' by Yates. *J. R. Stat. Soc., Suppl.* **2**, 235–42.
- Phillips, P. C. and Wade, M. J. (1991). The effects of inbreeding on heritability: dominance and overdominance. *Genetics*, in review.
- Provine, W. B. (1986). *Sewall Wright and evolutionary biology*. University of Chicago Press, Chicago.
- Rathie, K. A. and Barker, J. S. F. (1968). Effectiveness of regular cycles of intermittent artificial selection for a quantitative character in *Drosophila melanogaster*. *Aust. J. Biol. Sci.* **21**, 1187–1213.
- Robertson, A. (1952). The effect of inbreeding on the variation due to recessive genes. *Genetics* **37**, 189–207.
- (1966). A mathematical model of the culling process in dairy cattle. *Anim. Prod.* **8**, 65–108.
- Traxler, R. H. (1976). A sang in the history of factorial experiments. In *On the history of statistics and probability* (ed. D. B. Owen), pp. 283–95. John Wiley & Sons, New York.
- Wade, M. J. (1977). An experimental study of group selection. *Evolution* **31**, 134–53.

- (1982). Group selection: migration and the differentiation of small populations. *Evolution* **36**, 949–61.
- (1985). The effects of genotypic interactions on evolution in structured populations. In *Genetics: new frontiers. Proceedings of the XV international congress of genetics*, New Delhi, pp. 283–90. Oxford and IBH Publishing, New York.
- (1987). Measuring sexual selection. In *Sexual selection: Testing the alternatives* (ed. J. W. Bradbury and M. B. Andersson), pp. 197–207. John Wiley & Sons, New York.
- (1990). Genotype–environment interaction for climate and competition in a natural population of flour beetles, *Tribolium castaneum*. *Evolution* **44**, 2004–11.
- (1991). Genetic variance for rate of population increase in natural populations of flour beetles, *Tribolium spp.* *Evolution*, in press.
- and Arnold, S. J. (1980). The intensity of sexual selection in relation to male sexual behavior, female choice, and sperm precedence. *Anim. Behav.* **28**, 446–61.
- and Goodnight, C. J. (1991). Wright's Shifting Balance Theory: an experimental study. *Science*, in press.
- and McCauley, D. E. (1980). Group selection: the phenotypic and genotypic differentiation of small populations. *Evolution* **34**, 799–812.
- and McCauley, D. E. (1984). Group selection: the interaction of local deme size and migration in the differentiation of small populations. *Evolution* **38**, 1047–58.
- Williams, G. C. (1966). *Adaptation and natural selection*. Princeton University Press, New Jersey.
- Wright, S. (1930). Review of Genetical Theory of Natural Selection by R. A. Fisher. *J. Heredity* **21**, 349–56.
- (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.
- (1959). Physiological genetics, ecology of populations, and natural selection. *Perspect. Biol. Med.* **3**, 107–51.
- (1969). *Evolution and the genetics of populations*, vol. 2. University of Chicago Press, Chicago.
- Yates, F. (1935). Complex experiments. *J. R. Stat. Soc., Suppl.* **2**, 181–247.

Coevolution among competitors

MARK L. TAPER and TED J. CASE

1. INTRODUCTION

How frequently does coevolution occur? Does it influence the structure of communities? If it does shape communities, how does it shape them? These are questions that have been subjected to intense debate for decades. Half a decade ago, the evidence even as mustered in several book-long reviews (Thompson 1982; Futuyma and Slatkin 1983; Nitecki 1983) was somewhat equivocal. We believe that this will be an exciting decade in the study of coevolution in which recently developed empirical and theoretical tools will lead to insights into these questions. In this review we will focus our attention on coevolution among competitors, specifically the phenomenon of morphological character shifts, also called character displacement.

We reserve the use of the term ‘character displacement’ or ‘character shifts’ to refer to the joint evolution of morphological character differences between competing species resulting from selection pressures created by the species interactions. As we will elaborate, resource competition is only one of several mechanisms that might produce such selection. Throughout this paper, we refer frequently, but not exclusively, to the characters of body size and size of trophic apparatus. Displacement may also occur in other morphological (shape), behavioral (e.g. vocalizations, nest placement, habitat selection), physiological, or phenological (e.g. activity or blooming periods) traits. Species shifts in behavior, habitat, and resource use are more likely to have a large non-genetic component, and thus not necessarily represent evolved differences. By restricting ‘character displacement’ to morphological characters, we do not mean to degrade the role of coevolution in influencing these more phenotypically flexible traits but only to constrain our focus on characters whose shifts are clearly evolutionary.

Semantically, we can segregate coevolved character shifts into three patterns (Fig. 1).

1. Character divergence; characters of two species in sympatric populations diverge from their allopatric ancestors in opposite directions such that the character difference in sympatry is greater than in allopathy. This is sometimes called *character displacement* although others use ‘displacement’ more generically to mean coevolved

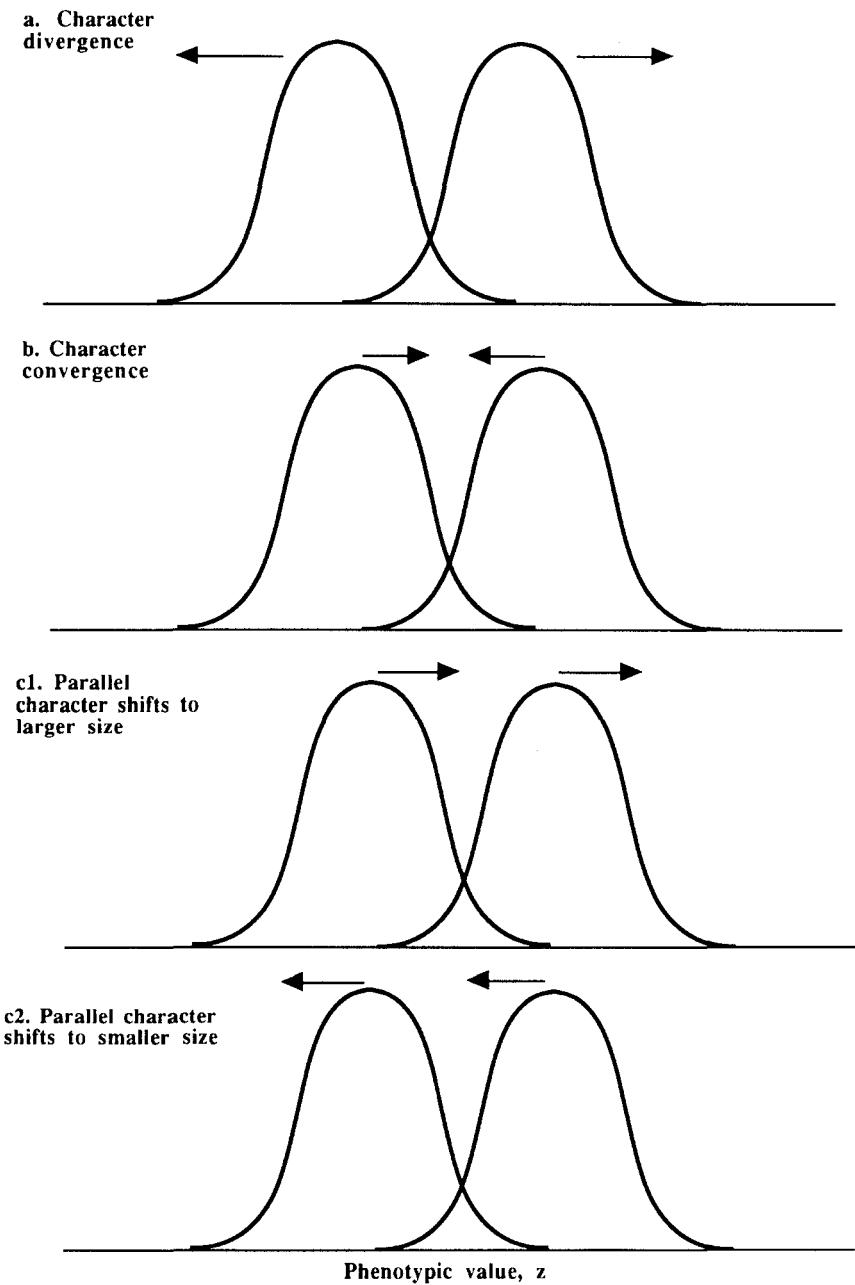


Fig. 1. An illustration of different modes of coevolutionary character shifts (i.e. character displacement). In each subfigure, the two Gaussian curves show the distribution of phenotypes in two hypothetical species and the arrows indicate the direction of selection: (a) character divergence; (b) character convergence; (c1) parallel character shifts to larger sizes; (c2) parallel character shifts to smaller sizes.

character shifts in any direction including convergence (e.g. Abrams 1987a). We will follow that convention here. Community ecologists often point to character divergence and the resulting divergence in resource use as an important aspect of community structure (Hutchinson 1959; Schoener 1965; Lack 1971; Grant 1972; Cody 1974; Case and Sidell 1983).

2. Character convergence; the character states of two species evolve towards a more central character state in sympatry compared with allopatry. It seems counter-intuitive that competition might sometimes lead to character convergence and so we spend some time explaining how this theoretically might arise.
3. Parallel character shifts; two species evolve in the same direction, either upward or downward in size, in sympatric populations relative to allopatric populations. Since the two species might evolve at somewhat different rates, the net character difference between them might be greater or lower in sympatry compared with allopatry. Parallel character shifts typically emerge in models when competition between phenotypes is asymmetric such that the competitive impact of large phenotypes on small is greater (or less) than the effect of small on large. It is hypothesized to play a central role in the character shifts of anoles in the Lesser Antilles (Rummel and Roughgarden 1985) and to generate a taxon cycle wherein competitive coevolution leads to extinction of a species, followed by vulnerability of the community to new species invasions, followed, in turn, by new bouts of coevolution, etc.

We will explore the details of the models that yield these parallel shifts in Section 5.

The term character displacement was originally coined to mean only character divergence (pattern 1) (Brown and Wilson 1956). However, subsequent authors (Grant 1972; Abrams 1987a) have expanded the term to mean coevolved shifts in any direction (i.e. patterns 1, 2, and 3). In this review we follow this latter convention.

Only recently have some of the central issues in character displacement been clarified or resolved. We focus here on four of these.

1. We examine the relationship between the evolution of character divergence in specific species pairs and the character structure of multi-species guilds. Does character divergence inevitably lead to predictable community structures and conversely can coevolutionary processes be inferred from the species size distributions in communities?
2. When coevolutionary character shifts occur in nature, how fast do they evolve?
3. Until very recently, there were no successful experimental demon-

strations of character displacement with laboratory organisms. We discuss the probable causes of previous failures and we summarize some very recent experimental results establishing character displacement with adzuki bean weevils.

4. Finally, we compare the various ways that character displacement has been modeled. Certain simplifying assumptions that make analytical models tractable can lead to false conclusions. This is particularly troublesome when making extensions from pair-wise coevolution to community assembly processes and evolutionary dynamics in multi-species competition communities (e.g. taxon cycles). We show how these pitfalls arise, how to avoid them, and, most importantly, what avenues future research should take to resolve still outstanding controversies.

2. THE ECOLOGY OF CHARACTER DISPLACEMENT

2.1 Ecological mechanisms

The ecological mechanism by which one species might create a directional selection differential for another species can be categorized according to how interspecific and interphenotypic interactions are mediated. (1) Selection differentials may be created by effects mediated by species at lower trophic levels, that is through resource utilization and the resulting interspecific competition for those resources. Similarly-sized species might compete more than dissimilarly sized species, thus providing some impetus for the evolution of character divergence (the ‘competition avoidance’ mechanism). (2) Size similarity might affect ecological interactions with higher trophic levels. Similar-sized species may be more heavily preyed on by search-image-forming predators, or share more parasites or pathogens than dissimilar-sized species, providing a directional selection for size divergence (the ‘predation avoidance’ mechanism). Other aspects of physical appearance may also affect search images, leading to character divergence in habitat, shape, or ‘aspect’ (Ricklefs and O’Rourke 1975). On the other hand, if prey are distasteful or have potent defenses, then similar-looking prey might experience less predation than dissimilar ones. This is the typical explanation for the evolution of Müllerian mimicry (Gilbert 1983). (3) Size similarity might affect mating or social interactions among species at the same trophic level. Size-similar phenotypes in different species might hybridize more readily; if the offspring of a hybrid mating are less fit than non-hybrid offspring, then this creates a situation favoring size divergence in the two species (the ‘hybrid avoidance’ mechanism; Bossert 1963; Spencer *et al.* 1986; Tsukagoshi 1988). This is often referred to as *reproductive character divergence* (or displacement).

Another way that interspecific social interactions might lead to character divergence is through agonistic behaviors. Theoretically, size-similar phenotypes in different species might agonistically interfere more readily than size-dissimilar species. This could create selection differentials for size divergence (Cody 1969, 1973; Abrams 1986). We discuss this further in section 2.4.

In the remainder of the paper we will focus almost exclusively on mechanism 1 above, which we will refer to as *competitive character divergence*, but it is well to keep the alternative mechanisms in mind when trying to interpret observed patterns in nature. The fact that competitive character displacement has been more actively modeled and invoked to explain field situations does not necessarily mean that it is universally more important. Recent work by Martin (1988a, b) on open-nesting passerine birds indicates that the predator avoidance mechanism might be operating with respect to the character 'microhabitat of nest placement.'

Field examples of reproductive character divergence are also available, for example, in the mating calls of *Litoria* frogs in Australia (Littlejohn 1965). Hybrid avoidance can set in motion a unique ecological dynamic which might limit its overall generality. First consider the simplest case where the two populations undergoing contact produce completely sterile or non-viable offspring and range contact is extensive. The rarer of the species suffers a relative disadvantage, all else being equal. This is because all the hybrid offspring that are produced have two parents – one from each species. Thus, if some number, x , of hybrid offspring are produced in one generation, x represents a greater proportion of the mating effort of the rare species than of the common species. Consequently, the rarer species' mean fitness is reduced more than the common species' by hybridization, and so in the next generation the rare species will decrease in relative frequency while the common species will increase (assuming hybrids have reduced fitness). This ecologically destabilizing process is completely unlike that in competitive character displacement, in which rare species, if anything, can be favored. This situation was modeled by Spencer *et al.* (1986); it was often the case that extinction would occur before reproductive character divergence could evolve to eliminate hybridization and 'rescue' the rare species.

In situations where range contact is due to the introduction of a small propagule of one species into the center of the other species' range, as might be the case when one species colonizes an island occupied by the other, this effect will be most dramatic. Now the introduced species is both rare and geographically surrounded by its congener. On the other hand, if range contact is due to the secondary formation of a narrow zone of overlap between two species that have extensive allopatric ranges elsewhere (as in many continental situations), the rare-species disadvantage

may not dominate the dynamics. However, we know of no models that contain explicit, spatially distributed dynamics depicting both the ecological and population genetic effects for this situation.

If introgression occurs (i.e. hybrids have reduced fitness but not zero fitness), then the resulting evolutionary trajectory is even more difficult to predict and the subject of much current debate (reviewed in Butlin 1987). Selection pressures for enhanced mating segregation (often called 'reinforcement') will favor character divergence in both species, but this diverging selective force is countered by interspecific gene flow, tending to homogenize the two species.

2.2 Conditions for competitive character displacement

Even if species modify selection differentials for one another, it is by no means certain that character displacement of any sort will occur. If competition is too severe, one or both species may become extinct before evolutionary changes can occur. There are also genetic constraints. There must be genetic variability in the character of interest and there must not be an all-purpose 'super' phenotype that has the highest fitness regardless of the frequencies of other phenotypes. This idea that a 'jack of all trades can be a master of none' is a fundamental assumption, tacit or explicit, of all models of character displacement based on competition avoidance, beginning with MacArthur and Levins (1967).

Without a trade-off in performance of specific phenotypes on alternate resources, character displacement cannot be expected to occur (see Via and Lande (1985) for a related discussion of adaptation to multiple habitats/hosts in the absence of trade-offs). This idea that each phenotype is best at using a specific resource type is *prima facie* so reasonable that it has not received the full attention it deserves, either empirically or theoretically. Correlations between character expression and resource utilization have been shown to occur in natural populations of a diverse array of organisms. Food item size is related to body size measures in lizards (Roughgarden 1972; Case 1979), mud snails (Fenchel 1975), and ants (Davidson 1978). The most detailed work of this sort has been done on the relationship between morphology and diet of Galápagos finches. Many researchers have reported on the feeding mechanics, food item profitability, and diet selection in these birds. Morphologically based trade-offs in the ability to use the available resource spectrum have been unequivocally demonstrated, and trade-offs have been clearly reflected in the actual diets. The finch morphology/diet literature is extensively reviewed in Grant (1986). On the other hand, Wiens and Rotenberry (1981) show that morphology does not predict diet in North American shrubsteppe birds (at least during the breeding season).

2.3 Detecting character displacement

Usually, one looks for evidence of character displacement by examining the characters of the same species in allopatry and in sympatry. Grant (1972, 1975) pointed out that parallel clinal variation in two species may produce a pattern resembling character divergence. Such pseudo-character divergence could arise when both species are declining in size across a cline solely for autecological reasons, but one species replaces the other geographically, with a narrow band of sympatry. In this band of sympatry the first species is at its smallest size in the cline, while the second is at its largest size, giving the appearance of character divergence. Yet the same size patterns for each species would appear even if the other were absent.

Another empirical way to detect character displacement is by careful use of the fossil record. However, rarely is the stratigraphy sufficiently fine-scaled to allow the analysis of size changes in the presence or absence of specific other species. Moreover, taxonomic problems will be more severe in the fossil record than with living organisms. Two similar species could be mistaken as a single species and this error would be more likely when species were convergent than divergent. However, Eldredge (1974) identified an example of character divergence (and character convergence) involving two species of trilobites in the middle Devonian. Kellogg (1980) presented evidence for character divergence in Miocene to Recent radiolarians from deep-sea sediments and Schindel and Gould (1977) discussed a possible case of character divergence in Pleistocene land snails from Bermuda.

In the absence of a clear fossil record, modern phylogenetic analysis using new molecular methods may be a great help in identifying which species, and even populations within species, are derived from which others. The largest difficulty is in inferring the state of the character in question in common ancestors. Nevertheless, phylogenetic analysis has the potential of making substantial contributions in the study of character displacement. As an example, Losos' (1990) phylogenetic analysis of Lesser Antilles anoles indicates that the large and small species of lizards form two separate phylogenetic groups. This makes it unlikely that ongoing character divergence is responsible for the striking pattern of size dissimilarity of sympatric lizards in the archipelago. The combination of phylogenetic methodology, molecular genetic techniques, and rigorous ecological measurements in the field will be an exciting merger. Long-standing questions about cause and effect in the evolution of ecological segregation (Lack 1944, 1971; Diamond 1986) should yield to this approach in the coming decade.

As a guide for further empirical studies, we offer the following six criteria to establish an unambiguous case for competitive character

divergence among a species pair (after Schluter and McPhail 1992). We begin with a statement of the pattern.

1. Morphological differences between a species pair in sympatry are *statistically* greater than those of allopatric populations for the same species pair. That is, sampling error should be ruled out as an explanation for the pattern.
2. The observed size differences between sympatric and allopatric populations have a genetic basis.
3. Differences in the character state of the species from sympatry to allopatry cannot be ascribed simply to different colonization sources of the populations. That is, in sympatry where one species is relatively large, this is not simply because the population there is descended from colonists that were large, but rather large size evolved *in situ*. This, for example, is one problem with inferring character divergence among whiptail lizards (*Cnemidophorus*) on islands in the Sea of Cortez. There are two sources of colonization for *C. tigris*: mainland Mexico (with a medium body-sized subspecies) and Baja California (which has a large body-sized subspecies). Perhaps the one-species island populations of *C. tigris*, which are intermediate in size, are all descended from a mainland colonist while *C. tigris* on two-species islands (with the small *C. hyperythrus*) are descended from the Baja Californian colonist.
4. The character state affects resource use so that differences in the character correspond to different resources being used to different degrees.
5. Species and phenotypes compete for the resources implicated in (4) and the strength of this competition is positively correlated with the degree of character similarity.
6. Differences in the mean character from sympatry to allopatry cannot be ascribed simply to changes in resource distributions from place to place.

The first three criteria validate the pattern as character divergence and the last three establish the mechanism generating that pattern as competition avoidance. While the last three criteria may seem unduly strict if applied to purported cases of character divergence in the literature, models show that even if all three criteria are satisfied, character divergence may still not occur (see section 5). Since in many field situations it will be difficult to verify all six conditions, we suspect that many actual cases of competitive character displacement may never be *unambiguously* proven.

Among invertebrates, which often have indeterminate growth, one can find large differences in the mean or maximum body size in a population simply because of environmental effects: differences in growth rate or

differences in age structure. This can create a situation of apparent character divergence without any real genetic change. Thus it is particularly important in these cases to document a genetic basis for the character shift. The evidence for such genetic changes may come from breeding or common garden studies, but for some organisms these would be impractical. Alternatively, we may be compelled to accept less direct evidence; if we find that, standardized for body size, the size of a trophic apparatus like tooth number differs between populations and species, then there is a high likelihood of genetic difference. Or a phenotypic character might be inferred to have a strong genetic component if it shows limited spatial variance despite substantial environmental heterogeneity. Unfortunately, this has not been the case in some of the textbook examples such as Fenchel's work on aquatic snails (1975) and Davidson's (1977) study of mandible size in desert ants of the southwest US. Both characters are very labile within populations and are non-standardized for overall size.

The most complete study to date, coping with all six of the criteria, is the work of Grant and colleagues on the Galápagos ground finches (summarized in Grant 1986). Less complete, but still relatively thorough studies of competitive character divergence are on the anoles of the Lesser Antilles (Roughgarden and Pacala 1989) and *Cnemidophorus* lizards on islands off Baja California, Mexico (Case 1979). A host of other studies, too numerous to enumerate, establish only criteria 1 and/or 2, albeit sometimes very convincingly (Dayan *et al.* 1989, 1990; Malmquist 1985).

As we have said, not all attempts to look for character divergence or even size regularity in guilds have been successful (see Simberloff and Boecklen 1981). The absence of these patterns may be due to an absence of interspecific interactions, or such interactions may be present but they yield selection for convergence or parallel character shifts. Even if divergence is the prevailing selective force, the strength and even direction of selection may be mitigated by biological variability of different sorts. Competition, for example, could be intermittent and not clearly related to any one specific character difference. Many characters might simultaneously be affected by competition but these characters lack genetic variation or they might have negative genetic covariances creating an evolutionary inertia to selection (see below). The set of species interacting may vary seasonally or yearly. Even if the competitors are constant, the set of contested resources may change such that formerly weak competitors become strong ones and vice versa. In this way, natural selection might not be able to screen out temporarily inferior species or genotypes before conditions change and new competitive relationships are established (Murray *et al.* 1987).

2.4 Character convergence

Character convergence is usually inferred in nature where sympatric species are more similar than allopatric ‘relatives’ (Cody 1969). As with character divergence, without a known phylogeny placing these relatives into an ancestral sequence, determining where and when convergence arose can be tenuous.

The idea that competitors might converge in characters seems counter to the underlying logic behind character divergence. Nevertheless, there are at least two very different reasons why convergence might evolve among resource competitors. The first is based on facilitating interspecific territoriality (Cody 1973), or flock aggregation, or predator mobbing behavior (Moynihan 1968). Typically, the characters most often converging in these examples are not morphological characters important in feeding, but instead are signals used in individual or species recognition: calls, colors, and patterns. A striking example of apparent character convergence occurs among the woodpeckers of southeast Asia. The species *Dinopium javanense* and *benghalense*) and *Chrysocolaptes lucidus* have broadly overlapping geographic distributions and are nearly indistinguishable in sympatry with respect to coloration and feather patterns. The only subspecies that are not strikingly similar are those in allopatry. Wallace (1869) described the remarkable case of apparent character convergence of an oriole (*Oriolus*) and friarbird (*Philemon moluccensis*) on the island of Beru in the Malay Archipelago. On the nearby island of Ceram the two species appear quite different from the Beru birds but again are convergent on each other. The same is true for this species pair in New Guinea (Cody 1969). In this example, however, it is not clear what the original state of the character was in these species.

Cody (1973) has developed a model for the evolution of character convergence among competing bird species to facilitate interspecific territoriality. When food overlap between a pair of species is high and territory size is large, both species will improve net energy profit by becoming mutually territorial, thus allowing smaller territories, which are easier to defend and less depleted of food. Character convergence, in this model, evolves to facilitate the behavioral responses required to maintain this advantageous interspecific territoriality. The model is a strategic optimization argument at the population level. It does not explicitly consider genetic variation in the morphological character or in the agonistic behavior; presumably they would be controlled by different loci. Explicit genetic models of this process are needed.

The second and very different explanation for convergence among competitors is due to Abrams (1987b). What is surprising about this model is that, like the models that yield character divergence, it has its roots in competition theory and the characters undergoing divergence do

affect competitive interactions. Suppose two consumer species each need two (or more) essential but non-substitutable resources. Each species must maintain a certain ratio of these discrete resources to fulfil its nutritional requirements. Depletion of one resource by the other competitor will mean that more time must be spent searching for this rarer resource. Any morphological characters that enhance a species' ability to find or process that most-limited resource will then be selected for. In this way, two species might both become more similar in sympatry than in allopatry, where the two resources are less depleted.

The model assumes that species have unique differences in their harvesting abilities to begin with (i.e. in allopatry) such that they would not converge on the same phenotype even in places where the resources were identically distributed. This might more often be the case among competitors that are in very different taxa with different physiological processes and constraints, rather than congeneric species. Unfortunately, attempts to search for character changes in very different taxa are few. The only documented case of character shifts between very different taxa (birds and bees) seems to demonstrate character divergence, not character convergence (Schluter 1986). The Galápagos finches *Geospiza fuliginosa* and *G. difficilis* are smaller in size and consume proportionally more nectar in their diet on islands lacking carpenter bees than on islands with carpenter bees. The bees do not occur in allopatry with respect to these finches so size shifts in bees cannot be examined. In this case the resource in question, nectar, is not an essential resource for the taxa in which divergence was measured.

Although experimental tests are lacking, the Abrams model seems cogent and deserves empirical investigation. Its usefulness will be limited if competition is more common for resources that are substitutable. This seems to be the case where competition has been documented among animals in different taxa, such as seed-eating ants and seed-eating rodents (Brown *et al.* 1986), or cone-eating finches and cone-eating squirrels (Benkman 1989; Pimm 1990).

Tilman (1988) makes a strong argument that, for plants, above-ground resources (light and carbon dioxide) and below-ground resources (water and soil nutrients) are fundamentally non-substitutable. Evolutionary studies are needed of geographic shifts in plants' investments in above-versus below-ground investments in resource acquisition, depending on the presence or absence of competitors. In common greenhouse experiments, Martin and Harding (1981) compared the competitive ability of two species of the annual plant, stork's bill (*Erodium*), collected either from sympatric or allopatric sites. They found that when grown together the total seed yield of plants whose origin was from sympatric populations was greater than that for allopatric populations, demonstrating the evolution of competitive ability. However, they did not examine whether

these changes were associated with the evolution of particular morphological, physiological, or life history features of the plants.

2.5 Community-level character patterns

2.5.1 From pairwise character divergence to guild patterns

The ecology literature has always put more stock in character divergence than character convergence or parallel character shifts. Consequently, the prevailing opinion is that if coexisting species reduce competition by using different resources, and if species differences in resource utilization are determined by key morphological or phenological characters like jaw size, bill length, or blooming periods, then actual guilds of competitors should exhibit a greater regularity in the characters mediating competition than hypothetical ‘null communities’ that are randomly assembled without regard to any limiting similarities (Strong *et al.* 1979; Gleeson 1981).

Ever since Hutchinson wrote his classic paper ‘Homage to Santa Rosalia’ (Hutchinson 1959), ecologists have sought three patterns in the size of competitors. If species in local guilds are ordered by size (body size or the size of trophic apparatus), then (1) the ratios of the sizes of adjacent species should be more constant than expected by chance; (2) local guilds should have larger minimum ratios than expected by chance; and (3) ratios of linear dimensions of adjacent species should cluster around the value of the Hutchinsonian ratio 1.3 (2.2 for mass ratios) (see references in Simberloff and Boecklen 1981). It is again important to emphasize that even if all three of these expectations are met, this does not necessarily indicate that evolution of character divergence has been the cause, since the tests do not compare the same species in sympatry and allopatry.

The Hutchinsonian ratio has only been weakly supported in that length ratios tend to fall between 1 and 2 (Roth 1981; Eadie *et al.* 1987). Furthermore, Eadie *et al.* show that even this weak concordance may be a statistical artifact. Body size ratios between 1 and 2 will almost always result if sets of species are randomly drawn from body size distributions that are lognormal and have a small standard deviation ($\sigma < 1$). In the more than 40 taxonomically or ecologically related groups of species reviewed by Eadie *et al.*, body sizes were almost invariably distributed lognormally with small standard deviations (but see Bowers and Brown 1982). Does this mean that the concept of a Hutchinsonian ratio is completely worthless? It may very well be, but we are not yet entirely convinced. Maiorana (1978) guesses that the logic may work in the opposite direction, that some limit to the similarity of competing organisms explains why the distribution of body sizes in taxonomically related species is lognormal with low variance in the first place (see also Bowers and

Brown 1982). In this light, it is interesting that at larger taxonomic scales (e.g. all mammals) body sizes are no longer lognormally distributed (Brown and Nicoletto 1991; Maurer *et al.* 1992).

Statistical rigor in the analysis of these size patterns began to develop in the early 1980s with the development of appropriate null models (Case *et al.* 1983; Schoener 1984; Colwell and Winkler 1984; Schlüter and Grant 1984*a, b*; Brandl and Topp 1985) and with the application to community analysis of the Barton and David (Simberloff and Boecklen 1981) and the Behrens-Fisher (Roth 1981) tests for constancy of size ratios, as well as the Irwin test (Simberloff and Boecklen 1981) for unexpectedly large minimum ratios.

Simberloff and Boecklen (1981) tested for statistical significance a large number of claims of constancy of size ratios and unexpectedly large minimum ratios. Using the Barton and David test for constancy of size ratios, Simberloff and Boecklen found that only four out of 21 studies rejected the null hypothesis (random selection of sizes from a log-uniform distribution) at the 5 per cent type I error level. Unexpectedly large size ratios fared even worse; only one out of 18 studies rejected the null hypothesis at the 5 per cent level, when tested with the Irwin test.

What do these failures to reject the null hypothesis mean? Simberloff and Boecklen (1981) themselves wondered if these tests might have low statistical power (i.e. high type II error rates). This supposition has been demonstrated with simulations by Losos *et al.* (1989). If this is the case then species assemblages that differ from the null hypothesis in important ways may not have been detected (see Toft and Shea (1983) for a discussion of the importance of power analysis in community ecology). Losos *et al.* reanalyzed the studies used in Simberloff and Boecklen in a fashion that increased statistical power. They use Fisher's combined *P* method (Sokal and Rohlf 1981) to effectively increase sample size. To avoid the bias introduced in consortium tests if negative results are not published (Simberloff 1983; Boecklen and NeSmith 1985), Losos *et al.* conservatively combined only probabilities from independent assemblages reported within single studies. In the tests of ratio constancy, 12 of the original 21 studies present data on more than one assemblage. When Fisher's method is applied to these 12 studies, six of 12 tests reject the null hypothesis at the 5 per cent level, compared with one out of 12 rejections found by Simberloff and Boecklen. In tests of minimum size ratios, six of 10 studies with multiple assemblages reject the null hypothesis, instead of one of 10.

So far we have focused on closely related species in single guilds. In isolated species-poor environments, certain taxa sometimes speciate profusely, forming great swarms of species that cross classical guild lines. These swarms, with divergence among species occurring mostly in functional characteristics such as trophic apparatus, foraging behavior,

and habitat use, have been termed ‘adaptive radiations.’ Classic examples include the Hawaiian honeycreepers (Bock 1970; Raikow 1977), the African cichlids (Fryer and Iles 1972; Greenwood 1981), and the Galápagos finches (Lack 1947; Grant 1986). Although competition may play a role in producing many specialized species rather than a few generalized ones (Grant 1986), the primary implication of these radiations is that diffuse competition has constrained the evolution of these taxa in their ancestral species-rich communities. Schlüter (1988) has analyzed this hypothesis quantitatively, comparing the degree of overall morphological differentiation within genera on islands and on continents, and has concluded that species from island genera are more broadly separated from one another in morphological space despite the fact that the island genera are typically younger. There is also anecdotal evidence that plant taxa can assume ecological niches occupied elsewhere by very different taxa. The repeated radiation of trees and shrubs from ancestral herbaceous lines on islands like the Canaries, St. Helenas, and Hawaii is a prime example (Carlquist 1965; Schlüter 1988).

2.5.2 *From guild patterns to inferences about character displacement*

Because of the bias stemming from the under-reporting of negative results, the analysis of Losos *et al.* (1989; discussed above in section 2.5.1) does not imply that 50–60 per cent of all species assemblages are non-randomly constructed with regard to species size. It does, however, demonstrate that highly regular size patterns in coexisting species are not rare. Other documented examples include the sizes of coexisting felids and mustelids (Kiltie 1984; Dayan *et al.* 1989, 1990) raptors (Schoener 1984), desert rodents (Bowers and Brown 1982; Brown and Kurzius 1987), West Indian bird communities (Case *et al.* 1983), and *Cepaea* land snails (Cowie and Jones 1987). The question as to why some guilds in some places show size regularity and others do not still needs attention. Systematic comparisons of size structure between guilds, taxa, and locations are needed. Are guilds that compete for food more likely to show size regularity than guilds more typically limited by predators, nest sites, parasites, etc? As yet there is no clear answer.

Regular size patterns by themselves are insufficient evidence for character divergence since a detectable regularity in such characters need not necessarily be due to coevolution among these species. There are at least two other alternatives. (1) Species may assort non-randomly or persist differentially such that dissimilar species are preserved more than similar ones (i.e. non-random community assembly). (2) The underlying distribution of resources (or predators) might itself be multi-modal with the resource peaks non-randomly regularly distributed. In this scenario, the consumers’ phenotypes are simply mapping onto an underlying multi-

modal distribution of resources. Probably the most thoroughly studied case of a guild that has strict size structure is that of the Galápagos ground finches. Here non-random species assortment, multi-modal resource peaks, and character divergence all seem to have contributed to producing the overall regularity in bill size of coexisting ground finches on islands in the archipelago (Bowman 1961; Case and Sidell 1983; Grant and Schlüter 1984; Schlüter and Grant 1984a). After measuring and accounting for non-random species associations and the differing shape of resource (seed size and hardness) distributions from island to island, character divergence between specific finch pairs is still evident. However, it is by no means common given all the potential species pairs and islands in the archipelago. The most convincing case is that originally offered by Lack (1947): *Geospiza fortis* and *G. fuliginosa* are more similar in allopatry on the islands of Los Hermanos and Daphne than they are in sympatry on Santa Cruz. The underlying distribution of seed sizes on these islands is not the same (Schlüter and Grant 1984b), but even after accounting for seed size differences, it is clear that the species' beak sizes are more divergent in sympatry than expected if the species were not competing for seeds. Although no hard evidence exists at this time for character displacement outside this Daphne system, it may well have occurred all over the Galápagos but not be as noticeable.

3. RATES OF CHARACTER DIVERGENCE OR RELAXATION

3.1 An outline of evidence

Given changes in selective pressures induced by the presence or absence of competitors, how quickly can character shifts take place? This is a difficult question to answer because the dates at which specific pairs of species have established contact are generally poorly known. However, several lines of evidence indicate that character displacement can occur very rapidly. A few examples of character displacement can be found with known limits on the maximum time available for evolution. These have shown remarkably rapid character changes (Section 3.2). In a laboratory setting, character displacement was observed in bean beetles in only nine generations (Section 4). Theoretical studies indicate that character displacement occurs with a broad range of rates, depending on community composition, but it can be extremely rapid (Section 5). Finally, one can study rates of morphological evolution on islands of known ages or colonization histories (Section 3.3). Here again, the evidence implies that morphological evolution can be very rapid.

3.2 Rapid character divergence

Schluter and McPhail (1992) have reported a remarkable case of character divergence among stickleback fish in lakes formed only 12 000 years ago in British Columbia. Lakes have either one or two species. If two species are present, one is a limnetic form (small with many gill rakers) and one a benthic form (generally larger with few gill rakers). Single-species lakes have fish that are intermediate in size and foraging mode.

Similarly, Diamond *et al.* (1989) discuss an apparent case of character divergence involving myzomelid honeyeaters on islands off New Guinea. Long Island is occupied by two species of myzomelids and the size of these birds is more different than that in allopatric populations on nearby islands. Since Long Island was devastated by a volcanic eruption about 300 years ago, these size changes have presumably evolved quite recently. The two species do not occur elsewhere in sympatry.

The character divergence found by Fenchel (1975) between the mud snails *Hydrobia ventrosa* and *H. ulvae* at a number of localities in Limfjord, Denmark, must have occurred since the fjord was opened to the sea in 1825. Thus no more than 150 generations were required to accomplish divergences in shell length of several standard deviations.

Pygmy shrews, *Sorex minutus*, and common shrews, *S. araneus*, show character divergence in northern Europe (Malmquist 1985). The allopatric populations of pygmy shrews are found on land-bridge islands having colonization windows between 8000 and 9000 years BP. In this case it is character relaxation after competitive release that has occurred quickly.

Not all examples of character divergence are known to be rapid. On islands in the Sea of Cortez, divergence between the lizards *Cnemidophorus hyperythrus* and *C. tigris* was not found on land-bridge islands (about 13 000 years old), but only on oceanic islands with ages over one million years (Case 1979).

3.3 Rapid evolution on islands

Although quantitative data are preferable, records of endemism can be used as crude markers for significant evolution. In many cases it is not known if the selective forces driving this evolution involve interspecific competition. Nonetheless, these studies do give an indication of how fast morphology can respond to selective pressures. One way to approach this is through the fossil record but this is usually not very precise. Another way is to make use of archipelagoes containing islands of varying ages and with species both in allopatry and sympatry on young and old islands. Human introductions of species in the recent past also offer evolutionary biologists unique opportunities to measure character evolution in different environmental settings and under different demographic conditions.

In general, distinctive island vertebrates are restricted to older oceanic islands. Holocene land bridge islands (connected to the mainland during the last glaciation and only about 12 000 years old) usually have few, if any, endemic species. For example, in the Sea of Cortez the average oceanic island has 35 per cent endemic reptile species and 69 per cent endemic mammals, while on land bridge islands endemism is only 5 and 16 per cent, respectively (Case 1983). The very largest of Holocene land bridge islands in the world, like New Guinea, Borneo, or Sri Lanka, contain much larger percentages of endemic species in birds, mammals, and reptiles, but these percentages are not substantially larger than similar percentages based on equal-sized parts of the mainland (Case, unpublished data). The generally low endemism on land bridge islands may stem from a number of causes. Clearly morphological evolution cannot happen overnight; the longer the time that directional selection operates, the greater the character evolution that can result. On islands this morphological evolution would lead to character differences constituting endemic status. Another hypothesis is that land bridge populations, beginning their existence with large populations, do not undergo founding population size bottlenecks, like that for oceanic islands that are colonized by oversea dispersal. Consequently, they may not have the right demographic environment for rapid evolution. A third factor is that on land bridge islands, the selective regime may not be as different from the mainland as that of oceanic islands. Land bridge islands typically begin their insular existence with nearly a full complement of mainland species, so the competitive environment may not generate selection pressures any different from those on the mainland. All three factors could simultaneously operate. If founder effects and divergent competitive selection regimes were more important for morphological evolution than elapsed time, then Holocene oceanic islands should exhibit much more morphological evolution than Holocene land bridge islands. However, only one well-studied archipelago allows this comparison. On the two Holocene volcanic islands in the Sea of Cortez (Raza and Tortuga), species endemism for reptiles and mammals is comparable to that on the much older oceanic islands and is much greater than that observed on land bridge islands similar in size (Case 1983). Unfortunately, there are very few other archipelagos in the world where all three types of islands exist in one place, so the generality of these results is not known. Certainly, the Long Island myzomelids discussed earlier (section 3.2) and the evidence from introduced species (see below) suggest that morphological evolution can be very rapid on islands when population bottlenecks and/or sharply different selective regimes are encountered.

Other island situations around the world reveal that morphological evolution may be very rapid, sometimes even on land bridge islands. In the last interglacial, after the island of Jersey (British Channel Islands)

lost its land connection to the mainland, red deer on Jersey became reduced to one-sixth of their original body weight in less than 6000 years (Lister 1989). The population later became extinct so we do not know the genetic basis of this dramatic body size reduction. This insular dwarfism is typical of many insular ungulates but is generally restricted to older oceanic islands (Case 1978); its selective causes are unknown but character displacement does not seem to play a role.

The carnivorous lizard *Varanus rosenbergi* inhabits the southern coast of South and West Australia and a number of land bridge islands offshore. The maximum body size on Kangaroo Island and Reevesby Island is substantially larger than that seen on the mainland (Case and Schwaner 1991). On Reevesby Island the population was introduced by man less than 100 years ago. While the size shifts in these varanids are not due to the presence or absence of competitors (i.e. character displacement) they do support the notion that body size can track resource availability; large body size forms are found on islands where prey are large-sized and abundant (Case and Schwaner 1992). Unfortunately, there is no solid evidence that these character changes are genetically based.

Some time around the early 1940s *Anolis bimaculatus leachi* was introduced from Antigua or Barbuda in the Lesser Antilles, to Bermuda (Wingate 1965). Today the body size of this anole is substantially larger than that on Antigua, but interestingly still smaller than the maximum size of this lizard seen in subfossil deposits 4000 years old (Pregill 1986). Again, it is uncertain to what extent this size change has a genetic component and character divergence is not implicated.

Laysan finches introduced to Pearl and Hermes Reef displayed significantly different beaks within 17 years in apparent adjustment to changes in the availability of seeds of different hardness (Conant 1988). Introduced populations of the European tree sparrow in North America and Australia show significant departures from ancestral European populations in morphology and genetics (Barlow 1973; St. Louis and Barlow 1988).

One of the most dramatic insular changes in body size in reptiles occurs in black tiger snakes (*Notechis ater niger*) on islands off the coast of southern Australia. Most of these islands were connected to one another or to the mainland between 6000 and 9000 years ago (Robinson *et al.* 1986), yet today the snakes on different islands are very different in scale counts, allozyme frequencies, and particularly body size (Schwaner 1985, 1990; Schwaner and Sarre 1988, 1990; Williams *et al.* 1988). The island of Hareby has snakes that are nearly five times heavier than those on the adjacent island of Roxby, even though the two islands were connected only 6000 years ago. Here again, the major selective agent affecting these body size shifts appears to be the sizes of available prey species on different islands; the mean (or maximum) adult size of the snakes is

closely correlated with the mean (or maximum) body size of available prey species (Schwaner 1985; Shine 1987). For tiger snakes, there is some breeding evidence that indicates that these size differences are at least partly genetic (Barnett and Schwaner 1985; Schwaner, unpublished data).

4. LABORATORY STUDIES

4.1 Laboratory studies of the evolution of competitive ability

Given the rapidity with which morphological characters can evolve in nature in response to the competitive milieu, it is surprising that there have been so few attempts to create situations in the laboratory that might lead to character divergence. The results of experimental investigations in the closely related topic of the evolution of competitive ability are surprising. The common result is that, counter to expectation, culturing different species or different lines of the same species together for extended periods of time does not lead to consistent or even significant changes in their competitive abilities (Park and Lloyd 1955; Futuyma 1970; Sokal *et al.* 1970; Dawson 1972; Sulzbach and Emlen 1979; Sulzbach 1980).

In many of these experiments the competing populations were very similar to start with: often they were strains of the same species. This was done intentionally to maximize niche overlap and hence competition. However, in the models of Taper and Case (1985) the same set of parameters describing the competitive universe can lead to two evolutionary end-states depending on the initial intensity of competition between the species. In particular, species that initially compete strongly (because they have highly overlapping phenotype distributions) will evolve to exactly the same phenotype, whereas species initialized with more different phenotypes and thus weaker interactions may diverge over time. (It is important here to point out that Taper and Case (1985) and Slatkin (1980) used the term 'convergence' as an adjective to explain this outcome. However, both species converging on the same size in these models is not 'convergence' in the sense of allopatric versus sympatric comparisons since both species would also evolve to this same size in allopatry; there is no convergence *relative* to allopatry only *relative* to the initial sizes of the species.) In the few laboratory cases where significant improvement of competitive ability appeared to have evolved (Seaton and Antonovics 1967; Bryant and Turner 1972), the competing organisms were deliberately established with initial character differences.

4.2 An experimental demonstration of character divergence

4.2.1 Experimental design

Recently Taper (1990) has published an experimental demonstration of character divergence between species of bean beetles. This study was motivated by the models of Taper and Case (1985) and the experimental studies of the evolution of competitive ability mentioned above. Consequently, the competitors were chosen with strong initial asymmetries. To simplify the design, observations, and analysis, the experiment was set up to cause an evolutionary response to competition in only one of the two species. This is termed the **target** species, while the other is considered the **driver**. The **target** species could use both of two resources, while the more specialised **driver** could use only one. Therefore, character displacement would be indicated if at the end of the experiment the ability of the **target** species from the experimental lines to use the resources is significantly different from that of the **target** species in the control lines. Character divergence or convergence could result.

The **driver** was an old highly monomorphic laboratory strain of *Callosobruchus maculatus*. The **target** was a strain of *C. chinensis* with high genetic variation created by crossing 11 laboratory strains of *C. chinensis*. *C. maculatus* is known to be superior to *C. chinensis* at larval interference competition (Yoshida 1966; Toquenaga and Fujii 1991; Taper, personal observations).

Two types of beans were used as resources, mung bean (*Vigna radiata*) and lentil (*Lens culinaris*). The adult beetles lay eggs on the beans; the larvae develop within and adults emerge in about 20 days. Both beans were novel to both strains. Initially both beetle strains grew extremely well on mung. Lentil, on the other hand, was essentially lethal to the **driver** and severely reduced the emergence of the **target** and retarded its development. Because the **target** can complete its development on either bean, these resources can be considered substitutable for the **target**. Control treatments included **target** alone on each bean separately and together, and a hard selection treatment in which the target was allowed to oviposit on both mung and lentil but all mung were discarded before beetle emergence. The hard selection treatment was included to assess the maximum possible effect of resource competition (complete exclusion from mung) on the evolution of the **target**. Because of the initial asymmetry between species both in resource utilization and in larval interference, one can predict that competition lines of the **target** should evolve an increased utilization of the resource not used by the **driver** species.

4.2.2 Characters measured

To avoid potentially confounding effects of larval environment on adult ovipositional preference or of maternal environment on larval development, all lines were reared for one generation on adzuki beans after termination of the selection portion of the experiment. The lines were then tested for performance on mung and lentil. Performance was tested in the absence of competition, in that beetles were reared one hatched egg to a bean.

After nine generations Taper (1990) measured several traits that reflect the ability of the beetles from the various lines to use the two bean types. These included oviposition preference, the proportion of hatched eggs developing to emergence, the development rate in the two beans, and a body size measure, the elytron (hardened forewing) length.

Oviposition preference is a resource capture trait. Proportion of hatched eggs developing to emergence and development rates are indirect indicators of resource utilization efficiency. Elytron length is a more complex trait to interpret. As a body size measure, it is strongly correlated with female potential fecundity. However, there are indications (Smith and Lessels 1985) that males and females are subject to differing selection pressures on body size.

4.2.3 Results, discussion, and further work

Evolution was detected in five of the six characters measured. A multivariate analysis revealed at least two independent underlying traits, a developmental time/proportion emergence trait, and a beetle wing length or body size trait. Character divergence did occur, in that the competition and no-competition treatments differed significantly, and, **target** lines raised on lentil plus mung in the presence of **driver** did better on lentil (greater proportion of emergence and faster development) than did **target** lines raised on lentil plus mung alone without the **driver**. After nine generations, elytra of the targets were statistically longer in competition lines than in control lines. Since the **driver** is larger than the **target** species, this evolution may represent a character convergence. There is some evidence that this change is not resource driven but is a response to extra-bean behavioral interference.

A surprise was that the **target's** oviposition preference for bean type did not evolve. This is particularly interesting in light of results of Wasserman and Futuyma's (1981) hard selection experiments using the aQ strain of *C. maculatus* (Taper's **driver**). They observed the evolution of preference but not physiological adaptation. Taper obtained the opposite results with his **target** species, *C. chinensis*. The resolution of this apparent discrepancy reveals aspects critical to both the design of laboratory studies and to ecological genetic models of character displacement. The history of the aQ strain is important. It is a very old

laboratory strain having been reared in the laboratory since 1946 – over 500 generations at the time of Wasserman and Futuyma's experiment. This represents strong and continuous selection for adaptation to a single host, the adzuki bean. Genetic variation in characters relating to physiological adaptations to host beans should be depleted. On the other hand, during aQ's long laboratory history many characters influencing behavioral preference would have been neutral since no alternative hosts were available. This may explain why aQ was able to respond to selection by shifting preference, not by physiological adaptation. Taper (1990) created a new outbred strain of *C. chinensis* with high genetic variation for his experiments, and this presumably explains why it was able to adapt physiologically.

It is still unclear, however, why behavioral preferences did not evolve. Taper (1990) hypothesized that a critical difference between the experiments was that female density relative to the number of beans was low in Wasserman and Futuyma's experiment, but unconstrained and quite high in Taper's. It is known that one of the primary influences on bean beetle oviposition preference is the number of eggs already on a bean; females avoid beans with eggs already present (Avidov *et al.* 1965; Mitchell 1975). At high beetle densities, intraspecific competition for prime oviposition sites (i.e. beans without eggs) would tend to cause females to spread their eggs between both types of beans. This spread of a female's eggs more or less evenly over both bean types reduces between-female fitness differences owing to site preference in the oviposition of early eggs, weakening directional selection for more divergent bean type preferences.

One feature noted by Taper (1990) is that the beetles did not appear to display any trade-off between the two bean species in their ability to 'process' the two resources. After nine generations, all lines appeared to use mung equally well, regardless of the selective regime to which they had been exposed. A character-based trade-off is central to all models of character displacement. Without such a trade-off a species should eventually evolve to use all resources perfectly despite the presence of competition. Recent reviews of other insect/plant systems suggest that the apparent absence of such trade-offs in this system is not an isolated instance (Futuyma and Moreno 1988; Jaenike 1990). If the absence of a trade-off was real, then the observed character displacement might have been a transient phenomenon. Competition may simply facilitate adaptation to lentil that would ultimately occur anyway.

Because of this uncertainty, Taper (manuscript in preparation) continued selection on all lines for an additional 16 generations (same treatments, but with modified containers and maintenance protocols). Measurements made at the termination of the experiment (generation 25) indicated that statistically significant differences remained in all the characters that had

diverged earlier. The competition lines were still significantly better adapted to lentil than were the controls in both the proportion emergence and developmental rate. However, trade-offs were indicated: differences among the treatments in both male and female development rates on mung were statistically significant, with the competition lines developing more slowly than control lines.

5. MODELING CHARACTER DISPLACEMENT AND COMMUNITY ASSEMBLY

5.1 Introduction

Three of the most complete field studies of character displacement have used the 'coevolutionary stable community' (or CSC) approach of Roughgarden (1976) to predict quantitatively the expected phenotypes of competing species in sympatry and allopatry. These are the studies of *Cnemidophorus* lizards on islands off Mexico (Case 1979); *Anolis* lizards in the Lesser Antilles (Rummel and Roughgarden 1983, 1985; Roughgarden and Pacala 1989); and Galápagos ground finches (Schluter and Grant 1984b; Grant 1986). This method involves the maximization of conditional population densities. Roughgarden and colleagues have extended the two-species model to deal with asymmetric competition and a scenario of multiple invasions and coevolution in island groups.

However, character displacement has also been modeled using two very different approaches: (1) evolutionarily stable strategies (ESS) involving the maximization of individual fitness under frequency-dependent selection but with no explicit genetics (Lawlor and Maynard Smith 1976; Slatkin and Maynard Smith 1979; Abrams 1986, 1987a, b; Brown and Vincent 1987a); and (2) quantitative genetic recursions (QGR) entailing the recursion of selection/response expressions linking genotypic interactions and fitness effects until an equilibrium is reached (Slatkin 1980; Milligan 1985; Taper and Case 1985; Case and Taper 1986). The models do not always produce the same results, particularly when competition is asymmetric. Yet comparison of the predictions of the various models has been hindered because the ecological details of the models analyzed have been very different.

In Taper and Case (1992) we compared these three approaches to modeling coevolutionary dynamics in a framework of common ecological details so that dynamics and equilibria can be compared and differences in model behavior ascribed to specific differences in underlying assumptions rather than difference choices for parameters. We were particularly interested in asymmetric competition for two reasons. First, some degree of asymmetry is a general feature of natural competition systems.

This follows from first principles based on the allometry of resource consumption and has been abundantly verified in nature (Brooks and Dodson 1965; Lawton and Hassell 1981; Schoener 1983; Case 1984; Spiller 1986). Second, it is under asymmetric competition that the three models are most likely to differ (Brown and Vincent 1987a; Taper 1988).

We were also interested in aspects of model behavior that would relate to observable features of community structure: are coevolved communities invadable? What niche packing features do they share? Do taxon cycles emerge and if so, do the different models predict the same sort of cycle?

First, we describe the common ecological setting in which the models are compared, then we review the mechanics and assumptions of the three alternative approaches, and then we discuss the very important differences in model behavior. This section draws heavily on Taper and Case (1992), and interested readers may wish to look there for greater detail and numerical examples. In Taper and Case (1992), there are a few changes in the quantitative genetic model from that of Taper and Case (1985) that must be made to put all three approaches in a common ecological setting. For example, Taper and Case (1985) allowed heritability to evolve as well as within-phenotype niche width, while in the papers of Roughgarden and colleagues and of Brown and Vincent these traits were not allowed to evolve. Consequently, for the sake of comparison, these traits were also restricted in the QGR model of Taper and Case (1992).

5.2 Placing competitors in a common ecological setting

For the sake of model comparison, the ecological setting has been drawn essentially intact from Rummel and Roughgarden (1985), who developed it to explain empirical patterns of body size of *Anolis* lizards in the Lesser Antilles (see also Roughgarden 1979). These species compete for resources that vary in type along some continuum (e.g. prey size). A phenotypic character, say body size (z), determines the ability of individuals to take resources of a certain type or size. Based on measurements of the abundance and productivity of these resources in the environment, along with knowledge of the rate of resource used by each phenotype z , one can derive a function representing the carrying capacity that a population of a single phenotype z would have, $K(z)$. A phenotype whose peak consumption rates are on abundant resources will have a higher K than one whose prime resources are less abundant. A phenotype that has higher consumption rates will have a higher K than one that eats the same set of resources but at lower rates. However, the details of consumption are not made explicit but are encapsulated in the $K(z)$ curve and the shape and magnitude of the competition function.

Following Rummel and Roughgarden (1985), $K(z)$ is assumed to be

Gaussian and to be constant and over time and identical from place to place. For convenience, the phenotype axis z is usually scaled so that phenotype 0 has the highest carrying capacity and deviations from 0 lead to reduced carrying capacity. The width of the carrying capacity curve is measured by σ_k . If slight deviations from the central phenotype yield rapid loss in carrying capacity, then σ_k is small. The wider σ_k , the more competitors that can ecologically coexist.

The fitness of individuals of a particular phenotype is the combination of the carrying capacity of that phenotype and the degree of competition experienced by individuals of that phenotype. The fitness function, $W(z)$, used in all the models is based on classical Lotka-Volterra competition.

$$W(z, \mathbf{N}, \mathbf{Z}) = 1 + r - \frac{r}{K(z)} \sum_{i=1}^n \int p_i(x) \alpha(z, x) N_i dx \quad (1)$$

The species in the community are indexed by i (which ranges from 1 to n); N_i is the population size of the i th species, and p_i is the phenotype distribution function for the i th species; r is the intrinsic rate of population growth and is the same for all species and z . Equation 1 is general and holds for all phenotypic positions of all species, even ones presently absent from the community, hence the absence of subscripts for W , z , and x . \mathbf{N} and \mathbf{Z} are the vectors of population size and phenotypic position for the community. The functional form of p_i differs among the three models but in all cases depends on z . Because of the differences in $p_i(z)$, $W(z, \mathbf{N}, \mathbf{Z})$ will differ among the models. It is an intrinsic element of frequency dependent selection that the phenotypic distribution influences the fitness function.

The competition function, $\alpha(z, x; \sigma_\alpha, \beta)$, measures the competition experienced by an individual of phenotype z from an individual of phenotype x . The functional form is derived from the MacArthur and Levins (1967) overlap formula (assuming Gaussian utilization functions; Roughgarden 1979, p. 531), and from a power-law relating consumption of resources to body size (Rummel and Roughgarden 1985). The function has a mean of zero and two other parameters: σ_α is the standard deviation and β is a parameter that measures the degree of asymmetry (explained below).

We plot values of $\alpha(z, x)$ in Fig. 2 using different values of β for the actual functional form used by Rummel and Roughgarden (1985). Their expression allows for competitive asymmetry (i.e. $\alpha(z, x) \neq \alpha(x, z)$) when $\beta \neq 0$. If β is positive then an individual at a higher niche position has a competitive advantage over a species at a lower niche position (see Fig. 2). This might happen because of agonistic dominance of large animals over small, but it is important to realize that these positive asymmetries are expected to arise purely in exploitation competition. All

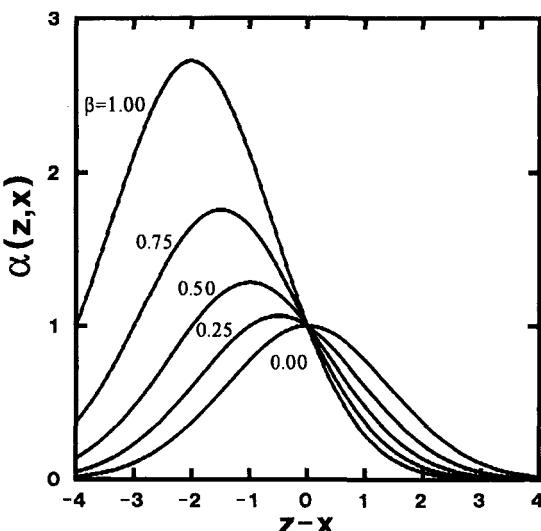


Fig. 2. Asymmetric α function. The curves plot the strength of competition as a function of phenotypic separation for different asymmetry parameters β ranging from 0 to 1. When β is greater than 0, a species with a higher phenotype has a competitive advantage. For all curves $\sigma_\alpha = 1$.

else being equal, as body size increases, $K(z)$ may decrease since fewer animals can be supported for the same level of resource productivity, but since larger animals eat more food per unit time and usually eat a wider range of foods as well, the food supply is more depleted by a large animal than by a small one foraging over the same time period. Thus, higher per capita competitive impacts of large on small than small on large phenotypes will result. As in the well known Lotka-Volterra competition equations, the ultimate outcome of competition depends jointly on relative K 's and α 's as expressed by eqn 1; neither alone can determine species success or phenotypic fitness (Persson 1985).

Extinction occurs when N_i falls below 1. As Rummel and Roughgarden (1985) point out, this scales the population in units of minimum viable population size.

As in Rummel and Roughgarden (1985), faunal build-up is considered a sequence of invasion and coevolution stages. We will first consider the case that species are homogeneous in phenotype. New species are introduced into the community by invasion, and then the community is allowed to coevolve to an apparent equilibrium before the next invasion occurs. Equilibrium is deemed to have been reached when no species has either a proportional change in N_i or an absolute change in phenotype z_i greater than an accuracy constant (10^{-7} for the runs reported here).

The species index i ranges from 1 to the number of species present in the community at a given time. This number will increase by 1 after an invasion and decrease by 1 after an extinction. The points along the phenotype axis at which invasions take place are chosen as the points of maximum mean fitness of an invasion propagule of size 1 (the minimum viable population size). Faunal build-up begins from an empty community with no species present. The first invasion is always at the peak of the carrying capacity curve.

5.3 The CSC approach

Derivation and use of this approach can be found in Roughgarden (1976, 1979, 1983) and Case (1982). Critiques of this derivation and discussion of the implications of the assumptions are given in Brown and Vincent (1987a), Taper (1988) and Abrams (1989). Community population dynamics are given by eqn 2, while the evolutionary dynamics are given by eqn 3.

$$\Delta N_i = [W(z_i) - 1] N_i \quad (2)$$

$$\Delta z_i = (\text{const}) \left. \frac{\partial W(z_i)}{\partial z_i} \right|_{\mathbf{Z}} \mathbf{N}^*(\mathbf{Z}) \quad (3)$$

Here, $\mathbf{N}^*(\mathbf{Z})$ is the vector of population sizes attained at ecological equilibrium given phenotypic position vector \mathbf{Z} . The vertical bar in eqn 3 indicates that the partial derivative to the left of the bar is to be evaluated under the conditions given to the right of the bar. The evaluation of the partial derivative of fitness at ecological equilibrium population sizes ($\mathbf{N}^*(\mathbf{Z})$) in effect separates ecological time from evolutionary time, and says that all ecological changes occur essentially instantaneously compared with evolutionary changes. We think that the distinction between evolutionary and ecological time is artificial, but we have retained the condition so as to represent accurately the CSC model as previously used (Case 1979, 1982; Pacala and Roughgarden 1982; Roughgarden 1983; Rummel and Roughgarden 1983, 1985; Pacala 1988).

In the CSC approach species are assumed to be completely monomorphic, and thus the p_i in eqn 1 is a delta function (i.e. a function whose value is 0 at all points other than z_i where it takes the value 1). We can think of $W(z_i)$ as the mean fitness of a species with niche position z_i , and the partial derivative of the fitness of species i is taken with respect to change in the phenotypic position of species i . The constant, which controls the response to selection and thus the speed of evolution, is proportional to heritability and is set to 1 to match the value used in Rummel and Roughgarden (1985).

The assumption of a monomorphic species is a necessary condition in the derivation of this model. The CSC approach is based on a theorem proved by Roughgarden (1976) that shows that in the absence of intraspecific frequency-dependent selection (interspecific frequency-dependence is allowed), communities will coevolve, with each species shifting its gene frequencies so as to maximize its own population density given the gene frequencies and population densities of all *other* species in the community. The assumption of monomorphic species precludes intraspecific frequency dependence and thus allows the application of this theorem. The contradiction implicit in this assumption has been pointed out several times (Taper 1988; Abrams 1989). A completely monomorphic species is incapable of evolution, but once variation is allowed, evolution does not necessarily result in the maximization of conditional population density. None the less, the critical question is not whether the monomorphy assumption is absolutely true, but whether the predictions of the model are robust to minor violations of the assumption.

5.4 The ESS approach

The development of this formulation can be found in the papers by Brown and Vincent (1987a, b). Roughgarden (1987) provides a critique. As in the CSC approach, population dynamics are determined by eqn 2. Evolutionary dynamics are given, however, by eqn 4.

$$\Delta z_i = (\text{const}) \frac{\partial W(z)}{\partial z} \Big|_{\begin{array}{c} z = z_i \\ \mathbf{N}^*(\mathbf{Z}) \\ \mathbf{Z} \end{array}} \quad (4)$$

Here the derivative controlling the dynamics of species i is the partial derivative of fitness with change in position on the phenotype axis evaluated at the phenotype of the i th species. As with the CSC the fitness derivative in the ESS is also evaluated at the ecological equilibrium.

Brown and Vincent (1987a) define the ESS in terms of equilibrium conditions and do not directly consider dynamics except to point out that they could be followed with eqn 4. The ESS is determined on the basis of three conditions. The first and second conditions are that eqns 2 and 4 are zero for all species, that is, at equilibrium neither the population size nor the phenotype of any species is changing. The third ESS condition is that the community cannot be invaded by a species of any phenotype. This condition is as important as the first two. In general, however, there are no analytical techniques to evaluate this last condition (although special cases may have tractable solutions) and usually this is done by numerically exhausting all viable invasion possibilities.

Comparing the CSC and the ESS approaches, there is a subtle but important difference between the partial derivatives in eqns 3 and 4. In the CSC model, eqn 3, the partial derivative measures the changes in species mean fitness with a shift in mean species position. However, the partial derivative in the ESS model, eqn 4, measures the difference between the fitness of an individual at a position only slightly removed from the species' present mean position and the fitness of an individual at the species' mean position. Equation 4 is assessing the ability of a rare mutant to invade the 'wild type' population. The critical assumption of the ESS model is that species are 'almost' completely monomorphic in the sense that the population contains only the 'wild type' at its equilibrium abundance and the mutant which is phenotypically close to the wild type and vanishingly rare. Evolution takes place by the rapid fixation of mutants. Strictly speaking, either a parthenogenetic or a haploid genetic system is implied. In effect, the ESS approach has a double standard regarding mutations. A new clone that appears infinitesimally close to a parental species is thought of as a mutant of that species on the road to finding the optimal strategy. On the other hand, a new mutant clone that appears measurably far from any existing species is thought of as a new species – a new player in the game, rather than an old player tentatively testing a new strategy. This does not constitute a real difference between the CSC and the ESS paradigms, because the phylogeny of invaders is not explicitly considered in either.

5.5 The QGR approach

This model was developed by Slatkin (1980) and discussed in detail by Taper and Case (1985). Here, for the first time, species are no longer assumed to be either completely or nearly monomorphic. Within a species, phenotypes are assumed have a Gaussian (normal) distribution with part of the phenotypic variance due to additive genetic variance (Falconer 1981). For simplicity the phenotypic standard deviation, σ_z , is taken here as a constant over all species and does not evolve. The population and evolutionary dynamics are given by eqns 5 and 6:

$$\Delta N_i = \int_{-\infty}^{\infty} [W(z) - 1] p_i(z) N_i dz \quad (5)$$

$$\Delta \bar{z}_i = (\text{const}) \left[\frac{\int_{-\infty}^{\infty} (W(z) p_i(z) z dz)}{\int_{-\infty}^{\infty} W(z) p_i(z) dz} - \bar{z}_i \right] \quad (6)$$

The standard name for the constant in QGR is the 'heritability' (h^2). Here, \bar{z}_i is the mean phenotype of the i th species and plays the same role as z_i in the other models. This is the first time we consider mean values of z because it is the first time that we have introduced population-level variation in z . The distribution of phenotypes is given by the function p_i , with $p_i(z)$ being the relative frequency of the i th species at phenotype z . Taper and Case (1992) set the constant (h^2) to 0.9, a value which Rummel and Roughgarden (1985) state is equivalent to a constant of 1 in the CSC model. The QGR model does not separate ecological and evolutionary time.

In the QGR model species are no longer monomorphic. We can think of the total resource utilization (niche width) of the species as being partitioned into two additive components: the average within-phenotype niche width and the between-phenotype niche width. To give a fair test of the monomorphism assumption, the between-phenotype component of the niche width was made quite small, 2 per cent in the QGR model. This means that each z is using a wide range of resources and very different z 's are using roughly similar sets of resources. In the context of these models (where resources are not explicit) the between-phenotype component of niche width can be identified simply with the phenotypic variance ($BPC = \sigma_z^2$), while the within-phenotype component is proportional to the variance of the competition function, which is Gaussian with variance, $WPC = \frac{1}{2}\sigma_\alpha^2$ (see Roughgarden (1979) and Taper and Case (1985) for discussions of the partitioning of niche width). In other words, if σ_α^2 is zero, only identical phenotypes compete. For two different phenotypes, z and x , the resource overlap, and thus competition between them, increases if they are closer together or if their within-phenotype niche widths increase. The latter process is simply mimicked by increasing σ_α^2 , since, all else being equal, the competition between two phenotypes, z and x , increases as σ_α^2 increases. Unlike the situation in Taper and Case (1985), for the purposes of comparison, we will now confine all phenotypes to have the *same* within-phenotype niche width and consider only evolution in mean phenotypic position. Some theory suggests that, at least for weak directional selection, an assumption of constant genetic variance will not lead one far astray (Lande 1979). However, it is by no means clear that this assumption is robust and how (or even in what direction) variances will change with directional selection (Turelli 1988; Turelli and Barton 1990). Empirically, at least under weak selection, the variance in a character usually changes less rapidly than the mean of the character (Sheridan 1988).

This model, too, has its share of assumptions (see Taper and Case 1991 for a fuller discussion). Nonetheless, of the three approaches, we feel that it is the most realistic genetic description with the fewest

constraining assumptions, and is in our opinion the standard by which to judge the success of the other two.

5.6 Model behavior

In what follows we describe qualitatively the similarities and differences among the models. For a more detailed discussion and numerical examples, the reader should refer to Taper and Case (1992).

As predicted by Brown and Vincent (1987a) and Taper (1988), for a single species when the asymmetry parameter β is set to 0, the CSC and the ESS models have exactly the same dynamics and equilibrium. The QGR model also behaves in a very similar fashion.

The next step is to consider the invadability of the single-species communities produced by the three models. When the asymmetry parameter β set to 0, σ_K to 1.6, and σ_α to 1.0, all three models predict that the single species community is invadable and after invasion they all yield a non-invadable two-species equilibrium. If σ_K is increased or σ_α decreased, however, more species may ultimately invade and be sustained in the system after coevolution (see Case 1982; Loeschke 1985; Taper and Case 1985). In all cases with β equal to 0 and σ_z small, the dynamics and equilibria of the CSC and ESS models are identical to each other and very similar to the QGR model.

However, with asymmetry (large phenotypes have a greater impact on the per capita growth of small phenotypes than vice versa), the models differ. The CSC approach predicts a single phenotype still centered under the peak of the $K(z)$ curve. On the other hand, even for a single species, the ESS and QGR models both predict that the species' mean phenotype should be displaced from the peak of the $K(z)$ curve towards larger phenotypes. The lack of displacement in the CSC model is a consequence of ignoring intraspecific frequency-dependence. Imagine that a population exists with z centered under the peak of $K(z)$. Now a mutant appears with a phenotype slightly larger than this central position. It will have a slightly lower K , but its competitive impact on the wild type is severe and larger than the reciprocal effect of the wild type on it. Depending on the magnitude of this discrepancy (i.e. the strength of asymmetry) relative to the increasing rapidity with which K declines with z , it is easy to imagine situations where this mutant is favored. Larger-sized phenotypes will be favored up to a point. Eventually, however, because of the nature of a Gaussian K curve, increasingly larger z 's correspond to rapid drops in $K(z)$, counterbalancing the asymmetric α advantage to larger phenotypes. Ultimately a stable equilibrium niche position is established somewhere to the right of the peak of $K(z)$.

Considering invasion, with mild competitive asymmetry ($\beta = 0.2$), the behavior of the CS model is as described by Rummel and Roughgarden

(1985). The first invader enters at the peak of the resources curve and remains there while it reaches its equilibrium population size. A second invader then enters at the optimal z value above the first species (i.e. this is the position with the highest invasion potential; smaller levels of z than the resident are not able to invade at all). The second and larger species evolves toward the peak of the resource curve, driving the first to progressively lower phenotypic values and eventually to extinction. The community then enters a cycle of invasion and coevolution and extinction involving no more than two species (again this is altered by changing σ_K/σ_α).

For the same parameters, the ESS and the QGR models behave similarly to each other, but radically different from the CSC. We illustrate this process in Fig. 3 using the actual results from the ESS simulations. As stated earlier, the first invader enters at the peak of the resource curve (Fig. 3a) and then, unlike the situation in the CSC, evolves to a larger phenotype where it eventually equilibrates (Fig. 3b). This creates a larger 'invasion window' for invading phenotypes that are below the peak of the K curve than those above (Fig. 3b). After invasion by a small z (Fig. 3c), both species shift to higher phenotypic values. They ultimately equilibrate at positions shifted somewhat to the right of center (Fig. 3d) and a third species invades from below as before (Fig. 3e). Again, all three species now shift upwards until a three-species equilibrium is reached (Fig. 3f). This three-species equilibrium is closed to further invasions by any z , so that it represents the final end-state. Cycles of invasion and extinction are also possible in the ESS and QGR models with higher values of β and/or σ_α .

Although by predicting continual turnover of species through successive colonization and extinction events, all three models support the concept of a taxon cycle (Wilson 1961; Ricklefs and Cox 1972, 1978; Roughgarden 1983), they make dramatically different predictions. The cycle moves in the opposite direction under the CSC model than it does under either of the ESS or QGR models. Specifically, the QGR and ESS models both predict: (1) solitary species will evolve upwards in body size; (2) subsequent invasions will be more successful by smaller species (not larger) species than the resident; and (3) both invader and resident should then evolve larger sizes (not smaller).

Taper and Case (1992) also consider the implications of relaxing the near-monomorphism assumption in the QGR model. Comparisons were made among the models with parameters set so that all models had the same total niche width, rather than the same within-phenotype niche width. (Thus σ_α was increased in the CSC and ESS models to compensate for increases in σ_z). Even under moderate polymorphism (BPC 33 per cent of total niche width) and asymmetry ($\beta = 0.2$), the models behave quite differently. The CSC, as before, enters an invasion/extinction cycle.

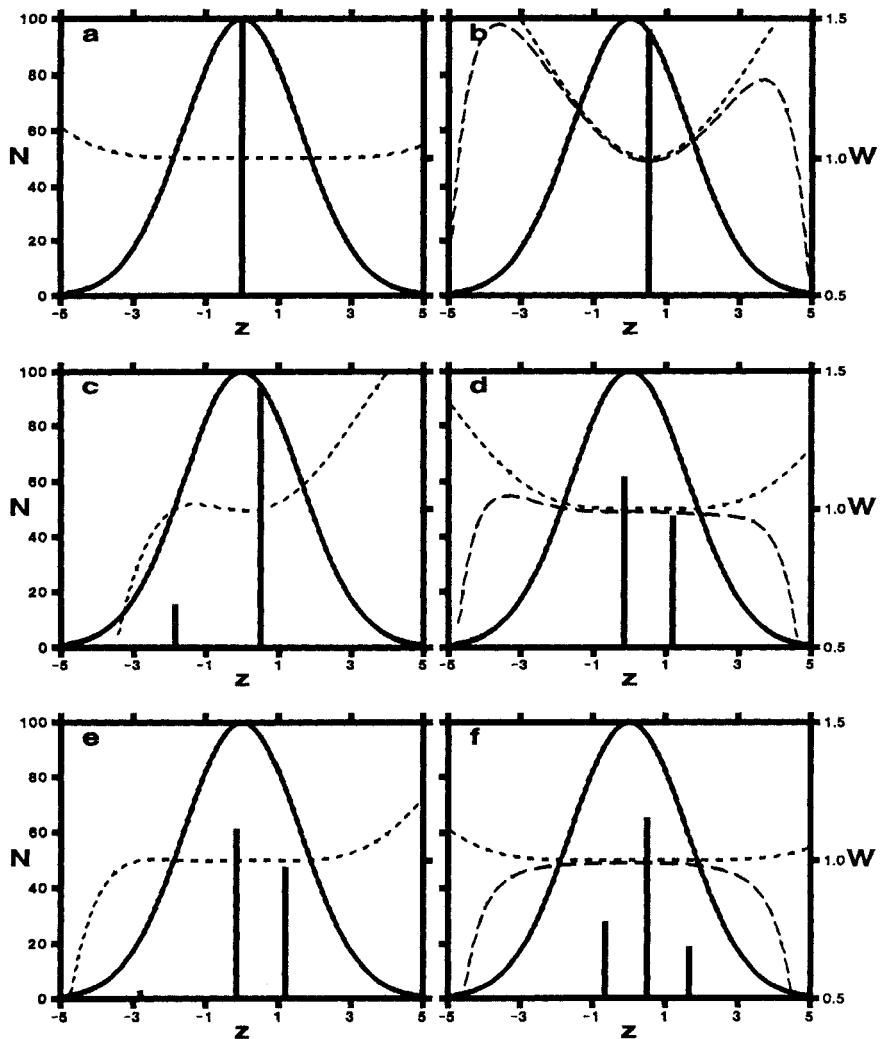


Fig. 3. Sub-figures (a–f) follow the faunal build up on an island from initial invasion to closed equilibrium community. Figures in the left-hand columns (a, c, and e) show the community shortly after invasion. Populations sizes have adjusted to their ecological equilibrium but no evolution has occurred since invasion. The righthand figures depict the communities at evolutionary equilibrium. The solid Gaussian curve represents the carrying capacity at phenotype z , while the solid vertical bars indicate species' phenotypic positions and population sizes. The light dotted line is the fitness function $W(z)$. The slope of this function at a species' phenotypic position dictates the evolution of that species in the immediate future. The light dashed lines show the fitness that an invading population of size 1 and phenotype z would have to have to invade successfully; this fitness must be greater than 1. The ESS model was used in calculating the positions and fitnesses in this figure. The parameters used were $K = 100$, $\sigma_K = 1.6$, $\sigma = 1$, and $\beta = 0.2$.

The ESS and QGR models both reach stable two-species communities. There is, however, a considerable difference between these communities. The displacement is more than twice as large in the QGR model than in the ESS model. Also, the mean phenotypic position of the entire community is closer to the peak of the carrying capacity in the QGR community than in ESS community, and the total population size of the QGR community is greater than that of the ESS community.

Furthermore, as can be seen by comparing the solid line in Fig. 4 to the dashed line in Fig. 3b, the introduction of within-species phenotypic variation (Fig. 4) influences the community invasion surface and reduces the probability that invasion can occur at phenotypes larger than the residents. Only by increasing K in the QGR model, is it possible to have any successful invasions for phenotypes larger than the resident (dashed line in Fig. 4).

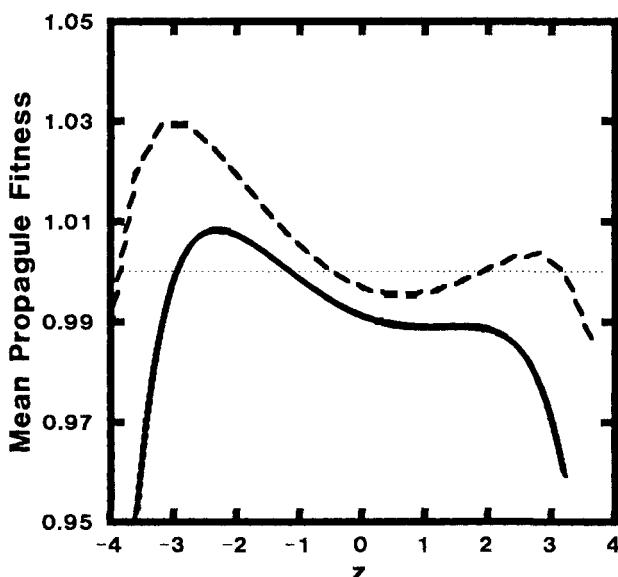


Fig. 4. Single-species community invadability with phenotypic variation. The mean fitness of an invasion propagule of size 1 is plotted against propagule phenotype. A propagule can invade if its mean fitness is greater than 1 (the light horizontal line). The heavy solid line shows the fitness of propagules invading the single species QGR model community with parameters as in Fig. 3 and $\sigma_z = 1$. The heavy dashed line indicates the fitness of propagules invading the one-species community found under the same parameters except that K has been doubled to 200.

5.7 Conclusions from modeling character displacement

5.7.1 *The CSC is not a viable technique for modeling character displacement*

The first conclusion is that, in general, the behavior of the CSC model is not consistent with either the ESS or QGR approaches. Since competitive asymmetries are probably the norm, rather than the exception (see section 5.1), maximization of equilibrium densities will not, in general, accurately predict the end-state of evolutionary trajectories. By disallowing intraspecific frequency-dependence, CSC disallows the core of the process it hopes to explain. If interspecific competition is mediated by phenotypic similarity then intraspecific competition must be also; yet if it is, then intraspecific frequency-dependence is inevitable and the maximization principle is no longer valid. This problem with CSC casts doubt on some of the conclusions in studies by Case (1979), Schluter and Grant (1984b) and Rummel and Roughgarden (1985).

The bias in the CSC that results under asymmetric competition can vary from slight to extreme, depending on the shape of the resource curve and the strength of asymmetry. As discussed in Taper (1988), the entire community will shift in the direction of asymmetry. That is, all species will be larger than predicted by the CSC if larger individuals are favored in competition.

5.7.2 *Limitations of the ESS*

The ESS and QGR approaches are more similar. When the between-phenotype component (BPC) of total niche width is low, the ESS model agrees closely with the QGR approach. The ESS model has a decided advantage in computational speed. Of course, the point of theory is to try to reach realistic conclusions about nature: if two models agree but one is simpler and faster, then more parameter space can be explored enhancing the range of predictions and stimulating more varied comparisons with nature. Unfortunately, as the BPC increases, so do the discrepancies between the models, suggesting caution in using the short-cut ESS approach. In nature, the BPC is typically greater than 30 per cent of the total niche width (Case 1982); in this range the differences between the two models are quantitatively significant, although the behavior of the models can still agree qualitatively.

Taper and Case (1992) show how the advantages of the ESS and QGR models may be welded into a new approach, which we term the 'quantitative genetic optimization,' (QGO). The ESS gains great conceptual appeal and computational alacrity by solving directly for the equilibrium. Because of the near-monomorphism assumption, it sacrifices realism in the fitness surface that it is optimizing over (see section 5.2). Recursive QGR models, albeit more realistic than ESS models, grind

agonizingly slowly towards their equilibria. But, if they have equilibria, why not solve for them directly? This approach has been successfully taken by Taper and Case (1992). The QGO solutions were arrived at by setting the left-hand side of eqns 5 and 6 to zero for all species and solving the resulting set of equations numerically for the vectors \mathbf{Z}^* and \mathbf{N}^* . Like ESS analysis, QGO does not record dynamics, nor is there any guarantee that the resulting solution is stable or can be reached from arbitrary initial conditions. Thus, a judicious combination of QGO and QGR analysis will probably provide the most insight into a given problem.

5.7.3 *On the nature of taxon cycles*

In these models, for a species to invade a community, it must be different from the residents. If the community has coevolved under positive asymmetric competition, then the model makes two predictions: (1) a solitary species will evolve larger body size, and (2) subsequent invasions are only likely for invaders smaller than the residents. It is still unclear whether these predictions are generally supported in nature. In the real world species are not necessarily confined to only invade at the most favorable invasion point. Figure 3b indicates that the single-species community is invadable almost anywhere, even though the most favorable invasion points are below the resident's size. Therefore evolutionary parallel shifts (at least in the two-species case) could occur either towards increasing or decreasing size. Although the evidence is still somewhat equivocal, phylogenetic analysis indicates that parallel shifts have probably occurred among the anoles of the Lesser Antilles and that the shift has been towards smaller size (Losos 1991; Roughgarden 1991).

Lack (1971) compared body size evolution in the largely insular passerine bird family Zosteropidae. Distant islands typically may have from one to three species. Where multiple species coexist (e.g. in the Carolines or Norfolk of the South Pacific), they are often different in body size. Taxonomic evidence suggests that the largest species arrived on the island first, since these are often endemic at the species if not generic level, while the smaller sympatric species are typically endemic only at the species or subspecies level, if at all. Remote islands typically support only a single species that is somewhat larger than the group average. 'This is a widespread trend in island songbirds, and the usual explanation is that it enables them to take a wider range of foods' (Lack 1971, p. 216).

To get a taxon cycle, the end state of coevolution must be vulnerable for further invasions. Strong asymmetric competition can produce this necessary feature and ecological evidence suggests that asymmetries are probably the rule rather than the exception. However, this alone is not sufficient for a taxon cycle to be realized in nature. A taxon cycle cannot be expected to support itself indefinitely in an isolated archipelago

comprised of similar islands (i.e. similar distributions of resources). Being far from a mainland colonization source, colonists are only available from other nearby islands. The number of species ultimately coexisting in a community depends on the breadth of resources available and the within-phenotype niche widths of the consumers. In any case, virtually all potential colonists will, like the residents, ultimately evolve large phenotypes. The result will be an absence of small species to invade, prematurely terminating the cycle. An empty island can of course be invaded, but then this one-species community will be effectively closed to invasion since only large species will be available to colonize. Thus, instead of a continually cycling community, an island in such a distant archipelago will reach a pseudo-stable equilibrium after only one cycle; it is not invadable by any of the available large colonists but could be invaded if smaller colonists entered.

Thus a taxon cycle may not be supportable in 'very' isolated archipelagos, unless interisland differences in resource distributions are large such that on some islands small size is favored while on other islands large size is favored. 'Very' near shore islands will probably not support a taxon cycle either because gene flow from mainland ancestors will retard local evolutionary divergence. Thus we suspect that taxon cycles will most often manifest themselves at intermediate levels of isolation. What constitutes an 'intermediate' level of isolation will depend on geography and on the vagility of the taxa under consideration.

5.8 Caveats and suggestions for further theoretical studies

Our focus in section 5 has been directed toward comparing modeling approaches and answering questions about model robustness. To facilitate this comparison, we used the same simplified ecological setting to analyze each model. We believe that this gives a fair comparison of the models, but that it limits the usefulness of specific predictions. We have only hinted at questions of the impact of asymmetric competition on character displacement and of the differences between invasion- and coevolution-structured communities. This is not because we feel these are not important questions – they are; however, we do not think that the ecological framework used in these models has sufficient reality to yield meaningful applications to real life situations.

One major difficulty is the use of a fixed carrying capacity curve. It has been pointed out and demonstrated repeatedly that such a formalization does not give results consistent with models with explicit resource dynamics, either ecologically or evolutionarily (Abrams 1975, 1980*a, b*, 1983*a, b*; Getz 1984; Taper and Case 1985; Glasser 1988; Glasser and Price 1988). The mechanistic details of how resources are harvested and how environmental conditions influence consumption and reproductive

rates will certainly be different for different taxa and these effects will influence the course of coevolution. Naganuma and Roughgarden (1990) have constructed a very mechanistic model for optimal body size in *Anolis* lizards; extensions of this model to *Anolis* coevolution would probably be more fruitful than applications of generic resource/consumer dynamical equations as in Taper and Case (1985).

The invasion scenario used is also worrisome. It seems unlikely that invaders will always have the optimum phenotype for invasion. What if colonists are not equally available at all phenotypes? Then it is possible, as suggested by Roughgarden *et al.* (1987) and Roughgarden and Pacala (1989), that the nature and supply of colonists may profoundly affect community structure. This is particularly likely to be true if invaders come from other islands in an archipelago rather than from a more diverse mainland pool.

Another problem with the invasion scenario is the artificial separation of evolution from invasion. Coevolution in frequency-dependent systems seems to proceed with a very wide range of rates. Evolution can be very rapid for a brief time following a change in community structure, but then the rate of evolution slows substantially (see Taper and Case 1992). Thus species can be far from equilibrium, but yet be evolving very slowly. This behavior is probably not an artifact of the fixed and Gaussian resource curve or other features of the simplified ecological model, because it has been seen in models with explicit resource dynamics as well (Taper and Case, unpublished results). This slow approach to equilibrium raises questions about the relative invadability of equilibrium and non-equilibrium communities.

6. WHERE DO WE GO FROM HERE?

We end by noting that, far from being a closed subject, the field is wide open. Empirical evidence is rapidly mounting that character displacement can affect community structure in important ways. Yet at this time very little is known about the mechanistic nature of trade-offs affecting fitness in alternative phenotypes, or of the effects of asymmetric competition, overlapping generations (Case and Taper 1986), and ontogeny of phenotypes (Werner 1986) on the coevolution of community structure. None the less, much confusion about appropriate modeling techniques has been cleared up, and powerful new techniques are now available. We expect that in the next decade the thrust of the research in this field will shift from trying to demonstrate or disprove the existence of character displacement to investigating more quantitative and mechanistic questions. How often and under what conditions do we expect character displacement to occur? How much displacement should we expect, in what direction,

and what effects does character displacement have on community properties such as invadability, resilience, or the degree of overall resource utilization?

The biggest barrier to making the models predictive for specific situations is the large number of parameters in these models that are nearly impossible to measure. The quantitative genetic models require knowledge of the competitive impact of specific phenotypes on one another and the heritability of those phenotypes. Yet in practice it is rarely possible to find identical sets of phenotypes so that competitive impacts can be measured and replicated, since in sexual species each individual is genetically unique. One experimental system with great potential for measuring specific interactions among phenotypes is offered by clonally reproducing species (Vrijenhoek 1978, 1984; Ellstrand and Antonovics 1984, 1985; Case 1990; Weeks and Sassman 1990). Molecular genetic techniques (allozymes, genetic fingerprints, mt-DNA restriction fragment maps) allow the resolution of genetic clones. One can then measure phenotypic variability within a clone; measure ecological differences correlated with genetic differences; and experimentally combine different clonal combinations to measure specific competitive phenotypic interactions. In this way it should be possible to test the underpinnings of evolutionary niche theory as applied to the problem of character displacement. Is it true that greater genetic variability leads to greater ecological variability and wider species niche widths? Do larger phenotypes have a competitive advantage? The tantalizing glimpses of answers to these questions seen in the studies described above warrant vigorous pursuit, both theoretical and empirical.

ACKNOWLEDGMENTS

This work has been partially supported by NSF grant # BSR-8807792 to Dr James H. Brown, NSF grant # BSR 8805969 to T.J.C., and a Japanese Science and Technology Agency grant to M.L.T. We are indebted in this work to many colleagues, who over the years have influenced our thinking on the topics we have discussed. We can recall specific discussions or correspondence with Peter Abrams, Jim Brown, Joel Brown, Martin Cody, Ann Evans, Koichi Fujii, Douglas Futuyma, Daniel Goodman, Yoh Iwasa, Jonathan Losos, Roger Mitchell, Trevor Price, Jonathan Roughgarden, Dolph Schlüter, Terry Schwaner, Nanoka Shigesada, Montgomery Slatkin, and Yukihiko Toquenaga. Jim Brown, Doug Futuyma, Ed Heske, Douglas Kelt, Pablo Marquet, Trevor Price, and Dolph Schlüter critically read an early version of this paper. We also thank Drs Fujii and Hokyo for providing aid to M.L.T. while in Japan.

REFERENCES

- Abrams, P. A. (1975). Limiting similarity and the form of the competition coefficient. *Theoret. Popul. Biol.* **8**, 356–75.
- (1980a). Are competition coefficients constant? Inductive vs. deductive approaches. *Am. Nat.* **116**, 730–35.
- (1980b). Some comments on measuring niche overlap. *Ecology* **61**, 44–9.
- (1983a). Arguments in favor of higher order interactions. *Am. Nat.* **121**, 887–91.
- (1983b). The theory of limiting similarity. *Ann. Rev. Ecol. Syst.* **14**, 359–76.
- (1986). Character displacement and niche shift analyzed using consumer-resource models of competition. *Theoret. Popul. Biol.* **29**, 107–60.
- (1987a). Alternative models of character displacement: I. Displacement when there is competition for nutritionally nonsubstitutable resources. *Evolution* **41**, 650–61.
- (1987b). Alternative models of character displacement and niche shift. 2. Displacement when there is competition for a single resource. *Am. Nat.* **130**, 271–82.
- (1989). The importance of intraspecific frequency-dependent selection in modelling competitive coevolution. *Evol. Ecology* **3**, 215–20.
- Avidov, Z., Applebaum, S. W. and Berlinger, M. J. (1965). Aspects of host specificity in the Bruchidae. II. Ovipositional preference and behavior of *Callosobruchus chinensis*. *Ent. Exp. Appl.* **8**, 96–106.
- Barlow, J. C. (1973). Status of the North American population of the European tree sparrow. *Ornithol. Monog.* **14**, 10–23.
- Barnett, B. & Schwaner, T. D. (1985). Growth in captive born tiger snakes (*Notechis ater serventyi*) from Chappel Island: Implications for field and laboratory studies. *Trans. Royal Soc. S. Austr.* **109**, 31–6.
- Benkman, C. W. (1989). On the evolution and ecology of island populations of crossbills. *Evolution* **43**, 1324–30.
- Bock, W. J. (1970). Microevolutionary sequences as a fundamental concept in macroevolutionary models. *Evolution* **24**, 704–22.
- Boecklen, W. J. and NeSmith, C. (1985). Hutchinsonian ratios and log-normal distributions. *Evolution* **39**, 695–8.
- Bossert, W. H. (1963). Simulation of character displacement in animals. PhD thesis, Harvard University, Cambridge, MA.
- Bowers, M. A. and Brown, J. H. (1982). Body size and coexistence in desert rodents: chance or community structure? *Ecology* **63**, 391–400.
- Bowman, R. I. (1961). Morphological differentiation and adaptation in the Galapagos finches. *Univ. Calif. Publ. Zool.* **58**, 1–302.
- Brandl, R. and Top, W. (1985). Size structure of *Pterostichus* spp. (Carabidae): Aspects of competition. *Oikos* **44**, 234–8.
- Brooks, J. L. and Dodson, S. I. (1965). Predation, body size and composition of plankton. *Science* **150**, 28–35.
- Brown, J. H. and Nicoletto, P. F. (1991). Spatial scaling of species composition: Body masses of North American land mammals; *Am. Nat.*, in press.
- Brown, J. H., Davidson, D. W., Munger, J. C. and Inouye, R. S. (1986).

- Experimental community ecology: The desert granivore system. In *Community ecology* (ed. J. Diamond and T. J. Case), pp. 41–62. Harper & Row, New York.
- Brown, J. H. and Kurzius, M. A. (1987). Composition of desert rodent faunas: combinations of coexisting species. *Ann. Zool. Fennici* **24**, 227–37.
- Brown, J. S. and Vincent, T. L. (1987a). Coevolution as an evolutionary game. *Evolution* **41**, 66–79.
- (1987b). A theory for the evolutionary game. *Theoret. Popul. Biol.* **31**, 140–66.
- Brown, W. L. Jr. and Wilson, E. O. (1956). Character displacement. *Syst. Zool.* **5**, 49–69.
- Bryant, P. S. and Turner, C. R. (1972). Rapid evolution of competitive ability in larval mixtures of the housefly. *Evolution* **26**, 161–70.
- Butlin, R. (1987). Speciation by reinforcement. *Trends in Ecol. Evol.* **2**, 8–13.
- Carlquist, S. (1965). *Island life: a natural history of the islands of the world*. Natural History Press, Garden City, New York.
- Case, T. J. (1978). A general explanation for insular body size trends in terrestrial vertebrates. *Ecology* **59**, 1–18.
- (1979). Character displacement and coevolution in some *Cnemidophorus* lizards. *Fortschr. Zool.* **25**, 235–82.
- (1982). Coevolution in resource-limited competition communities. *Theoret. Popul. Biol.* **21**, 69–91.
- (1983). The reptiles: ecology. In *Island biogeography in the Sea of Cortez* (ed. T. J. Case and M. L. Cody), pp. 159–209. University of California Press, Berkeley, CA.
- (1984). Niche overlap and resource weighting terms. *Am. Nat.* **124**, 604–8.
- (1990). Patterns of coexistence in sexual and asexual species of *Cnemidophorus* lizards. *Oecologia* **83**, 220–7.
- and Schwaner, T. D. (1992). Insular body size differences in Australian varanid lizards. *Oecologia*, in press.
- and Sidell, R. (1983). Pattern and chance in the structure of model and natural communities. *Evolution* **37**, 832–49.
- and Taper, M. L. (1986). On the coexistence and coevolution of asexual and sexual competitors. *Evolution* **40**, 366–87.
- , Faaborg, J. and Sidell, R. (1983). The role of body size in the assembly of West Indian bird communities. *Evolution* **37**, 1062–74.
- Cody, M. L. (1969). Convergent characteristics in sympatric populations: a possible relation to interspecific territoriality. *Condor* **71**, 222–39.
- (1973). Character convergence. *Ann. Rev. Ecol. Syst.* **4**, 189–211.
- (1974). *Competition and the structure of bird communities*. Princeton University Press, Princeton, NJ.
- Colwell, R. K. and Winkler, D. W. (1984). A null model for null models in biogeography. In *Ecological communities: Conceptual issues and the evidence* (ed. D. R. Strong, D. Simberloff, L. G. Abele and A. B. Thistle), pp. 344–59. Princeton University Press, Princeton, NJ.
- Conant, S. (1988). Geographic variation in the Laysan finch (*Telespyza cantans*). *Evol. Ecol.* **2**, 270–82.
- Cowie, R. H. and Jones, J. S. (1987). Ecological interactions between *Cepaea*

- nemoralis* and *Cepaea hortensis*: competition, invasion, but no niche displacement. *Funct. Ecol.* **1**, 91–7.
- Davidson, D. W. (1977). Species diversity and community organization in desert seed-eating ants. *Ecology* **58**, 711–24.
- (1978). Size variability in the worker caste of a social insect (*Veromessor pergandei* Mayr) as a function of the competitive environment. *Am. Nat.* **112**, 523–32.
- Dawson, P. S. (1972). Evolution in mixed populations of *Tribolium*. *Evolution* **26**, 357–65.
- Dayan, D., Simberloff, D., Tchernov, E. and Yom-Tov, Y. (1989). Inter and intraspecific character displacement in mustelids. *Ecology* **70**, 1526–39.
- (1990). Feline canines: community-wide character displacement among the small cats of Israel. *Am. Nat.* **136**, 39–60.
- Diamond, J. (1986). Evolution of ecological segregation in the New Guinea montane avifauna. In *Community ecology* (ed. J. Diamond and T. J. Case), pp. 98–125. Harper & Row, New York.
- , Pimm, S. L., Gilpin, M. E. and LeCroy, M. (1989). Rapid evolution of character displacement in myzomelid honeyeaters. *Am. Nat.* **134**, 675–708.
- Eadie, J. M., Broekhoven, L. and Colgan, P. (1987). Size ratios and artifacts: Hutchinson's rule revisited. *Am. Nat.* **129**, 1–17.
- Eldredge, N. (1974). Character displacement in evolutionary time. *Amer. Zool.* **14**, 1083–97.
- Ellstrand, N. C. and Antonovics, J. (1984). Experimental studies of the evolutionary significance of sexual reproduction. I. A test of the frequency-dependent selection hypothesis. *Evolution* **38**, 103–15.
- (1985). Experimental studies of the evolutionary significance of sexual reproduction. II. A test of the density dependent selection hypothesis. *Evolution* **39**, 657–66.
- Falconer, D. S. (1981). *Introduction to quantitative genetics*, 2nd edn. Longman, London.
- Fenchel, T. (1975). Character displacement and coexistence in mud snails (*Hydrobiidae*). *Oecologia* **20**, 19–32.
- Fryer, G. and Iles, T. D. (1972). *The cichlid fishes of the Great Lakes of Africa*. Oliver and Boyd, Edinburgh.
- Futuyma, D. J. (1970). Variation in genetic response to interspecific competition in laboratory populations of *Drosophila*. *Am. Nat.* **104**, 239–52.
- and Moreno, G. (1988). The evolution of ecological specialization. *Ann. Rev. Ecol. Syst.* **19**, 207–34.
- and Slatkin, M. (eds) (1983). *Coevolution*. Sinauer Associates, Sunderland, MA.
- Getz, W. M. (1984). Population dynamics: a per capita resource approach. *J. Theor. Biol.* **108**, 623–43.
- Gilbert, L. E. (1983). Coevolution and mimicry. In *Coevolution* (ed. D. J. Futuyma and M. Slatkin), pp. 263–81. Sinauer Associates, Sunderland, MA.
- Glasser, J. W. (1988). Interference (and facilitation) among species that exploit alternative resources. *Ecolog. Modelling* **40**, 111–29.
- and Price, H. J. (1988). Evaluating expectations deduced from explicit hypotheses about mechanisms of competition. *Oikos* **51**, 57–70.

- Gleeson, S. K. (1981). Character displacement in flowering phenologies. *Oecologia* **51**, 294–5.
- Grant, P. (1972). Convergent and divergent character displacement. *Biol. J. Linnean Soc.* **4**, 39–68.
- (1975). The classical case of character displacement. *Evol. Biol.* **8**, 237–337.
- (1986). *Ecology and evolution of Darwin's finches*. Princeton University Press, Princeton, NJ.
- and Schlüter, D. (1984). Interspecific competition inferred from patterns of guild structure. In *Ecological communities: Conceptual issues and the evidence* (ed. D. R. Strong, D. Simberloff, L. G. Abele, and A. B. Thistle), pp. 201–33. Princeton University Press, Princeton, NJ.
- Greenwood, P. H. (1981). *The Haplochromine fishes of the East African Lakes*. Cornell University Press, Ithaca, New York.
- Hutchinson, G. E. (1959). Homage to Santa Rosalia, or why are there so many kinds of animals? *Am. Nat.* **93**, 145–59.
- Jaenike, J. (1990). Host specialization in phytophagous insects. *Ann. Rev. Ecol. Syst.* **21**, 243–74.
- Kellogg, D. E. (1980). Character displacement and phyletic change in the evolution of the radiolarian subfamily Artiscinae. *Micropaleontology* **26**, 196–210.
- Kiltie, R. A. (1984). Size ratios among sympatric neotropical cats. *Oecologia* **61**, 411–6.
- Lack, D. (1944). Ecological aspects of species formation in passerine birds. *Ibis* **86**, 260–86.
- (1947). *Darwin's finches*. Cambridge University Press, Cambridge.
- (1971). *Ecological isolation in birds*. Harvard University Press, Cambridge, MA.
- Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* **33**, 402–16.
- Lawlor, L. R. and Maynard Smith, J. (1976). The coevolution and stability of competing species. *Am. Nat.* **110**, 79–99.
- Lawton, J. H. and Hassell, M. P. (1981). Asymmetrical competition in insects. *Nature* **289**, 793–5.
- Lister, A. M. (1989). Rapid dwarfing of red deer on Jersey in the last interglacial. *Nature* **342**, 539–42.
- Littlejohn, M. J. (1965). Premating isolation in the *Hyla ewingi* complex (Anura: Hylidae). *Evolution* **19**, 234–343.
- Loeschke, V. (1985). Coevolution and invasion in competitive guilds. *Am. Nat.* **126**, 505–20.
- Losos, J. B. (1990). A phylogenetic analysis of character displacement in Caribbean *Anolis* lizards. *Evolution* **44**, 558–69.
- (1991). A critical comparison of the taxon cycle and character displacement models for Lesser Antilles *Anolis*. *Copeia*, in press.
- , Naeem, S. and Colwell, R. K. (1989). Hutchinsonian ratios and statistical power. *Evolution* **43**, 1820–6.
- MacArthur, R. H. and Levins, R. (1967). The limiting similarity, convergence, and divergence of coexisting species. *Am. Nat.* **101**, 377–85.

- Maiorana, V. C. (1978). An explanation of ecological and developmental constants. *Nature* **273**, 375–7.
- Malmquist, M. (1985). Character displacement and biogeography of the pygmy shrew in Northern Europe. *Ecology* **66**, 372–7.
- Martin, M. M. and Harding, J. (1981). Evidence for the evolution of competitive ability between two species of plants. *Evolution* **35**, 975–87.
- Martin, T. E. (1988a). On the advantages of being different: nest predation and the coexistence of bird species. *Proc. Nat. Acad. Sci. USA* **85**, 2196–9.
- (1988b). Processes organizing open-nesting bird assemblages: competition or nest predation. *Evolutionary Ecol.* **2**, 37–50.
- Maurer, B. A., Brown, J. H. and Rusler, R. D. (1992). The micro and macro in body size evolution. *Evolution*, in press.
- Milligan, B. G. (1985). Evolutionary divergence and character displacement in two phenotypically variable competing species. *Evolution* **39**, 1207–22.
- Mitchell, R. (1975). The evolution of oviposition tactics in the bean weevil, *Callosobruchus maculatus*. *Ecology* **56**, 696–702.
- Moynihan, M. (1968). Social mimicry: character convergence versus character displacement. *Evolution* **22**, 315–31.
- Murray, K. G., Feinsinger, P., Busby, W. H., Linhart, Y. B., Beach, J. H. and Kinsman, S. (1987). Evaluation of character displacement among plants in two tropical pollination guilds. *Ecology* **68**, 1283–93.
- Naganuma, K. H. and Roughgarden, J. D. (1990). Optimal body size in Lesser Antillean *Anolis* lizards – a mechanistic approach. *Ecol. Monog.* **60**, 239–56.
- Nitecki, M. H. (ed.) (1983). *Coevolution*. University of Chicago Press, Chicago, IL.
- Pacala, S. W. (1988). Competitive equivalence: the coevolutionary consequences of sedentary habit. *Am. Nat.* **132**, 576–93.
- and Roughgarden, J. (1982). The evolution of resource partitioning in a multidimensional resource space. *Theoret. Pop. Biol.* **22**, 127–45.
- Park, T. and Lloyd, M. (1955). Natural selection and the outcome of competition. *Am. Nat.* **89**, 235–40.
- Persson, L. (1985). Asymmetrical competition: are larger animals competitively superior? *Am. Nat.* **126**, 261–6.
- Pimm, S. L. (1990). The decline of the Newfoundland crossbill. *Trends Ecol. Evol.* **5**, 350–1.
- Pregill, G. K. (1986). Body size of insular lizards: A pattern of Holocene dwarfism. *Evolution* **40**, 997–1008.
- Raikow, R. J. (1977). The origin and evolution of the Hawaiian honeycreepers (*Drepanididae*). *Living Bird* **15**, 95–117.
- Ricklefs, R. E. and Cox, G. W. (1972). Taxon cycles in the West Indian avifauna. *Am. Nat.* **106**, 195–219.
- (1978). Stage of taxon cycle, habitat distribution, and population density in the avifauna of the West Indies. *Am. Nat.* **112**, 875–95.
- Ricklefs, R. E. and O'Rourke, K. (1975). Aspect diversity in moths: A temperate-tropical comparison. *Evolution* **29**, 313–24.
- Robinson, A. C., Carty, P. D., Mooney, P. A. and Rudduck, P. M. (1986). *South Australia's offshore islands*. Australian National Parks and Wildlife Service Special Publication.

- Roth, V. L. (1981). Constancy in the size ratios of sympatric species. *Am. Nat.* **118**, 394–404.
- Roughgarden, J. (1972). Evolution of niche width. *Am. Nat.* **106**, 683–718.
- (1976). Resource partitioning among competing species – a coevolutionary approach. *Theoret. Popul. Biol.* **9**, 388–424.
- (1979). *Theory of population genetics and evolutionary ecology: An introduction*. Macmillan, New York.
- (1983). Coevolution between competitors. In *Coevolution* (ed. D. J. Futuyma and M. Slatkin), pp. 383–403. Sinauer, Sunderland, MA.
- (1987). Community coevolution: a comment. *Evolution* **41**, 1130–4.
- (1991). Character displacement versus taxon loop. *Copeia*, in press.
- Gaines, S. D. and Pacala, S. W. (1987). Supply side ecology: the role of physical transport processes. In *Organization of communities: Past and present* (ed. P. Giller and J. Gee), pp. 459–86. Blackwell Scientific, London.
- and Pacala, S. (1989). Taxon cycle among *Anolis* lizard populations: review of the evidence. In *Speciation and its consequences* (ed. D. Otte and J.A. Endler), pp. 403–32. Sinauer, Sunderland, MA.
- Rummel, J. D. and Roughgarden, J. (1983). Some differences between invasion-structured and coevolution-structured competitive communities: a preliminary theoretical analysis. *Oikos* **41**, 477–86.
- (1985). A theory of faunal build-up for competition communities. *Evolution* **39**, 1009–33.
- Schindel, D. E. and Gould, S. J. (1977). Biological interaction between fossil species: character displacement in Bermudian land snails. *Paleobiology* **3**, 259–69.
- Schlüter, D. (1986). Character displacement between distantly related taxa? Finches and bees in the Galapagos. *Am. Nat.* **127**, 95–102.
- (1988). Character displacement and the adaptive divergence of finches on islands and continents. *Am. Nat.* **131**, 799–824.
- and Grant, P. R. (1984a). The distribution of *Geospiza difficilis* in relation to *G. fuliginosa* in the Galapagos Islands: tests of three hypotheses. *Evolution* **36**, 1213–26.
- and Grant, P. R. (1984b). Determinants of morphological patterns in communities of Darwin's finches. *Am. Nat.* **123**, 175–96.
- and McPhail, J. D. (1992). Ecological character displacement and speciation in sticklebacks. *Am. Nat.* (in press).
- Schoener, T. W. (1965). The evolution of bill size differences among sympatric congeneric species of birds. *Evolution* **19**, 189–213.
- (1983). Field experiments on interspecific competition. *Am. Nat.* **122**, 240–85.
- (1984). Size differences among sympatric bird-eating hawks: a worldwide survey. In *Ecological communities: Conceptual issues and the evidence* (ed. D. R. Strong, D. Simberloff, L. G. Abele and A. B. Thistle), pp. 254–81. Princeton University Press, Princeton, NJ.
- Schwaner, T. D. (1985). Population structure of black tiger snakes, *Notechis ater niger*, on offshore islands of South Australia. In *Biology of Australian frogs and reptiles* (ed. G. Grigg, R. Shine and H. Ehrmann), pp. 35–46. Surrey Beatty and Sons, Sydney.
- (1990). Geographic variation in scale and skeletal anomalies of tiger snakes

- (Elapidae: *Notechis scutatus-ater* complex) in Southern Australia. *Copeia*, 1138–43.
- and Sarre, S. D. (1988). Body size of tiger snakes in Southern Australia, with particular reference to *Notechis ater serventyi*. *J. Herpetology* **22**, 24–33.
- and Sarre, S. D. (1990). Body size and sexual dimorphism in mainland and island tiger snakes. *J. Herpetology* **24**, 320–23.
- Seaton, A. P. and Antonovics, J. (1967). Population interrelationships. I. Evolution of mixtures of *Drosophila* mutants. *Heredity* **22**, 19–33.
- Sheridan, A. K. (1988). Agreement between estimated and realized genetic parameters. *Ann. Breeding Abst. No. 11*.
- Shine, R. (1987). Ecological comparisons of island and mainland populations of Australian tiger snakes. *Herpetologica* **43**, 233–40.
- Simberloff, D. (1983). Sizes of coexisting species. In *Coevolution* (ed. D. J. Futuyma and M. Slatkin), pp. 404–430. Sinauer, Sunderland, MA.
- and Boecklen, W. (1981). Santa Rosalia reconsidered: size ratios and competition. *Evolution* **35**, 1206–28.
- Slatkin, M. (1980). Ecological character displacement. *Ecology* **61**, 163–77.
- and Maynard Smith, J. (1979). Models of coevolution. *Quart. Rev. Biol.* **54**, 233–63.
- Smith, R. H. and Lessells, C. M. (1985). Oviposition, ovicide and larval competition in granivorous insects. In *Behavioral Ecology: Ecological consequences of adaptive behavior* (ed. R. M. Sibly and R.H. Smith), pp. 423–47. Blackwell, Oxford.
- Sokal, R. R., Bryant, H. and Wool, D. (1970). Selection for changes in genetic facilitation: negative results in *Tribolium* and *Musca*. *Heredity* **25**, 299–306.
- and Rohlf, F. J. (1981). *Biometry*. W. H. Freeman, San Francisco, CA.
- Spencer, H. G., McArdle, B. H. and Lambert, D. M. (1986). A theoretical investigation of speciation by reinforcement. *Am. Nat.* **128**, 241–62.
- Spiller, D. A. (1986). Consumptive-competition coefficients: an experimental analysis with spiders. *Am. Nat.* **127**, 604–14.
- St. Louis, V. L. and Barlow, J. C. (1988). Genetic differentiation among ancestral and introduced populations of the Eurasian tree sparrow (*Passer montanus*). *Evolution* **42**, 266–76.
- Strong, D. R., Szyska, L. A. and Simberloff, D. (1979). Tests of community-wide character displacement against null hypotheses. *Evolution* **33**, 897–913.
- Sulzbach, D. S. (1980). Selection for competitive ability: negative results in *Drosophila*. *Evolution* **34**, 431–6.
- and Emlen, J. M. (1979). Evolution of competitive ability in mixtures of *Drosophila melanogaster*: Populations with an initial asymmetry. *Evolution* **33**, 1138–49.
- Taper, M. L. (1988). The coevolution of resource competition: appropriate and inappropriate models of character displacement. *Bull. Popul. Ecol. Soc.* **44**, 45–53.
- (1990). Experimental character displacement in the adzuki bean weevil, *Callosobruchus chinensis*. In *Bruchids and legumes: Economics, ecology and coevolution* (ed. K. Fujii), pp. 289–301. Kluwer Academic Publishers, Netherlands.

- and Case, T. J. (1985). Quantitative genetic models for the coevolution of character displacement. *Ecology* **66**, 355–71.
- and Case, T. J. (1992). Models of character displacement and the theoretical robustness of taxon cycles. *Evolution* **46**, 317–33.
- Thompson, J. N. (1982). *Interaction and coevolution*. Wiley-Interscience, New York.
- Tilman, D. (1988). *Dynamics and structure of plant communities*. Princeton University Press, Princeton, NJ.
- Toft, C. A. and Shea, P. J. (1983). Detecting community-wide patterns: Estimating power strengthens statistical inference. *Am. Nat.* **122**, 618–25.
- Toquenaga, Y. and Fujii, K. (1991). Contest and scramble competition in *Callosobruchus maculatus*. I. Larval competition curves and resource sharing patterns. *Res. Popul. Ecol.*, in press.
- Tsukagoshi, A. (1988). Reproductive character displacement in the ostracod genus *Cythere*. *J. Crust. Biol.* **8**, 563–75.
- Turelli, M. (1988). Phenotypic evolution, constant covariances, and the maintenance of additive variance. *Evolution* **42**, 1342–7.
- and Barton, N. H. (1990). Dynamics of polygenic characters under selection. *Theoret. Popul. Biol.* **38**, 1–57.
- Via, S. and Lande, R. (1985). Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution* **22**, 19–33.
- Vrijenhoek, R. C. (1978). Coexistence of clones in a heterogeneous environment. *Science* **199**, 549–52.
- (1984). Ecological differentiation among clones: The frozen niche variation hypothesis. In *Population biology and evolution* (ed. K. Wohrmann and V. Loeschke), pp. 217–31. Springer-Verlag, Heidelberg.
- Wallace, A. R. (1869). *The Malay archipelago*. Macmillan, London.
- Wasserman, S. S. and Futuyma, D. J. (1981). Evolution of host plant utilization in laboratory populations of the southern cowpea weevil, *Callosobruchus maculatus* (Coleoptera:Bruchidae). *Evolution* **35**, 605–17.
- Weeks, S. C. and Sassaman, C. (1990). Competition in phenotypically variable and uniform populations of the tadpole shrimp, *Triops Longicaudatus* (Notostaca Triopsidae). *Oecologia* **82**, 552–9.
- Werner, E. E. (1986). Species interactions in freshwater fish communities. In *Community ecology* (ed. J. Diamond and T. J. Case), pp. 344–58. Harper & Row, New York.
- Wiens, J. A. and Rotenberry, R. T. (1981). Morphological size ratios and competition in ecological communities. *Am. Nat.* **117**, 592–9.
- Williams, V., White, J., Schwaner, T. D. and Sparrow, A. (1988). Variation in venom proteins from isolated populations of tiger snakes (*Notechis ater niger*, *N. scutatus*). *Taxicon* **26**, 1067–75.
- Wilson, E. O. (1961). The nature of the taxon cycle in the melanesian ant fauna. *Am. Nat.* **95**, 169–93.
- Wingate, D. W. (1965). Terrestrial herpetofauna of Bermuda. *Herpetologica* **21**, 202–18.
- Yoshida, T. (1966). Studies on the interspecific competition between bean weevils. *Mem. Fac. Lib. Arts Educ., Miyazaki Univ.* **20**, 59–98.

This page intentionally left blank

DNA variation in humans and its implications for human evolution

WEN-HSIUNG LI and LORI SADLER

1. INTRODUCTION

A knowledge of the extent and pattern of genetic variability in natural populations is a prerequisite for understanding the mechanism of evolution. For this reason, electrophoresis was used extensively from the late 1960s to the early 1980s as a simple, convenient technique for studying the genetic variability in natural populations (see reviews by Lewontin 1974; Nei 1975). In view of the tremendous progress in recombinant DNA techniques one would expect that some of these techniques would have been heavily used to study DNA variation in natural populations. Unfortunately, this has not been the case. Some systematic studies of DNA variation have indeed been made in *Drosophila* populations (see, for example, Langley *et al.* 1982; Kreitman 1983; Aquadro *et al.* 1988; Aquadro 1991; Kreitman and Hudson 1991), but these studies are not extensive enough to provide a detailed assessment of the level and pattern of DNA variation in even the best studied *Drosophila* species, *D. melanogaster*. Unexpectedly, in humans, a better assessment can be made (Li and Sadler 1991) – not as a consequence of any systematic study but as a consequence of the fact that many human genes have been carefully sequenced from two or more individuals. We review this assessment and make a comparison with the *Drosophila* data.

Owing to the curiosity of humans about their own past, human evolution has received better attention in terms of the application of recombinant DNA techniques. Most of these studies are concerned mainly with two issues. One issue is whether the human species has ever gone through any severe bottleneck. Some authors (e.g. Brown 1980; Wainscoat *et al.* 1986; Jones and Rouhani 1986) suggest that a severe bottleneck has occurred during human evolution whereas others (e.g. Mayer *et al.* 1988; Xiong *et al.* 1991) argue for the contrary. The other issue concerns the origin of anatomically modern humans. There are two major competing models. The ‘complete replacement’ or ‘Out of Africa’ model postulates an African origin for modern man and a rapid, complete replacement of indigenous populations in other regions of the world by the African stock (see, for example, Stringer and Andrews 1988) whereas the ‘multiregional evolution’ model postulates a gradual, simultaneous transformation of

archaic regional human populations into modern ones by gene flow and natural selection (Wolpoff *et al.* 1984). Cann *et al.* (1987a) and Wilson *et al.* (1987) argue that their mitochondrial DNA (mtDNA) data support the former view whereas Xiong *et al.* (1991) contend that the sequence data from two apolipoprotein C-II deficiency alleles are more compatible with the latter view. We review molecular studies on the above two controversies.

2. DNA VARIATION IN HUMANS

2.1 Nuclear DNA

A commonly used measure of genetic variability at the nucleotide level is the nucleotide diversity. This measure is defined as the number of nucleotide differences per site between two randomly chosen sequences from a population (Nei and Li 1979); it is the same as the proportion of different nucleotides between two randomly chosen sequences from a population. When more than two sequences are available, the average nucleotide diversity is computed as the arithmetic mean of all pairwise comparisons.

Searching through literature and data banks one can find many human genes, particularly their coding regions, that have been sequenced two times or more. However, the majority of these sequences are not suitable for the study of nucleotide diversity because their sequence accuracy has not been examined carefully. Fortunately, in some cases a pair of sequences have been obtained from the same laboratory and have been carefully checked against each other; they were usually one cDNA and one genomic sequence and the cDNA and genomic libraries used were constructed from different individuals. In some other cases, one laboratory has compared their sequence (usually genomic) carefully with a published sequence. Using such sequences Li and Sadler (1991) have been able to obtain an approximate estimate of the nucleotide diversity in humans (Table 1).

In Table 1, coding and non-coding regions are treated separately. The non-coding regions are divided into 5' and 3' untranslated (UT) regions; flanking (untranscribed) regions are not considered because of the paucity of data. In coding regions, a site is labelled non-degenerate if all changes at that site are non-synonymous (amino acid-changing), two-fold degenerate if one of the three possible changes is synonymous, and four-fold degenerate if all three possible changes are synonymous. The calculation can easily be done by using the computer program of Li *et al.* (1985).

In most of the cases in Table 1 no difference was found between the

Table 1
Nucleotide diversity in humans

Gene ^a	Non-coding regions		Coding regions			
	5'UT	3'UT	Non-degenerate	2-fold degenerate		4-fold degenerate
				non-syn.	syn.	
Acid glycoprotein α1	0/78	0/143	0/383	0/134	0/134	0/83
Adenosine deaminase	0/72	0/311	0/701	0/212	0/212	2/173
α amylase (salivary)	0/199	0/30	0/1017	0/301	0/301	0/215
Aldose reductase	NA	0/371	0/612	0/200	0/200	0/133
Androgen receptor	0/77	0/139	0/1773	0/510	0/510	0/471
Angiogenin	NA	0/175	0/271	0/100	0/100	0/67
Angiotensinogen	0/39	0/602	1/919	0/268	0/268	0/265
Apolipoprotein A-I	0/86	0/57	0/508	0/164	0/164	0/126
Apolipoprotein A-II	0/9	0/112	0/188	0/60	0/60	0/49
Apolipoprotein E	0/61	0/142	1/571	0/158	3/158	0/168
Apo ferritin H	0/91	0/161	0/349	0/134	0/134	0/66
Calcitonin/CGRP	0/74	NA	0/235	0/80	0/80	0/66
Cathepsin G	0/8	0/81	0/466	0/167	0/167	0/129
Complement C1 inhibitor	NA	0/264	0/625	0/190	0/190	0/154
Elongation factor 1, α	0/53	0/295	0/911	0/245	0/245	0/227
Erythropoietin	0/179	0/565	0/359	0/103	0/103	0/114
Factor VIII	0/109	0/1800	0/4558	1/1527	0/1527	0/965
Factor IX	NA	0/1389	1/900	0/292	0/292	0/188
Factor X	NA	NA	0/927	0/300	0/300	0/201
Fibrinogen γ	0/80	0/241	1/858	0/288	1/288	2/162
Gdx	0/35	0/1815	0/321	0/113	0/113	0/91
Granulocyte colony stimulating factor	0/31	0/853	0/389	0/112	0/112	0/117
Hepatic lipase	0/4	0/46	1/964	0/301	0/301	1/229
Interleukin 1, β	0/87	0.7/599	0/525	0/175	0/175	0/104
Interleukin 5	0/44	0.7/357	0/245	0/97	0/97	0/57
Keratin 18	0/47	1/68	0/811	0/268	0/268	0/208

Table 1
Continued

Gene ^a	Non-coding regions		Non-degenerate	Coding regions			
	5'UT	3'UT		2-fold degenerate		4-fold degenerate	
				non-syn.	syn.		
Alpha lactalbumin	NA	0/272	0/271	0/106	0/106	0/46	
Lactate dehydrogenase A	0/25	0/565	0/649	0/195	0/195	0/149	
Lactate dehydrogenase B	0/7	0/200	0/655	0/195	0/195	0/149	
Ly1-1	0/258	0/286	0/496	0/130	0/130	0/172	
Pancreatic polypeptide	0/56	0/76	0.5/173	0/53	0/53	0/56	
Parathyroid hormone	0/74	0/348	0/221	0/72	0/72	0/49	
Phosphoglycerate kinase	0/79	0/434	0/813	0/230	0/230	0/205	
Phosphoglycerate mutase	0/35	0/36	1/495	0/141	0/141	1/121	
Plasmin inhibitor α_2	NA	1/729	1/933	0/277	0/277	0/254	
Prolactin	0/4	0/145	0/430	0/141	0/141	0/107	
Protein C	0/98	0/360	0/899	0/274	0/274	0/207	
Renin	0/42	0/172	0/795	0/217	0/217	0/203	
Ribosomal protein S14	0/33	0/45	0/294	0/74	0/74	0/85	
Steroid 21-hydroxylase	0/32	0/492	0/932	0/287	0/287	0/260	
Superoxide dismutase (Cu/Zn)	NA	0/94	0/297	0/87	0/87	0/75	
Thrombomodulin	0/150	0/1117	0/1115	0/291	0/291	0/316	
Tissue plasminogen activator	0/218	0/755	0/1077	0/358	1/358	1/248	
Thymidine kinase	1/57	0/659	0/450	0/133	0/133	0/116	
Triosephosphate isomerase	0/367	0/718	0/487	0/129	0/129	1/128	
Tumor antigen p53	0/135	NA	1/762	0/210	0/210	0/204	

Table 1
Continued

Gene ^a	Non-coding regions		Coding regions			
	5'UT	3'UT	Non-degenerate	2-fold degenerate		4-fold degenerate
				non-syn.	syn.	
Tumor necrosis factor	0/152	0/789	0/440	0/130	0/130	0/126
Vasoactive intestinal polypeptide	0/176	4/767	0/325	0/108	0/108	0/74
Von Willebrand factor	0/163	0/94	2/1474	0/450	0.5/450	1.3/359
Total	1/3624	7.4/19769	10/34869	1/10787	5.5/10787	9.3/8537
π	0.0003	0.0004	0.0003	0.0001	0.0005	0.0011
	$\pm 0.0003^b$	± 0.0001	± 0.0001	± 0.0001	± 0.0002	± 0.0036

Taken from Li and Sadler (1991).

^a The numbers of sequences used are three for each of the genes for acid glycoprotein α_1 , interleukin 1 β , and interleukin 5, two for the pancreatic polypeptide gene, eight for the von Willebrand factor gene, and two for each of the other genes.

^b The standard errors are computed by assuming binomial sampling of nucleotides.

sequences compared. In the 5' UT region, only one nucleotide difference is observed; it is in the gene for thymidine kinase. (Note that this region is short in most genes and no adequate data are available for many of the genes under study.) In the 3' UT region, variation is noted in the genes for interleukin 1 β , interleukin 5, keratin 18, plasmin inhibitor α , and vasoactive intestinal polypeptide. At the non-degenerate sites, variation is seen in the genes for angiotensinogen, apolipoprotein E, factor IX, fibrinogen γ , hepatic lipase, phosphoglycerate kinase, plasmin inhibitor α , tumor antigen p53, and von Willebrand factor. At the two-fold degenerate sites, non-synonymous variation is observed only in the gene for factor VIII while synonymous variation is observed in the genes for apolipoprotein E, fibrinogen γ , tissue plasminogen activator, and von Willebrand factor. At the four-fold degenerate sites, variation is observed in the genes for adenosine deaminase, fibrinogen γ , hepatic lipase, phosphoglycerate mutase, tissue plasminogen activator, triosephosphate isomerase, and von Willebrand factor.

The average level of nucleotide diversity (π) computed from the pooled data is low (Table 1). The highest level is only 0.11 per cent, which is

observed at the four-fold degenerate sites. The second highest level is observed at the two-fold degenerate sites; the synonymous and non-synonymous components are 0.05 and 0.01 per cent, respectively, and the total diversity is 0.06 per cent. The third highest level, $\pi = 0.04$ per cent, is observed in the 5' untranslated (UT) regions, though it is not significantly different from the π values in the 3' UT regions and at the non-degenerate sites, both being 0.03 per cent. The weighted average for all transcribed regions is 0.04 per cent.

Li and Sadler (1991) have also compared sequences that have not been checked carefully against each other and found that the level of nucleotide diversity is about 10 times higher than that for carefully checked sequences. Therefore, unchecked sequences are not suitable for studying nucleotide diversity.

The results in Table 1 may be taken as representing the approximate level of diversity in the American white population, because most of the DNA libraries used were constructed from white Americans. However, it probably represents an upper estimate for two reasons. First, some of the libraries were from Japanese, Europeans, Australians, or other groups. Second, the data used are probably not completely free of sequencing errors.

The relative levels of diversity in different DNA regions shown in Table 1 are consistent with the relative degrees of sequence divergence obtained from between species comparisons and, like the latter observations, can be explained by the relative stringencies of selective constraints in different regions (see Li *et al.* 1985). For example, in coding regions, non-degenerate sites and four-fold degenerate sites would be subjected to, respectively, the strongest and the weakest selective constraints and would show, respectively, the lowest and the highest level of diversity. In fact, this is the case and the level of diversity at four-fold degenerate sites is about four times higher than that at the non-degenerate sites, which is about the same ratio obtained from between species comparisons (Li *et al.* 1985). As another example, the 5' UT region is in general more conservative than the 3' UT region and so the diversity at the 5' UT region would tend to be lower than that at the 3' UT region, though in the present case the difference is not significant because the sample is not large enough.

It is clear that the nucleotide diversity in humans is low. There are two possible explanations for this observation: either the effective population size of the human species has been relatively small in the past, or the species has gone through a severe bottleneck in the recent past. As will be discussed later, there is strong evidence against the second explanation.

2.2 Mitochondrial DNA

A number of authors have studied mtDNA variation in human populations (e.g. Brown 1980; Greenberg *et al.* 1983; Johnson *et al.* 1983; Wallace *et al.* 1985; Horai and Matsunaga 1986; Cann *et al.* 1987a); most of these studies used restriction enzyme techniques. To date, the most extensive study seems to be the one by Cann *et al.* (1987a) and their result is cited in Table 2. In this table the diagonal elements represent within-population diversity. Caucasians, Australians, and New Guineans have the lowest diversity, about 0.25 per cent and the Africans have the highest diversity, 0.47 per cent. Since the mt genome contains very little non-coding DNA, we compare these values with the average diversity for transcribed regions in the nuclear genome, which has been estimated above to be 0.04 per cent for the American whites. Thus, for Caucasians the average diversity for mtDNA is about six times (0.23 per cent/0.04 per cent) the average diversity for the transcribed regions in the nuclear genome. The higher diversity in mtDNA is probably due to a higher rate of mutation.

2.3 Comparison with *Drosophila* nuclear DNA

We now compare the nucleotide diversity in man with those in *Drosophila* populations. Table 3 represents a summary of the nucleotide diversity in three *Drosophila* species estimated from restriction enzyme data. Although restriction enzymes do not detect all variation between sequences, the level of nucleotide diversity estimated by this method for the Adh region (Langley *et al.* 1982) is the same as that obtained by direct nucleotide sequencing (Kreitman 1983). Moreover, results of detailed surveys of the

Table 2
MtDNA divergence within and between five human populations^a

Population	Sequence divergence (%)				
	1	2	3	4	5
African	0.47	0.04	0.04	0.05	0.06
Asian	0.45	0.35	0.01	0.02	0.04
Australian	0.40	0.31	0.25	0.03	0.04
Caucasian	0.40	0.31	0.27	0.23	0.05
New Guinean	0.42	0.34	0.29	0.29	0.25

Taken from Cann *et al.* (1987a).

^a Values of the mean pairwise divergence between individuals within populations (δ_s) appear on the diagonal. Values below the diagonal (δ_{xy}) are the mean pairwise divergences between individuals belonging to two different populations, X and Y. Values above the diagonal (δ) are interpopulation divergences, corrected for variation within those populations.

Table 3
Nucleotide diversity in species of Drosophila^a

Regions ^b	<i>D. pseudoobscura</i>		<i>D. simulans</i>		<i>D. melanogaster</i>	
	Length (kb)	π	Length (kb)	π	Length (kb)	π
<i>Adh</i>	32	0.026	13	0.015	13	0.006
<i>Amy</i>	26	0.019	—	—	15	0.008
<i>rosy</i>	5	0.013	100	0.018	100	0.005
Average (weighted)		0.022		0.018		0.005

^a Data compiled by Aquadro (1991).

^b Only autosomal regions are used so that a comparison can be made with the nucleotide diversity in Table 1, which is estimated mostly from autosomal genes.

rosy region by four-base recognition enzymes in both *D. pseudoobscura* (Riley *et al.* 1989) and *D. melanogaster* (C. F. Aquadro, personal communication) gave similar estimates as those obtained from six-base recognition enzymes. We may therefore assume that the values in Table 3 are reasonably accurate estimates of the levels of nucleotide diversity in the regions studied. The average of the estimates in each species may be taken as the average nucleotide diversity in non-coding regions because the three regions contain mostly sequences that do not code for amino acids.

In *D. pseudoobscura* the average nucleotide diversity for the three regions is 2.2 per cent (Table 3). Since this represents largely the diversity in non-coding regions, it cannot be directly compared with the estimates for humans shown in Table 1. However, comparative analyses have shown that in mammalian genes the rate of nucleotide substitution at four-fold degenerate sites is only slightly lower than that in pseudogenes, suggesting that four-fold degenerate sites are subject to only weak selective constraints (see Li and Graur 1991). Therefore, the nucleotide diversity at four-fold degenerate sites in human genes ($\pi = 0.11$ per cent) is unlikely to be two-fold lower than that in non-coding regions and we may conclude that in non-coding regions the nucleotide diversity in humans is one order of magnitude lower than in *D. psuedoobscura* ($\pi = 2.2$ per cent). This is in sharp contrast to the observation that at the electrophoretic level, humans and *D. pseudoobscura* show the same level of genetic variability (see Nei and Graur 1984).

Why should the level of nucleotide diversity differ so much between the two species, though at the electrophoretic level the average heterozygosity is about the same in the two species? A simple explanation

is to assume: (1) the majority of nucleotide changes in non-coding regions are selectively neutral or almost neutral whereas the majority of electrophoretic variants are slightly deleterious, and (2) the long-term effective population size in *D. pseudoobscura* is considerably (say, 10 times) larger than that in *Homo sapiens*. Under these two assumptions the level of nucleotide diversity will be much higher in *D. pseudoobscura* than in *H. sapiens* because a larger population can accumulate more neutral mutations than a smaller one. On the other hand, the former species may not necessarily have a higher heterozygosity for electrophoretic variants than the latter because selection against slightly deleterious mutations is more effective in a large population than in a small one. If the above two assumptions are correct, then the contrast in the extent of genetic variability at the two levels between the two species supports Ohta's (1974) hypothesis of slightly deleterious mutation.

The average nucleotide diversity in *D. melanogaster* is 0.5 per cent (Table 3). This is much lower than in *D. pseudoobscura*, probably because of a smaller effective population size (Choudhary and Singh 1987; Aquadro 1991). However, it is considerably higher than that in humans. Since among all the *Drosophila* species studied to date, *D. melanogaster* has the lowest level of nucleotide diversity (see the review by Aquadro 1991), we may conclude that the level of nucleotide diversity in human populations is much lower than those in *Drosophila* populations.

3. ORIGINS AND EVOLUTION OF MODERN HUMANS

3.1 The issue

It is widely agreed among anthropologists that living populations of *Homo sapiens* differ substantially in certain anatomical features (e.g. cranial size and shape) from the more archaic Upper Pleistocene representatives of the species. However, there has been much disagreement with respect to the origin of anatomically modern humans. The question has been whether all living populations have a single recent (Late Pleistocene) origin or whether they evolved in many different regions from local archaic populations of *H. sapiens* (for reviews, see Howells 1976; Wolpoff *et al.* 1984; Stringer and Andrews 1988).

The single origin hypothesis is essentially a punctuation model and assumes that modern humans emerged in a local region as a consequence of isolation and speciation and that this new species completely replaced all archaic populations as they migrated throughout the Old World. This hypothesis has been known as the 'Noah's Ark' hypothesis (Howells 1976). The source region has been suggested by various authors to be in Europe, China, the Near East or western Asia, sub-Saharan Africa, or

Australia (see the review by Wolpoff *et al.* 1984). The earliest fossils of anatomically modern humans have been found at Omo in Ethiopia (see Butzer *et al.* 1969; Leakey *et al.* 1969), Border Cave in South Africa (see Cooke *et al.* 1945; de Villiers 1976; Beaumont *et al.* 1978; Butzer *et al.* 1978), and Klasies River Mouth in South Africa (see Singer and Wymer 1982). These fossils have been dated to be 90000 years BP or older. Although the interpretation of these fossils has not been unanimously accepted (see, for example, Wolpoff 1980; Binford 1986), it is now commonly accepted that modern humans originated in Africa (see Stringer and Andrews 1988). This version of the 'Noah's Ark' hypothesis is known as the 'Out of Africa' hypothesis (see Stringer 1990). This view is shown schematically in Fig. 1a.

An alternative model for the origin of modern *Homo sapiens* is that modern populations evolved in different areas from already differentiated ancestral groups of archaic *H. sapiens* or *H. erectus* (see the review by Wolpoff *et al.* 1984). This hypothesis stemmed from Weidenreich's (1943, 1946) view that fossils from North China and Australasia indicate a continual transition in morphological characters from archaic to modern humans in these areas. He believed that this was also the case in Africa and Europe. Since this model stresses the local regional continuity in

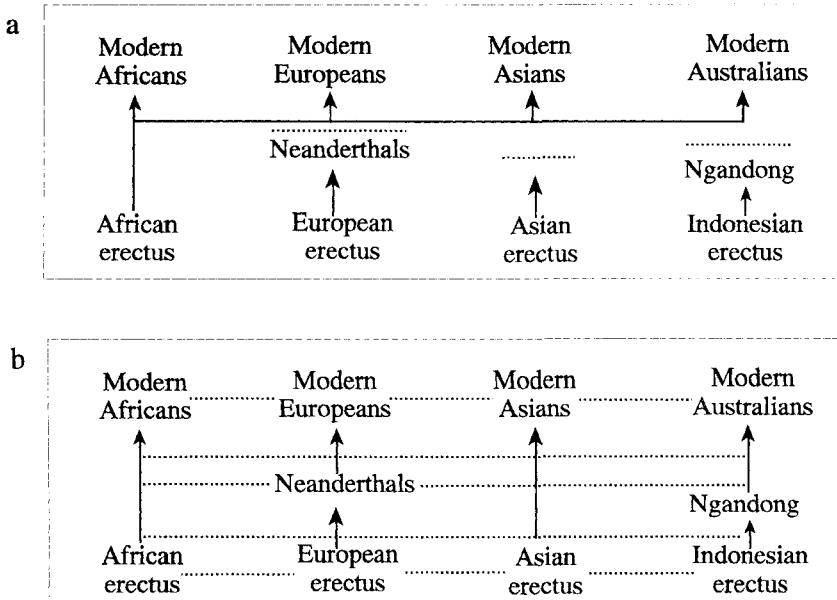


Fig. 1. (a) The 'Out of Africa' model and (b) the multiregional model for the origin of anatomically modern humans. Dashed lines represent migration or gene flow. Taken from Stringer (1990).

morphological evolution, it has been known as the 'regional continuity' model. Although Weidenreich did perceive the possibility of gene flow between regions, Coon (1962) proposed that different regions evolved independently in isolation. This version of the model requires independent species-changing mutations in different regions, which is extremely unlikely. To emphasize the importance of gene flow, Thorne and Wolpoff (1981) and Wolpoff *et al.* (1984) modified the 'regional continuity' model to the 'multiregional evolution' model, which postulates a gradual transformation of archaic regional populations to modern ones by gene flow and natural selection. Fig. 1b shows a schematic representation of this view.

There are important differences between the 'Out of Africa' and the 'multiregional' models (see Wolpoff *et al.* 1984; Stringer and Andrews 1988). The former stresses a complete replacement of all indigenous populations in the rest of the Old World by an African stock. Under this model, any anatomical features that might have evolved in different regions of the world in earlier *Homo erectus* or archaic *H. sapiens* populations would have been lost, so that there would be no specific regional continuity of anatomical characteristics from, say, a million years ago through to the modern world. In contrast, the multiregional model predicts continuity of regionally developed characteristics. In terms of genetic changes, the differences between the two models are even more dramatic. Since the 'Out of Africa' model assumes that the original African stock represented a new species, it could not hybridize with any indigenous regional populations. Therefore, no archaic humans or indigenous regional populations had contributed to the gene pool of modern humans. On the other hand, the multiregional model denies any complete replacement of local gene pools and predicts regional genetic continuities.

It should be noted that the above two schools of thought represent two extreme views of the origin and evolution of modern humans. In our opinion, it is quite possible that the true story is somewhere between the two extremes. For example, a third possible scenario is as follows. Humans with modern anatomical features first appeared in a local region, say in Africa, but the group did not represent a new species and still could 'hybridize' with other local populations. Without complete replacement but through gene flow and migration, and through natural selection, modern humans gradually spread through the world. This model differs from the 'Out of Africa' model in that the latter assumes that the original stock represented a new species and this species had completely replaced the archaic species. The new model is quite similar to the multiregional model but differs from the latter in one important aspect: it assumes that modern humans first emerged in a local region and gradually spread to other regions whereas the multiregional model assumes

that the transformation of archaic to modern humans occurred in parallel in many different regions of the Old World. However, like the multiregional model, the new model predicts regional continuities in both morphological and genetic changes.

It should also be clear from the above discussion that locating the origin of modern humans is different from proving the 'Noah's Ark' or the 'Out of Africa' hypothesis, for the latter requires additional evidence that the original stock had completely replaced all other populations. In other words, to establish the 'Out of Africa' hypothesis one must show not only that the root of the human tree is in Africa but also that there is no continuity in both the fossil record and genetic change.

3.2 Molecular data

Many authors (e.g. Cavalli-Sforza and Bodmer 1971; Nei and Roychoudhury 1982) have used molecular data to study the relationships among human populations. Here we shall review only studies that are directly related to the issue of the origin of modern humans.

Johnson *et al.* (1983) used restriction enzyme cleavage techniques to study the radiation of mtDNA among five groups: Caucasians (US Whites and Europeans), Asians (Taiwan, mainland China, and Japan), Bushmen in Botswana, and Warao Indians in Venezuela. Their tree is shown in Fig. 2. If one assumes roughly equal rates of evolution among the branches of the tree, the root of the tree would be placed between Bushmen and all other groups and this would suggest that modern humans

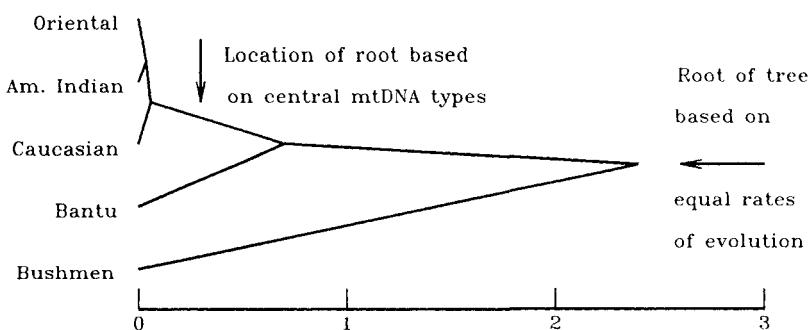


Fig. 2. The evolutionary tree for five human groups constructed from genetic distances estimated from restriction enzyme data. There are two possible roots: one based on the assumption of equal rates of evolution, and the other based on the assumption of identifying the root as one of the three central mtDNA types. The units represent genetic distances. Bushmen diverge at 2.4 units, Bantu at 0.7, Caucasians at 0.06, and Asians diverge from Amerindians at 0.038 units of the genetic distance. Taken from Johnson *et al.* (1983).

originated in Africa. However, Johnson *et al.* (1983) preferred another approach of rooting the tree. They identified three mtDNA types as 'central' to ethnic radiations because of their high frequencies, their appearance in more than one ethnic group, or their presence in other primate species. If one of these central types in fact represents the ancestral type, then the root of the tree would be between African groups and the non-African groups. Under this assumption the Bushmen lineage would have evolved three to five times faster than the non-African lineages, a rate difference that seems unreasonably large. It is difficult to judge which of the two hypotheses is more plausible because there are uncertainties involved in both approaches of rooting the tree.

Wainscoat *et al.* (1986) studied the evolutionary relationships among eight populations of Europeans, Asian Indians, Asiatics, and Africans by examining the presence (+) or absence (-) of each of five closely linked polymorphic restriction enzyme sites on a segment of the β -globin cluster. Noting that three restriction cleavage haplotypes, (+ - - -), (- + - + +), and (- + + - +), are common in Europe and Asia, while the type (- - - +) is common in Africa and absent elsewhere, they proposed that a small founder population of modern humans migrated from Africa to Eurasia, losing by genetic drift the characteristic African type (- - - +) en route to Eurasia. Giles and Ambrose (1986) pointed out that while the geographic distribution of the haplotypes is compatible with the hypothesis that modern humans originated from Africa, as Wainscoat *et al.* (1986) and Jones and Rouhani (1986) have contended, it is equally compatible with the position that modern humans originated in Eurasia and migrated to Africa, or perhaps other alternatives.

Cann *et al.* (1987a) conducted a detailed restriction cleavage analysis of mtDNA types in five groups of humans: Africans, Asians, Australians, New Guineans, and Caucasians. Based on the tree of minimum length (Fig. 3), they inferred that Africa is the likely source of the present-day human mitochondrial gene pool because one of the two primary lineages leads exclusively to African mtDNAs and the other lineage to some Africans and all other populations. This proposal minimizes the number of intercontinental migrations needed to account for the geographic distribution of mtDNA types and is consistent with the observation that Africans have higher mtDNA diversity than other populations (Table 2, Fig. 3). Their inference is reasonable, if the rate of nucleotide substitution in mtDNA is roughly constant among different human lineages. Nevertheless, this inference should be taken with caution because inferring the root of a tree is usually difficult. Indeed, the root in Fig. 3 has been challenged by Saitou and Omoto (1987) [see the rebuttal by Cann *et al.* (1987b)]. More recently, Excoffier and Langaney (1989) reanalyzed published restriction cleavage data of mtDNA and concluded that modern humans originated in Europe. However, their conclusion needs to be

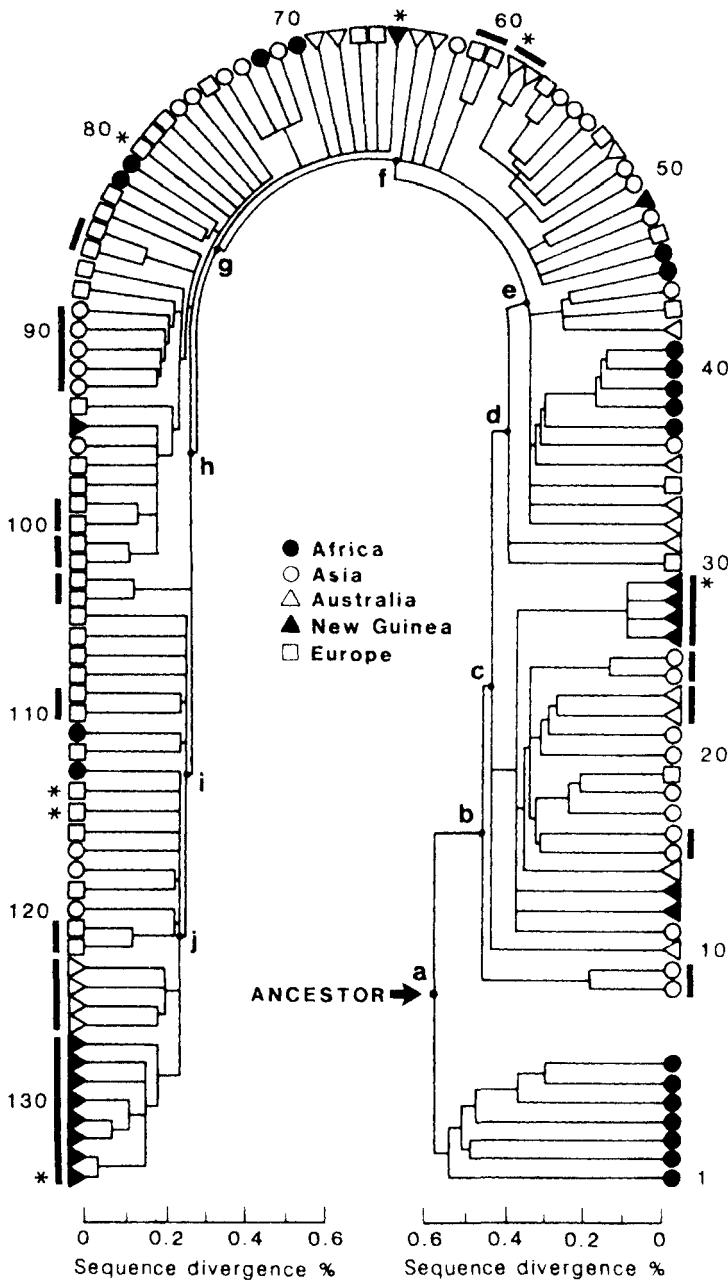


Fig. 3. Genealogical tree for 134 types of human mitochondrial (mt) DNA based on parsimony inference of restriction site differences. The root was placed at the mid-point of the longest path connecting the two lineages. The numbers refer to mtDNA types. Black bars, clusters of mtDNA types specific to a given geographic region; asterisks, mtDNA types found in more than one individual. Taken from Cann *et al.* (1987a).

substantiated by other types of data because they used essentially a parsimony analysis, which provides only the topology but not the root of a tree.

Assuming a rate of 1–2 per cent nucleotide substitution per lineage per million years, Cann *et al.* (1987a) also estimated that all present-day mtDNA types stem from one woman (point 'a' in Fig. 3) who lived about 200 000 years ago; this woman has been dubbed our 'common mother' (Wilson *et al.* 1987). However, there are at least two problems with this estimate, besides the uncertainty of the assumption of a constant rate of evolution of mtDNA among human lineages. First, the rate of nucleotide substitution they used was calibrated from the extent of differentiation within clusters specific to New Guinea (Table 2), Australia (Stoneking *et al.* 1986), and the New World (Wallace *et al.* 1985), assuming a minimum time depth of colonization: 30 000 years ago for New Guinea, 40 000 years ago for Australia, and 12 000 years ago for the New World. This estimate is higher than that (0.7 per cent per million years) estimated from mtDNA sequences from humans and apes (Nei 1985). Second, mtDNA is equivalent to a single gene and estimates of divergence time obtained from a single gene are usually subject to large standard errors (see Nei 1985).

Cann *et al.* (1987a) claimed that their data are incompatible with the multiregional hypothesis. Under this hypothesis, one would expect genetic differences of great antiquity within widely separated parts of the modern pool of mtDNAs, but the greatest divergences within clusters specific to non-African parts of the world (Fig. 3) correspond to times of only 90 000–180 000 (their estimates). For this reason they suggested that the early Asian *Homo erectus* (such as Java man and Peking man) contributed no surviving mtDNA lineages to the gene pool of our species. Further, they argued that if there was hybridization between the resident archaic forms in Asia and anatomically modern forms emerging from Africa, one would expect to find extremely divergent types of mtDNA in present-day Asians, more divergent than any mtDNA found in Africa, but there was no evidence for these types of mtDNA among the Asians studied (Johnson *et al.* 1983; Horai *et al.* 1984; Bonné-Tamir *et al.* 1986). However, there are other factors that must be taken into consideration. First, as mentioned above, the evolutionary history inferred from a single sequence such as mtDNA is subject to large stochastic and sampling effects. The importance of sampling effects can be seen by a comparison of Cann *et al.* (1982) with Cann *et al.* (1987a): in the former, 100 individuals were studied and the Australian aborigines were the most divergent, whereas in the latter, 146 individuals were studied and the Africans were the most divergent. On the other hand, stochastic effects resulting from random genetic drift can lead to loss of archaic types of mtDNA (Avise *et al.* 1984; Nei 1987). Second, mtDNA is but a tiny part

of our genetic make-up and so an inference drawn from this molecule may not represent the evolutionary history of our genetic make-up.

The recent study of Xiong *et al.* (1991) suggests that there has been a continuity in genetic change in Asia. They determined the nucleotide sequences of a Japanese and a Venezuelan apolipoprotein (apo)C-II deficiency allele, of a normal Japanese apoC-II gene, and of a chimpanzee apoC-II gene. The normal Japanese sequence is identical to --- and the chimpanzee sequence differs by only three nucleotides from --- a previously published normal Caucasian sequence. In contrast, the two human mutant sequences each differ from the normal apoC-II gene sequence by several nucleotides, including deletions (Table 4). Assuming that humans and chimpanzees diverged 5 million years ago and that the two human mutant alleles have evolved as fast as the η globin pseudogene, Xiong *et al.* (1991) estimated that both mutant alleles have persisted in the human population for more than 2 million years. Allowing stochastic effects, they estimated that each allele has persisted for at least 500 000 years, i.e. since the middle of the time span of *Homo erectus*. The two mutant alleles might have been derived from an ancestral mutant gene because they shared a deletion of a C in exon 3, which caused a shift in the reading frame and was probably the primary cause for their deficiency. They could be both of Mongoloid origin because the Venezuelan sequence could have been derived from Amerindians --- although the sequence was from a Caucasian family, this family has some Amerindian ancestry. Of course, one could also argue that the two alleles were introduced into the Japanese population and the Venezuelan population from Caucasians,

Table 4
Degree of divergence between sequences

Sequence pairs	Deletions				Total divergence
	Base substitutions	No. of deletions	Total no. of bases involved		
ApoC-II _{Jap} vs. Hum ApoC-II	1/965 (0.001)	3/965 (0.003)	4/965 (0.004)		0.005
ApoC-II _{Ven} vs. Hum ApoC-II	4/965 (0.004)	1/965 (0.001)	1/965 (0.001)		0.005
ApoC-II _{Jap} vs. ApoC-II _{Ven}	5/965 (0.005)	2/965 (0.002)	3/965 (0.003)		0.008
Hum vs. Chimp Apoc-II	3/965 (0.003)	0	0		0.003
Hum vs. Chimp η globin pseudogene	32/2156 (0.015)	5/2156 (0.002)	10/2156 (0.005)		0.020

though this explanation is more involved than that of Mongoloid origin. At any rate, it is difficult to argue that they were of African origin, because this explanation would involve many assumptions. If the two mutant alleles were either of Mongoloid or Caucasoid origin, their antiquity suggests that there was a regional continuity in genetic change either in Asia or in Europe.

In summary, two conclusions can be drawn from DNA analysis of human populations: first, there is a major division of human populations into African and non-African groups, and second, the African group shows a higher nucleotide diversity than the non-African group. If the rate of nucleotide substitution does not vary greatly among populations, then these two observations would support the view of an African origin for modern man but would be in disfavor of the multiregional model. On the other hand, the nucleotide sequences of the two apoC-II deficiency alleles support a regional continuity in genetic change in Asia or Europe and are not in line with the prediction of the 'Out of Africa' hypothesis. Although no definite conclusion can be drawn at the present time, the molecular data reviewed above suggest that both the 'Out of Africa' model and the multiregional model are too extreme and the alternative model we proposed above might be closer to the true story of modern human origins.

4. HAS THERE EVER BEEN A SEVERE BOTTLENECK DURING HUMAN EVOLUTION?

There has long been controversy over whether the human species ever went through a severe bottleneck. In 1972 Haigh and Maynard Smith studied whether the present monomorphism of hemoglobin α chain locus in man is consistent with the neutral mutation hypothesis or not. By using the branching process method, they showed that, if the neutral mutation hypothesis is correct, the frequency of mutant alleles that have occurred at this locus in the last 50 000 generations should amount to about 5 per cent, which is much higher than the observed frequency of about 0.1 per cent. From this result, they concluded that either the neutral mutation hypothesis is false or the human population went through a bottleneck recently. The branching process method is, however, not appropriate for studying neutral mutations because it assumes an infinitely large population so that neutral mutations may be accumulated indefinitely. Using the diffusion method, which assumes a finite population size, Nei and Li (1975) showed that the monomorphism of the hemoglobin α chain locus in man can be accommodated with the neutral mutation hypothesis without the assumption of a bottleneck.

Noting that the rate of nucleotide substitution in mammalian mtDNA

is extremely high, estimated to be 0.5–1.0 per cent per lineage per million years, whereas the observed nucleotide diversity is only 0.36 per cent in humans, Brown (1980) suggested that *Homo sapiens* might have passed through a severe population constriction as recently as 180 000 years ago. This suggestion was based on the calculation that under the estimated rate of nucleotide substitution the observed nucleotide diversity for mtDNA could have been derived from a single pair that existed $180 - 360 \times 10^3$ years ago. This reasoning is not persuasive because in any species all the present-day mtDNAs can be traced back to a single female in the past, but this does not mean that there was then a severe bottleneck.

As mentioned above, the observation that the restriction haplotype (----+) of the β -globin complex is common in Africa but absent elsewhere has been taken as an indication that the human population has gone through a bottleneck, i.e. this haplotype became lost in the founder population during its migration from Africa to Eurasia (Wainscoat *et al.* 1986; Jones and Rouhani 1986). More recently, the rather low mtDNA diversity of about 0.3–0.56 per cent per nucleotide (Table 2) has been taken as evidence for a bottleneck during recent human evolution (Wills 1990). However, as discussed below, these suggestions are not supported by other data.

The studies of Lawlor *et al.* (1988), Mayer *et al.* (1988) and others on the major histocompatibility complex (MHC) class I genes in humans and chimpanzees have provided strong evidence against the occurrence of any severe bottleneck during human evolution. These genes encode ubiquitous cell-surface glycoproteins that serve as recognition elements for cytotoxic T lymphocytes in allograft rejection or in the destruction of virus-infected cells (Klein 1986). Some of the MHC loci are known to be highly polymorphic. For example, in humans 23 alleles have thus far been identified at the HLA-A locus and 47 alleles at the HLA-B locus. Interestingly, Mayer *et al.* (1988) and Lawlor *et al.* (1988) found that many of these polymorphisms pre-date the divergence of humans and chimpanzees. This point is illustrated in Fig. 4, which shows a comparison of the α_2 domains of two chimpanzee alleles (ChLA-A126 and ChLA-A108) and a number of human HLA-A alleles. In this domain the human HLA-A11E allele differs from the chimpanzee ChLA-A108 allele by only three amino acids but differs from the other human alleles by four or more amino acids. Thus, the divergence between HLA-A11E and any of the other human alleles would have pre-dated the human–chimpanzee divergence. Such a situation also occurs at the HLA-B locus (Lawlor *et al.* 1988). Had a severe bottleneck occurred during human evolution, the presence of so many trans-species alleles at each of the two loci would not be possible, although these alleles are probably subject

ChLA-A126	-----L-F-----S-----								
ChLA-A108	-----S-----			A	Q-----	L-----T-----			
	100	110	120	130	140	150	160	170	180
HLA-A11E	GSHTIQIMYGCDVGPDPGRFLRGYRQDAYDGKDYIALNEDLRSWTAADMAAQITKRKWEAAHAAEQQRAYLEGRCVEWLRRYLENGKETLQRT								
HLA-A3.1	S-----					E-----L-----D-T-----			
HLA-A3.2	S-----					V-----D-T-----			
HLA-A2.1	V-R-----S-W-----H-Y-----K-----			T-H-----V-L-----T-----					
HLA-A2.2F	L-R-----S-W-----H-Y-----K-----			T-H-----V-W-----T-----					
HLA-A2.2Y	L-R-----S-W-----H-Y-----K-----			T-H-----V-W-----T-----					
HLA-A2.3	V-R-----S-W-----H-Y-----K-----			T-H-----T-E-----W-----T-----					
HLA-A2.4.1	V-R-----S-W-----H-Y-----K-----			T-H-----V-L-----T-----					
HLA-A2.4.2	V-R.-----S-W-----H-Y-----K-----			T-H-----V-L-----T-----					
HLA-A28	.-----S-----		K-----.	T-H-----V-----T-----.					
HLA-Aw24	L-M-F-----S-----H-Y-----K-----			V-----T-DG-----A-----					
HLA-A32	M-----L-----Q-----			Q-----RV-----L-----T-----					
HLA-Aw68.1	M-----S-----		K-----	T-H-----V-W-----T-----					
HLA-Aw68.2	R-----	H-Y-----	K-----	T-H-----V-W-----T-----					
HLA-Aw69	V-R-----S-W-----H-Y-----K-----			T-H-----V-L-----T-----					

Fig. 4. Amino acid sequences of the $\alpha 1$ domain of alleles at the A locus of the major histocompatibility complex (MHC) class 1 loci. The prefixes ChLA and HLA refer to chimpanzee and human alleles, respectively. ‘-’ indicates identity with the HLA-A11E allele. Taken from Mayer *et al.* (1988).

Table 5

Mean extinction time (t) for a defective allele in a population of size N . The selection coefficients against the heterozygote and the mutant homozygote are h and s ; $h < 0$ means a heterozygote advantage. The frequency of the allele is p

N	p	h	s	t (generations)
250	0.004	0	0.05	8.6
500	0.004	0	0.05	15.8
1000	0.004	0	0.05	29.0
1000	0.001	0	0.05	9.5
1000	0.001	-0.005	0.05	11.0
1000	0.001	-0.010	0.05	13.1

to overdominant selection (Hughes and Nei 1988). This can in fact be shown mathematically (see Takahata 1990).

In the previous section we mentioned that both the Japanese and the Venezuelan apoC-II deficiency alleles have probably persisted in the human population for more than 500 000 years. Both mutant alleles are non-functional and result in a complete absence of apoC-II protein in mutant homozygotes. As can be seen from Table 5, a defective allele can persist in a population for only a short time if a bottleneck occurs. For example, under the conditions that the population size is $N = 500$ (and constant over time), that the selective disadvantage against the homozygote is $s = 0.05$, and that the selective disadvantage against the heterozygote is $h = 0.00$, the mean persistence time (t) is only about 16 generations or $16 \times 20 = 320$ years. This is relatively short. For $N \leq 1000$, a heterozygote advantage has only a mild effect on t (Table 5). Therefore, even if the allele has a heterozygote advantage, it has a high probability of becoming lost in a short time following a severe bottleneck. Since the mean time for one of two alleles to become lost is even shorter, i.e. roughly half of the mean time for a single allele, the antiquity of the two apoC-II deficiency alleles strongly suggests the absence of a severe bottleneck during human evolution.

ACKNOWLEDGMENTS

We thank Y. Zhong and Winston Hide for help. This study was supported by NIH grants.

REFERENCES

- Aquadro, C. F. (1991). Molecular population genetics of *Drosophila*. In *Molecular approaches to pure and applied entomology* (ed. J. Oakeshott and M. Whitten), in press. Springer-Verlag, New York.
- , Lado, K. M. and Noon, W. A. (1988). The rosy region of *Drosophila melanogaster* and *Drosophila simulans*. I. Contrasting levels of naturally occurring DNA restriction map variation and divergence. *Genetics* **119**, 875–88.
- Avise, J. C., Neigel, J. E. and Arnold, J. (1984). Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *J. Mol. Evol.* **20**, 99–105.
- Beaumont, P. B., de Villiers, H. and Vogel, J. C. (1978). Modern man in sub-Saharan Africa prior to 49000 years B. P.: A review and evaluation with particular reference to Border Cave. *S. Afr. J. Sci.* **74**, 409–19.
- Binfold, L. R. (1986). Reply to 'On Binfold on klasies River Mouth: Response of the excavators' R. Singer and J. Wymer. *Curr. Anthropol.* **27**, 57–62.
- Bonné-Tamir, B., Johnson, M. J., Natali, A., Wallace, D. C. and Cavalli-Sforza, L. L. (1986). Human mitochondrial DNA types in two Israeli populations – a comparative study at the DNA level. *Am. J. Hum. Genet.* **38**, 341–51.
- Brown, W. M. (1980). Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc. Natl. Acad. Sci. USA* **77**, 3605–9.
- Butzer, K. W., Brown, F. H. and Thurber, D. L. (1969). Horizontal sediments of the lower Omo valley: The Kibish formation. *Quaternaria* **11**, 15–29.
- , Beaumont, P. B. and Vogel, J. C. (1978). Lithostratigraphy of Border Cave, KawaZulu, South Africa: A Middle Stone Age sequence beginning c. 195 000 B. P. *J. Archaeol. Sci.* **5**, 317–41.
- Cann, R. L., Brown, W. M. and Wilson, A. C. (1982). Evolution of human mitochondrial DNA: A preliminary report. In *Human genetics. Part A* (ed. B. Bonné-Tamir), pp. 156–165. Alan R. Liss, New York.
- Cann, R. L., Stoneking, M. and Wilson, A. C. (1987a). Mitochondrial DNA and human evolution. *Nature* **325**, 31–6.
- (1987b). Reply to 'Disputed African origin of human populations.' *Nature* **329**, 111–2.
- Cavalli-Sforza, L. L. and Bodmer, W. F. (1971). *The genetics of human populations*. Freeman, San Francisco.
- Choudhary, M. and Singh, R. S. (1987). Historical effective size and the level of genetic diversity in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Biochem. Genet.* **25**, 41–51.
- Cooke, H. B. S., Malan, B. D. and Wells, L. H. (1945). Fossil Man in the Lebombo Mountains, South Africa: The 'Border Cave', Ingwavuma District, Zululand. *Man* **45**, 6–13.
- Coon, C. S. (1962). *The origin of races*. Knopf, New York.
- De Villiers, H. (1976). A second adult human mandible from Border Cave. Ingwavuma District, KwaZulu, South Africa. *S. Afr. J. Sci.* **72**, 212–5.
- Excoffier, L. and Langaney, A. (1989). Origin and differentiation of human mitochondrial DNA. *Am. J. Hum. Genet.* **44**, 73–85.

- Giles, E. and Ambrose, S. H. (1986). Are we all out of Africa? *Nature* **322**, 21–2.
- Greenberg, B. D., Newbold, J. E. and Sugino, A. (1983). Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene* **21**, 33–49.
- Haigh, J. and Maynard Smith, J. (1972). Population size and protein variation in man. *Genet. Res.* **19**, 73–89.
- Horai, S. and Matsunaga, E. (1986). Mitochondrial DNA polymorphism in Japanese. II. Analysis with restriction enzymes of four or five base pair recognition. *Hum. Genet.* **72**, 105–17.
- Horai, S., Gojobori, T. and Matsunaga, E. (1984). Mitochondrial DNA polymorphism in Japanese. I. Analysis with restriction enzymes of six base pair recognition. *Hum. Genet.* **68**, 324–32.
- Howells, W. W. (1976). Explaining modern man: Evolutionists, versus migrationists. *J. Hum. Evol.* **5**, 477–96.
- Hughes, A. L. and Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167–70.
- Johnson, M. J., Wallace, D. C., Ferris, S. D., Rattazzi, M. C. and Cavalli-Sforza, L. L. (1983). Radiation of human mitochondrial DNA types analyzed by restriction endonuclease cleavage patterns. *J. Mol. Evol.* **19**, 255–71.
- Jones, J. S. and Rouhani, S. (1986). How small was the bottleneck? *Nature* **319**, 449–50.
- Klein, J. (1986). *Natural history of the major histocompatibility complex*. John Wiley and Sons, New York.
- Kreitman, M. (1983). Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* **304**, 412–7.
- and Hudson, R. R. (1991). Inferring the evolutionary histories of the Adh and Adh-dup loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* **127**, 565–82.
- Langley, C. H., Montgomery, E. and Quattlebaum, W. F. (1982). Restriction map variation in the Adh region of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **79**, 5631–5.
- Lawlor, D. A., Ward, F. E., Ennis, P. D., Jackson, A. P. and Parham, P. (1988). HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature* **335**, 268–71.
- Leakey, R. E. F., Butzer, K. W. and Day, M. H. (1969). Early *Homo sapiens* remains from the Omo River region of South-west Ethiopia. *Nature* **222**, 1132–8.
- Lewontin, R. C. (1974). *The genetic basis of evolutionary change*. Columbia University Press, New York.
- Li, W.-H. and Graur, D. (1991). *Fundamentals of molecular evolution*. Sinauer Associates, Sunderland, MA.
- Li, W.-H. and Sadler, L. A. (1991). Low nucleotide diversity in man. *Genetics* **129**, 513–23.
- Li, W.-H., Wu, C.-I. and Luo, C.-C. (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering

- the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2**, 150–74.
- Mayer, W. E., Jonker, M., Klein, D., Ivanyi, P., van Seventer, G. and Klein, J. (1988). Nucleotide sequences of chimpanzee MHC class I alleles: Evidence for trans-species mode of evolution. *EMBO J.* **7**, 2765–74.
- Nei, M. (1975). *Molecular population genetics and evolution*. North-Holland, Amsterdam.
- (1985). Human evolution at the molecular level. In *Population genetics and molecular evolution* (ed. T. Ohta and K. Aoki), pp. 41–64. Japan Sci. Soc. Press, Tokyo.
- (1987). *Molecular evolutionary genetics*. Columbia University Press, New York.
- and Graur, D. (1984). Extent of protein polymorphism and the neutral mutation theory. In *Evolutionary biology* (ed. M. K. Hecht, B. Wallace and G. T. Prance), pp. 72–118. Plenum, New York.
- and Li, W.-H. (1975). Probability of identical monomorphism in related species. *Genet. Res. Camb.* **26**, 31–43.
- and Li, W.-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **76**, 5269–73.
- and Roychoudhury, A. K. (1982). Genetic relationship and evolution of human races. *Evol. Biol.* **14**, 1–59.
- Ohta, T. (1974). Mutational pressure as the main cause of molecular evolution and polymorphism. *Nature* **252**, 351–4.
- Riley, M. A., Hallas, M. E. and Lewontin, R. C. (1989). Distinguishing the forces controlling genetic variation at the *Xdh* locus in *Drosophila pseudoobscura*. *Genetics* **123**, 359–69.
- Saitou, N. and Omoto, K. (1987). Time and place of human origins from mtDNA data. *Nature* **327**, 288–288.
- Singer, R. and Wymer, J. (1982). *The Middle Stone Age at Klasies River Mouth in South Africa*. University of Chicago Press, Chicago.
- Stoneking, M., Bhatia, K. and Wilson, A. C. (1986). Rate of sequence divergence estimated from restriction maps of mitochondrial DNAs from Papua New Guinea. *Cold Spring Harb. Symp. Quant. Biol.* **51**, 433–9.
- Stringer, C. (1990). The Asian connection. *New Scientist* **17**, 33–7.
- and Andrews, P. (1988). Genetic and fossil evidence for the origin of modern humans. *Science* **239**, 1263–8.
- Takahata, N. (1990). A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proc. Natl. Acad. Sci. USA* **87**, 2419–23.
- Thorne, A. G. and Wolpoff, M. H. (1981). Regional continuity in Australasian Pleistocene hominid evolution. *Am. J. Phys. Anthropol.* **55**, 337–49.
- Wainscoat, J. S., Hill, A. V. S., Boyce, A. L., Flint, J., Hernandez, M., Thein, S. L., Old, J. M., Lynch, J. R., Falusi, A. G., Weatherall, D. J. and Clegg, J. B. (1986). Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. *Nature* **319**, 491–3.
- Wallace, D. C., Garrison, K. and Knowler, W. C. (1985). Dramatic founder effect in Amerindian mitochondrial DNAs. *Am. J. Phys. Anthropol.* **68**, 149–55.

- Weidenreich, F. (1943). The 'Neanderthal Man' and the ancestors of '*Homo sapiens*.' *Am. Anthropol.* **45**, 39–48.
- (1946). *Apes, giants, and man*. University of Chicago, Chicago.
- Wills, C. (1990). Population size bottleneck. *Nature* **348**, 398.
- Wilson, A. C., Stoneking, M., Cann, R. L., Prager, E. M., Ferris, S. D., Wrisschnik, L. A. and Higuchi, R. G. (1987). Mitochondrial clans and the age of our common mother. In *Human genetics* (ed. F. Vogel and K. Sperling), pp. 157–64. Springer-Verlag, Berlin, Heidelberg.
- Wolpoff, M. H. (1980). *Paleoanthropology*. Knopf, New York.
- , Zhi, W. X. and Thorne, A. G. (1984). Modern *Homo sapiens* origins: A general theory of hominid evolution involving the fossil evidence from East Asia. *The origins of modern humans: A world survey of the fossil evidence* (ed. F. H. Smith and P. Spencer), pp. 411–83. Alan R. Liss, New York.
- Xiong, W., Li, W.-H., Posner, I., Yamamura, T., Yamamoto, A., Gotto Jr, A. M. and Chan, L. (1991). No severe bottleneck during human evolution: Evidence from two apolipoprotein C-II deficiency alleles. *Am. J. Hum. Genet.* **48**, 383–9.

Evolutionary studies in cultivated rice and its wild relatives

HIROKO MORISHIMA, YOSHIO SANO and
HIKO-ICHI OKA

1. INTRODUCTION

Crops are the products of man-mediated plant evolution. For an understanding of crop evolution, an integration of different disciplines of science is required. Reviewing the evolutionary studies that have so far been carried out in various crops, we notice that these studies have different tints as influenced by biological features of each crop as well as by the scientific interest of each researcher. Our intention in writing this article is to explore some implications of rice evolution from the viewpoints of genetics and ecology. Our arguments are mostly based on our own observations in natural and experimental populations of wild and cultivated rice. Core problems to be addressed are:

1. What genetic changes are associated with domestication processes in rice?
2. What factors are responsible for those evolutionary changes?

Among cereal crops, rice is unique in that (1) it is basically an aquatic plant, differentiated into various types adapted to different water conditions, (2) a variation ranging from outbreeding perennials to inbreeding annuals exists within its primary gene pool, (3) domestication proceeded independently in Asia and in Africa, and (4) varietal differentiation at a subspecies level occurs in Asian cultivars. Our target species dealt with in this article are two cultivated species and their wild relatives, which share the same genome A. Consequently, our principal concern is the dynamics of differentiation at the genic level.

A book entitled *Origin of Cultivated Rice* written by one of us was recently published (Oka 1988). This article does not cover the whole scope of rice evolution as dealt with in the book. But we will try to incorporate some newly obtained information, including molecular diversity, which will enable us to make precise characterization of genic diversity between and within rice taxa. Combining new data with a wealth of information accumulated in the past, we hope to have a better understanding of the dynamics of genetic changes involved in rice evolution.

2. CULTIVATED RICES AND THEIR WILD RELATIVES

2.1 Taxonomic status and the gene pool concept

Two cultivated and about 20 wild species are enumerated in the genus *Oryza*. The current status of taxonomy of the genus was discussed by Vaughan (1989) referring to many taxonomic studies previously done. The common rice grown throughout the world is *O. sativa* L. which is of Asiatic origin and rich in varietal diversity. The African rice endemic to West Africa, *O. glaberrima* Steud. is distinguished from *O. sativa* by some key characters. These two cultigens are isolated from each other by F_1 pollen-sterility.

Among the wild species, close relatives of the cultivated rices are *O. rufipogon* Griff. distributed in tropical Asia and *O. barthii* A. Chev. (often called *O. breviligulata* A. Chev. et Roehr.) distributed mainly in West Africa. The former species, varying in life cycle from perennial to annual types, is thought to be the wild progenitor of *O. sativa*, and the latter species, which is annual, is thought to be the wild progenitor of *O. glaberrima*. The annual form of *O. rufipogon* is called *O. nivara* Sharma et Shastry by some workers. But we lump both perennial and annual types as one species *O. rufipogon*, because the variation between the two types is continuous.

Wild taxa related to *O. rufipogon* are distributed throughout the humid tropics of the world. The African relative, *O. longistaminata* A. Chev. et Roehr. is perennial and rhizomatous. The plants of this taxon are partly self-incompatible and allogamous (Chu *et al.* 1969b; Ghesquiere 1986), but isolated from their sympatric relatives by a crossing barrier (Chu and Oka 1970a). The Australian relative is annual and has been named *O. meridionalis* Ng. The related taxon distributed in Latin America, which varies in perenniability, is called *O. glumaepatula* Steud. (called *O. cubensis* Ekman by some workers). In our early papers, we have designated all these taxa related to *O. rufipogon* as geographical races of the *O. perennis* complex. The two cultigens and all their wild relatives mentioned above are diploids ($2n = 24$) and share the same genome A, as their hybrids show no significant disturbance in chromosome pairing. But, F_1 sterility and other hybrid dysgeneses develop between these taxa.

O. officinalis Wall. ex Watt and its relatives form another large species group. Some of them are tetraploids. They are characterized by a relatively small size of spikelets and are distributed in different continents. Many of them occur in shady habitats. Cytogenetic studies (reviewed by Nayar 1973) revealed that they have genomes B, C, D, and E. In addition, there are several other species that are distantly related to all the species mentioned above.

Nomenclature of genus *Oryza*, in particular that of wild relatives of

the cultigens, is still in confusion to some extent. For a student of genetics and evolution, however, genetic relationships are more important than Latin epithets. The classification of gene pools as suggested by Harlan and de Wet (1971) is useful to describe genetic relationships between cultivated rices and their wild relatives. They partitioned the whole variation range genetically reached by cultigens into three categories: primary gene pool (races yielding reasonably fertile hybrids with cultivars), secondary gene pool (races crossable with cultivars but gene flow is restricted), and tertiary gene pool (crosses with cultivars are difficult, and gene transfer is not possible without special procedures).

According to the classification of gene pools, *O. sativa* and *O. rufipogon* belong to the same biological species and form a primary gene pool. Partial F_1 sterility occurs even within *O. sativa*, but does not hamper gene flow. Similarly, *O. glaberrima* and *O. barthii* form another primary gene pool. Between *O. sativa* and *O. glaberrima*, which are often grown in mixture by African peasant farmers, natural hybrids are rare (Sano 1989a). However, experimental hybridization is easy, and gene transfer can be made by backcrossing (Sano *et al.* 1980a). They may be considered as mutually secondary gene pools. The other wild relatives of *O. rufipogon* having the AA genome may be regarded as the secondary gene pool for the two cultigens.

O. officinalis and other species having genomes B, C, and D are crossable with the cultigens with some difficulty, and produce completely male sterile hybrids. Hybrids between *O. sativa* (AA) and *O. officinalis* (CC) were backcrossed with *O. sativa*. Most of the BC₁ progenies were allotriploids (AAC), as the unreduced gametes often produced in the F_1 were fertilized by *O. sativa* (Li *et al.* 1964; Ho and Li 1965). In the further backcrossed progenies with *O. sativa* (BC₂), Jena and Khush (1990) identified 11 monogenic traits derived from *O. officinalis*, which included resistances to many races of brown planthoppers and white-backed planthoppers. Monosomic alien addition lines from *O. punctata* (BB) to *O. sativa* were also produced by Yasui and Iwata (1991). Some other species can also be crossed with cultivars although the rate of success is lower. Hybrids of *O. brachyantha* A. Chev. et Roehr (FF), *O. meyeriana* Baill and *O. ridleyi* Hook with *O. sativa* were reported (Li 1964; Katayama and Onizuka 1979). These distantly related species may be regarded as tertiary gene pools for *O. sativa*. Wide hybridization will broaden the range of the tertiary gene pool from which alien gene can be transferred.

2.2 Delimitation of wild ancestors of the cultigens

Among different wild taxa with the A genome, only the Asian common wild rice *O. rufipogon* yields fertile hybrids with *O. sativa*. This species

is distributed from West India or Pakistan to Indonesia and China. The northern-most site of its distribution is known to be Dongxiang (28°N), China. From India to Indochina and China, many populations of this species grow in the proximity of rice fields, and are introgressed with genes of *O. sativa*. At present, it is not easy to find populations of this species completely isolated from rice cultivars by a long distance. In Malaysia, Indonesia, and the Philippines, this species is rather rare and most of its populations are distributed distant from rice fields.

In Africa, *O. barthii* often grows in close contact with the African rice *O. glaberrima* producing various intermediate plants. Both species are completely annual and share the same taxonomic characters such as short and tough ligules. No other wild taxa show such a close relation with *O. glaberrima*. These two species are often sympatric with another African wild species *O. longistaminata* and with *O. sativa* introduced from Asia. Although *O. longistaminata* is isolated by a crossing barrier, introgression across the barrier sometimes produces natural hybrids with *O. sativa* (Chu and Oka 1970b). In tropical America and Australia, the wild rices related to *O. sativa* are separated from rice fields by a long distance with a few exceptions.

The Asian common wild rice, *O. rufipogon*, varies between perennial and annual types which differ in life-history traits, mating system and habitat preference (Morishima *et al.* 1984a). The perennials are partly allogamous and grow mainly in deepwater habitats, while the annuals have a lower rate of outcrossing and grow in drier habitats than the perennials, as will be mentioned later.

It has been a subject of discussion whether the perennial or the annual type is the immediate ancestor of *O. sativa*. Based on similarity in botanical and ecological characters, Sampath and Rao (1951) as well as Sampath and Govindaswamy (1958) considered the perennial type to be the progenitor. We have postulated that the perennial type would be the progenitor, because (1) perennial populations contained more genetic variability, and accordingly would have higher evolutionary potential than annual populations; (2) *O. sativa* cultivars are essentially perennial plants; and (3) the intermediate wild-cultivated plants from the Jeypore Tract, India, appeared to form a bridge connecting the perennial type of *O. rufipogon* with *O. sativa* in various traits (Oka 1964a, 1974b).

On the other hand, Chatterjee (1951), following Roschevitz (1931), considered the annual type which he referred to as *O. sativa* f. *spontanea* to be the progenitor. Later, Chang (1976) postulated that *O. sativa* was derived from the annual type, which he referred to as *O. nivara*, because it has a high seed productivity and some other characters similar to *O. sativa*, and because the evolution of cultivars from annual wild forms is the pattern generally recognized in various annual cereals.

During the evolutionary process of wild rice, the annual type has

probably been derived from the primitive perennial type as an outcome of adaptation to drought stress, similarly to intraspecific differentiation of annual types in many grass species in monsoon Asia as discussed by Whyte (1972). The annual populations generally have small genetic variability as expected from their reproductive system associated with predominant selfing, and accordingly do not seem to have a high evolutionary flexibility to derive new forms. On the other hand, typically perennial populations have a low seed productivity and rarely release their potential genetic variability. We found that intermediate perennial-annual populations, which regenerate by both ratoons and seeds, had a moderately high seed productivity and were highly polymorphic genetically (Sano *et al.* 1980b). Further, they seemed to be adapted to disturbed habitats. We are then inclined to assume that such intermediate types are most likely to be the immediate ancestor of *O. sativa*.

2.3 Intermediate wild-cultivated plants

In Asia and Africa, cultigens and their wild ancestors often grow in close proximity. Many intermediate wild-cultivated types are found in such situations. They grow in direct-seeded rice fields, in the form of hybrid swarms that exhibit a continuous variation (Oka and Chang 1961) or as weeds that are distinguishable from neighboring cultivars (Oka and Chang 1959). They are also found in disturbed wet habitats adjacent to crop fields.

The intermediate wild-cultivated plants such as found from the Jeypore Tract, India, have provided us with information on the process of domestication of wild rices. Those plants showed continuous variation between wild and cultivated types as well as between Indica and Japonica types (Oka and Chang 1962). The intermediate plants that we now see are most probably the products of natural hybridization and selection, and have come to exist after the occurrence of cultivated types. But we may assume that it matters little whether they have played a primary or secondary role in the historical sense, since the same forces may act in both the primary and secondary processes of domestication. A continuous variation between wild and cultivated types was also found among and within populations in the *glaberrima–barthii* complex in Africa (Morishima and Oka 1970).

The intermediate types that persist in crop fields can be called weedy forms. They cannot survive in natural habitats. They generally have a high rate of seed shedding, seed dormancy, tolerance to various adverse conditions, and other traits of wild plants such as black hull, red pericarp, and awnedness. Their soil-buried seeds germinate after the rice seeds are sown. They look like cultivars in the vegetative stage, and reach maturity usually earlier than the neighboring cultivars.

The weedy forms occurring in the localities where wild rices are distributed were most probably derived from natural hybrids between wild and cultivated forms (Oka and Chang 1959). There are other types of weedy rice that are found at some sites in China, Korea, Nepal, and other countries where no wild rice exists. Their origin is still unknown.

There are two contrasting views as to the role of weedy forms in crop evolution, one assuming them to be the immediate progenitor of domesticates and the other considering them to have evolved together with crops side by side from a common progenitor (Harlan 1965). However, this matters little insofar as crop evolution often depends on repeated differentiation-hybridization cycles (Harlan 1966). Weedy forms or intermediate wild-cultivated forms contain a large amount of genetic variability and serve as a gene reservoir for cultivars. They may be considered, together with truly wild forms, to be part of the primary gene pool of cultigenes.

3. GENETIC DIVERSITY INFERRED FROM PHENOTYPIC AND MOLECULAR EVIDENCE

3.1 Variation among and within taxa having genome AA

Variation among and within taxa having the AA genome has been studied with the hope of elucidating their phylogenetic relationships. Various quantitative and qualitative characters in *O. rufipogon* and its related wild taxa were surveyed, and the taxa distributed in different continents were distinguished by numerical taxonomic methods (Morishima 1969). Later, extensive variation studies of isozymes revealed the genetic structure of those taxa (Second 1982, 1985), which generally agreed with that previously anticipated from orthodox and numerical taxonomic studies. Nevertheless, the nomenclature of this species group has been often confused in the literature. The discrepancies as to species recognition seem to have originated from the paucity of species characteristics, even if the taxa are isolated by reproductive barriers.

More recently, polymorphisms in Fraction I protein and some repeated DNA sequences (Pental and Barnes 1985), length and sequence of the intergenic spacer of ribosomal DNA repeats (Cordesse *et al.* 1990; Sano and Sano 1990), and restriction pattern of chloroplast DNA (Ichikawa *et al.* 1986; Ishii *et al.* 1988; Dally and Second 1990) were uncovered. Those molecular studies generally showed a close genetic relationship between *O. sativa* and *O. rufipogon*, as well as between *O. glaberrima* and *O. barthii*, supporting that the former two and the latter two share the same primary gene pool, respectively. The intergenic spacer region of rDNA is expected to evolve rapidly compared with ctDNA since its

variation pattern generally shows a high level of family homogeneity within species but a high level of heterogeneity between species (Appels and Honeycutt 1986). Restriction enzyme maps in the AA genome species revealed that length heterogeneity results from repetition of short repeated sequences in the intergenic spacer region and the spacer region greatly differs among reproductively isolated taxa with respect to the length and the sequence (Fig. 1).

The levels of genetic diversity within taxa greatly differed from taxon to taxon. As shown in Table 1, *O. rufipogon* generally showed the highest diversity followed by *O. sativa* and *O. longistaminata*. Sano and Sano (1990) identified 18 different spacer length variants among 243 accessions of species with genome AA, of which 13 variants were found in *O. rufipogon*, and seven of those were found in *O. sativa*. Dally and Second (1990) found 32 different ctDNA restriction patterns among 247 accessions including species having the B, C, D, and E genomes. In their study, *O. rufipogon* exhibited seven different cytotypes, which were all found in *O. sativa*. Both rDNA and ctDNA, despite a large amount of diversity carried by *O. rufipogon*, did not show significant association with the perennial vs. annual variation, which represents the most conspicuous differentiation occurring in this taxon. Direct sequencing of three genes, phytochrome introns, 10 kd prolamin gene family, and chloroplast DNA, conducted by Barbier *et al.* (1991) detected only a few base substitutions

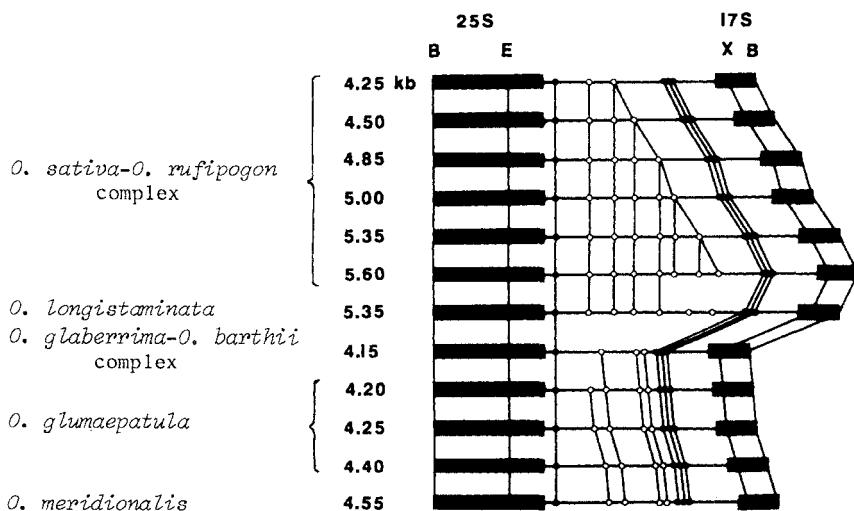


Fig. 1. Restriction endonuclease cleavage maps in the intergenic spacer regions of ribosomal DNA of *Oryza* species with the AA genome as determined by the method of indirect end labeling (Sano and Sano 1990). B, *Bam*HI; E, *Eco*RI; X, *Xba*I; •, *Hinf*I; o, *Sa*II.

Table 1
Genetic variation within AA genome species at isozyme and DNA levels

Species	Isozyme (Recomputed from Second 1985)	rDNA (Sano & Sano 1990)	ctDNA (Computed from Dally & Second 1990)
	H^a	H^b	H^c
<i>O. sativa</i>	0.34 (64) ^d	1.32 (76)	1.50 (74)
<i>O. rufipogon</i>	0.50 (110)	2.21 (77)	1.60 (24)
<i>O. longistaminata</i>	0.26 (20)	2.07 (22)	0 (18)
<i>O. meridionalis</i>	0.11 (12)	0 (16)	0.32 (10)
<i>O. glumaepatula</i>	0.28 (16)	0.90 (20)	0.69 (2)
<i>O. glaberrima</i>	0.04 (6)	0 (20)	0 (6)
<i>O. barthii</i>	0.21 (20)	0 (12)	0.35 (9)

^a Average gene diversity based on eight polymorphic isozyme loci.

^b Diversity index ($H' = -\sum p_i \ln p_i$) based on length polymorphism of spacer region of rDNA.

^c Diversity index based on restriction pattern of chloroplast DNA.

^d Figures in parentheses are the number of accessions tested.

between the perennial and annual accessions of *O. rufipogon*, although *O. longistaminata* was found to be diverged from them.

O. sativa includes diverse cultivars that are classified into different varietal groups, e.g. Indica vs. Japonica types, upland-lowland-deepwater types adapted to different water regimes of habitats, and Aman-Aus-Boro groups grown in different cropping seasons. Varietal differentiation within *O. sativa* is too complex to deal with thoroughly in the present article. Only the Indica vs. Japonica differentiation which exhibits differences comparable to that between subspecies will be discussed later in some more detail. It should be pointed out here, however, that the Indica and Japonica types which were originally defined on the basis of character association pattern (Oka 1958), tended to show significant differences in isozyme and DNA variations. Variant types found at isozyme and DNA levels in *O. sativa* were all found in *O. rufipogon*. In other words, no gene or allele was confined to *O. sativa*. This suggests that a part of variability carried by *O. rufipogon* contributed to the formation of *O. sativa*. It should be noted that, however, major types characterizing the Indica and Japonica types are not always major types in *O. rufipogon*.

O. glaberrima cultivars are less diversified as compared with the Asian counterpart in characters (Morishima *et al.* 1962), in nuclear genes (Second 1985; Sano and Sano 1990), and in organellar genomes (Dally and Second 1990). Together with its wild progenitor *O. barthii*, they showed a narrow genetic variation that is distinct from other AA genome

taxa. This may be accounted for by the paucity of genetic variability in *O. barthii* owing to its annual and predominantly inbreeding nature.

3.2 Differences between wild and domesticated forms

As pointed out by Vavilov (1951) and discussed more fully by Harlan *et al.* (1973), domestication processes, particularly of cereals, have many homologous features in common. Looking at two cultivated rice species, we find that similar selection pressures produced similar domesticated forms in different phylogenetic lineages. The most essential difference between wild and domesticated forms lies in that the former can reproduce by itself, while the latter can reproduce only with the aid of man.

Adaptation to cultivated fields is attained by non-shedding of seeds, rapid and uniform germination, determinate growth, increased seed productivity, and seedling vigor. Those traits are 'adaptive syndromes resulting from automatic selection due to planting harvested seeds' (Harlan *et al.* 1973). This holds true also in cultivated rice. Our early studies demonstrated that 'cultivation pressure' brought about a rapid change of population genotype towards the cultivated type (Oka and Morishima 1971).

Degree of seed shedding is the most critical trait disruptively selected in natural and cultivated fields. The number of major genes underlying seed shedding does not seem to be many, although some modifiers are involved. Eiguchi and Sano (1990) found that one of the two dominant genes conferring high seed shedding on *O. rufipogon*, *Sh-2*, is tightly linked with a gene for spreading panicle, *Spr-3(t)*, and located between *lg* (liguleless) and *Ph* (phenol reaction) on chromosome 4. Spreading panicle is an important adaptive trait in wild rice because it confers a high rate of seed dispersal.

Disruptive selection experiments in cultivated (non-shedding) × wild (shedding) hybrid populations consistently brought about an increase in the frequency of individuals with red pericarp (probably *Rd* on chromosome 1) and positive phenol reaction (*Ph* on chromosome 4) in plots selected for shedding, and conversely a decrease in their frequency after selection for non-shedding. Wild rices are known to have *Ph* and *Rd* exclusively, while cultivars are polymorphic. These results suggest that at least two independent shedding genes exist, each being linked with genes that are differentially distributed between wild and cultivated forms. Since the inheritance of seed shedding seems to be relatively simple as in other cereals, it is reasonable to conceive that picking up non-shedding mutants in wild rice populations is the most likely and simplest step towards domestication.

A high seed productivity characterizes cultivated rice. In this respect, the annual type of *O. rufipogon* is similar to cultivars in yielding more

seeds than the perennial type. Studying isozyme variation, we noticed that allele *Pox-1¹* (slow band) is predominant in the annuals (and exclusively found in cultivars), while the opposite allele *Pox-1²* (fast band) is specific to the perennials. In F₃ plants from an annual × perennial cross of *O. rufipogon*, Morishima (1991) found that allelic variation at *Pox-1* was associated with some quantitative traits, such as reproductive allocation, flowering time, and anther length (correlated with outcrossing rate). These traits are coadaptive characters responsible for differentiation between annuals and perennials. The direction of association between allelic variation at *Pox-1* and these adaptive traits agreed with that consistently observed in nature. During the experimental process of isolating a chromosomal segment marked by *Pox-1* on isogenic backgrounds, their associations decayed at different rates depending on characters. These results suggest that some of the quantitative trait loci (QTL) for perennial–annual syndrome are clustered on a chromosome segment marked by *Pox-1*, but linkages between *Pox-1* and the QTLs are not very tight.

Isozyme variations associated with ecotypic differentiation could be explained either by hitchhiking or by direct selection. If *Pox-1* is located on a certain gene block surrounded by QTLs that are mutually coadapted, its persistent association with QTLs can be expected even if it is selectively neutral. Yet, the selective role of *Pox-1* itself cannot be ruled out, judging from a strong and consistent association between *Pox-1¹* and the annual habit or seed productivity generally observed in other AA-genome species.

Growth habit is usually more determinate in the domesticates than in the wild forms. Cultivated rices show synchronous tillering ensuring uniform seed maturation. In wild rices, the annual type is more determinate than the perennial type. The genetic control of determinate vs. indeterminate growth is little studied in rice. In the F₂ of annual × perennial crosses, Barbier (1990) found negative correlation between the number of panicles per reproductive tiller (degree of branching) and the number of vegetative tillers developed after seed maturation per reproductive tiller. This suggests a genetic trade-off between determinate growth of monocarpic annuals and sustained growth of persistent perennials, which represents an aspect of differentiation in resource allocation strategy.

3.3 Diversification by human selection

In addition to the adaptive syndrome, which is automatically selected by cultivation, we find various variant types of seed size, color, aroma, and grain quality in cultivated rices, which are not found in wild rices. Seed length of *O. sativa* cultivars ranged from 4.4 to 12.3 mm (625 accessions), while that of *O. rufipogon* ranged from 7.5 to 9.9 mm (85 accessions). Coloration of organs in rice is known to be under a complex genic control

as reviewed by Takahashi (1982). A number of alleles at multiple loci, including some inhibitors, determine the enormous color diversity found in cultivars. Estimation for alleles at seven loci for apiculus and stigma coloration showed that the average gene diversity was $H = 0.351$ for *O. sativa*, and $H = 0.052$ for *O. rufipogon*. Negative phenol reaction (*ph*) and glutinous endosperm (*wx*) are also particular traits that are specifically found in cultivars but not found in wild rice.

Amylose content is a major factor in determining eating quality of rice. It ranges from nearly zero (glutinous) to 30 per cent in some Asian rice cultivars, although the wild form shows generally a high amylose content. The *wx* locus specifies a starch-bound 60 kd protein, so-called 'waxy protein,' which is responsible for the amylose synthesis (Sano 1984, Sano *et al.* 1986). The wide variation in amylose content was controlled mainly by a series of alleles at the *wx* locus, which regulate the quantitative level of the gene product, suggesting that the *wx* locus has a potential to produce diversified phenotypes through changes of gene regulation (Sano *et al.* 1991). Since the allelic diversity at the *wx* locus is confined to cultivars (Sano *et al.* 1985), various alleles seem to have been selected from mutations in the wild form during the domestication process.

Drastic increase in phenotypic diversity accompanied by domestication is a general trend found in almost all crops. In man-made habitats, relaxation of natural selection might have allowed various mutant types to survive, which were subjected to human selection. It may be considered that newly emerged alleles and their new combinations contributed to generating diverse phenotypes observed in cultivars.

3.4 Intrapopulational genetic variation

Variation among individuals within a population provides the raw material on which selection operates. It will develop into variations between populations, and finally into speciation.

Intrapopulational genetic variation in metric traits was evaluated in terms of genetic variances of single characters and generalized genetic variance for multiple characters (Morishima and Oka 1970). Our data indicated that the perennial type of *O. rufipogon* and *O. longistaminata* contained greater genetic variability within populations than the annual type of *O. rufipogon* and *O. barthii*. These differences may be accounted for by differences in their life history and mating system. It was also found that an appreciable amount of genetic variability in various characters are preserved in land races of *O. sativa* and *O. glaberrima*.

Isozyme studies confirmed the above results, and revealed the genetic structure of natural populations of *O. rufipogon* more clearly. In particular, perennial populations showed higher gene diversity (H_S), a larger number of polymorphic loci, and higher average number of alleles per locus than

the populations of wild annuals and cultivated rice (Table 2). They are characterized by a high frequency of heterozygotes and a low fixation index as compared with annuals. On the other hand, the annual populations showed a higher index of interpopulational gene differentiation (G_{ST}) than the perennial ones. Intermediate perennial–annual populations and weedy populations were intermediate. The above results were obtained from the first generation plants derived from seeds collected in natural habitats. Extant populations in the field showed a slightly different feature of their genetic structure as will be mentioned later.

Indigenous cultivars or land races are known to be highly polymorphic within populations in contrast to modern improved varieties which are nearly pure genetically. It was demonstrated that in metric characters including yield components as well as in isozymes, land races contained an appreciable amount of variability within populations (Oka 1988, pp. 67–72). Among land race populations, the amount of variation in characters and in isozymes was not always correlated. Morishima (1989) examined a lowland and an upland land race population managed by the same farmer from Yunnan, China. She found that the upland population was phenotypically more diverse (with plant types varying from Indica to Japonica and from upland to lowland types) than the lowland population (which was an Indica and lowland type), but the two populations contained similar levels of isozyme polymorphism.

3.5 Geographical distribution of genetic diversity

Genetic diversity in *O. sativa* is not evenly distributed geographically. Surveying variations in esterase isozymes, amylose content in the endosperm starch, seed size, and other traits, Nakagahra and his group advocated the areas extending from Assam to Burma, northern Thailand, and Yunnan as a diversity center of Asian rice (Nakagahra 1978; Nakagahra *et al.* 1986). Glaszmann (1987) made an extensive survey of Asian rice cultivars based on 15 polymorphic isozyme loci, and found a high isozyme diversity in West and South Asia including Burma, particularly in Bangladesh. Hakoda *et al.* (1990) examined isozyme variations in a number of Asian deepwater rices, and found that deepwater rices distributed in Bangladesh, India, and Burma were highly polymorphic, some approaching the Japonica type, although many were Indica types.

A large area including the so-called foothills of the Himalayas and neighboring districts has generally been considered as the diversity center of *O. sativa*. There are also independent diversity centers (possibly secondary centers) like the Jeypore Tract, India, although the geographical distribution of genetic diversity is not known in detail.

The diversity centers are characterized by the coexistence of diverse

Table 2

*Gene diversity and other parameter values estimated for populations of Asian common wild rice, *O. rufipogon*, and land races of *O. sativa**

Ecotype	No. of populations observed	Gene diversity			Polymorphic loci ^a	Average no. of alleles ^b	Heterozygotes frequency	Fixation index \bar{F}
		H_T	H_S	G_{ST}				
<i>O. rufipogon</i>								
Perennial	10	0.389	0.235	0.396	0.58	1.82	0.136	0.437
Intermediate	6	0.350	0.249	0.289	0.67	1.86	0.120	0.513
Annual	13	0.140	0.056	0.600	0.21	1.21	0.011	0.877
Weedy and hybrids	7	0.379	0.276	0.272	0.64	1.74	0.129	0.596
<i>O. sativa</i>	15	0.373	0.086	0.772	0.33	1.33	0.004	0.991

H_T = average gene diversity for all populations; H_S = average gene diversity within populations; G_{ST} = index of gene differentiation, $(H_T - H_S)/H_t$.

^a Proportion of polymorphic loci per population, to total loci.

^b Average number of alleles per locus, for all loci: data for six polymorphic loci; sample size, 21 plants per population on average.

types including Indica, Japonica, and intermediate types. In other words, the Indica–Japonica distinction is obscured in this area, genes and characters being associated with each other rather randomly. Degree of non-random association among isozyme loci in land races collected from the Himalayan foothills was assessed in terms of squared correlation coefficient of allelic frequencies R^2 (Hill and Robertson 1968). For a two-locus (A,B) two-allele (1,2) model, R^2 is defined as $R^2 = (X_1 - pq)^2 / p(1-p)q(1-q)$, where p and q are frequencies of alleles A¹ and B¹, and X_1 stands for frequency of A¹B¹ gametes. When averaged over six polymorphic isozyme loci, the value was 0.115 for 151 hill rices, and 0.382 for 101 control cultivars representing the whole of Asia, although their average gene diversities (H) were similar (0.471 and 0.493, respectively). Whether this situation was brought about by hybridization between distantly related cultivars, or represents a chaotic state preceding differentiation, is a matter of discussion. However, the tendency of subsistence farmers to prefer some diversity in their planting material could also play a role in preserving diverse types.

Geographical distribution of genetic diversity in wild rices is less affected by human factors. There are no recognizable diversity centers in *O. rufipogon* corresponding to those of cultivars. In general, a large amount of genetic variability can be preserved by various mechanisms, such as diversifying selection, different breeding systems, and other internal mechanisms. In domesticated plants, geographical variation and distribution of genetic diversity are jointly affected by both human factors and adaptation to environmental conditions. Therefore, we should be cautious in interpreting the origin of genetic diversity from the currently observed variation. A high diversity observed in a particular region cannot be interpreted only by the antiquity of cultivation in that region. To delimit the site or sites of incipient domestication as an historical incidence, archaeological evidence is needed in addition to biological information.

3.6 Multilocus system

We cannot examine natural variation without noticing particular patterns of association between states of different characters or between alleles at different loci. Such non-random association or multilocus covariation is called gametic disequilibrium. In rice, the variation between wild and domesticated forms, that between perennials and annuals in wild forms, and that between the Indica and Japonica types in cultivars are good examples for evolution of multilocus systems. A trend to gametic disequilibrium can be recognized at different levels of genetic organization ranging from phenotype to DNA variants. Questions arise as to whether or not ecotypical or varietal differentiation originally defined by phenotypic traits is associated with variation at the molecular level.

Let us summarize the evidence for gametic disequilibrium found in the *rufipogon-sativa* complex, within *O. rufipogon*, and within *O. sativa*. First, in the *rufipogon-sativa* complex, seed shedding, seed productivity and other adaptive traits are associated with one another generating a clear difference between wild and domesticated forms. Linkages found between shedding gene and other genes distributed differentially between wild and cultivated rices can be understood in the same context. Nevertheless, in molecular variation including isozymes, no clear tendency to disequilibrium has so far been detected despite the presence of a high degree of polymorphism in this plant group. Second, among races of *O. rufipogon*, various traits related to life history and mating system are significantly correlated with each other resulting in ecotypic differentiation into perennials and annuals. However, molecular variation did not show any association with this differentiation. In addition to ecological variation, *O. rufipogon* exhibits geographical variation. In this regard, characters (Morishima and Gadrinab 1987), isozymes (Second 1985) and rDNA (Sano *et al.* 1989) seemed to be associated with one another to some extent. Third, among cultivars of *O. sativa*, several characters and genes are strongly associated with each other exhibiting Indica-Japonica differentiation. In contrast to the other cases, the distinction of Indica-Japonica types was consistently found in isozyme, nuclear DNA RFLP, rDNA, ctDNA, and mtDNA.

Major factors in generating gametic disequilibrium are, as discussed by Hedrick *et al.* (1978), natural selection of coadapted sets of genes, linkage, founder effect (drift), migration, hitchhiking, and inbreeding. Inbreeding has an effect to slow the rate of decay of disequilibria which otherwise do not persist long. These factors are separable with rigorous tests. A coadaptive property of the disequilibrium is suggested by several features, such as a trend of disequilibrium to increase over time, consistent disequilibrium in different populations, epistatic interaction between loci for fitness, and presence of gene complexes composed of functionally or morphologically related genes. Our observational and experimental results indicate that natural selection for a coadapted set is a main agent in forming multilocus systems for wild vs. cultivated forms as well as for perennial vs. annual forms. An unanswered problem is whether or not linkages and gene complexes involving functionally related genes are the outcome of coadaptation favored during evolutionary processes.

4. LIFE HISTORY, MATING SYSTEM AND POPULATION DYNAMICS

4.1 Life history traits and habitat selection

Life history traits basically characterize fecundity/survivorship schedules of individuals. The Asian common wild rice, *O. rufipogon*, exhibits a

wide variation in life history traits within species, although its cultivated form *O. sativa* is usually grown as an annual crop. To learn about the variation in life history traits, we investigated various components of seed propagating ability, characters related to vegetative propagation, age at maturity (flowering date), stress tolerance, competitive ability, and others. Natural populations of *O. rufipogon* showed a continuous variation in these characters, but multivariate analysis of the data revealed a tendency toward differentiation into the perennial and annual types. This implies that the life history traits are intercorrelated among natural populations (Oka and Morishima 1967; Oka 1976; Sano and Morishima 1982). The two types can be easily distinguished in the field, although there are intermediate types.

The perennial (polycarpic) type is characterized by a low seed productivity, high regenerating ability, tall stature, late flowering, and strong competitive ability. In contrast, the annual (monocarpic) type is characterized by a high seed productivity, high seed dispersability, pronounced seed dormancy, short stature, early flowering, and stress tolerance. Reproductive allocation (the proportion of total seed weight to total plant weight) is a good measure of sexual (annual) vs. asexual (perennial) reproduction. Among the races of *O. rufipogon*, it varied from 40 to 60 per cent in annuals and 0 to 20 per cent in perennials (Morishima *et al.* 1984a). Plants having intermediate reproductive allocation were found, but seemed to be less frequent.

The propagating system of *O. rufipogon* in natural habitats was observed in some Thailand populations (Oka and Sano 1981, Morishima *et al.* 1984b). The numbers of seedlings and ratooned plants per unit area were counted at the recruitment stage in the early rainy season. It was confirmed that the populations judged to be perennial on the basis of morphological characters, regenerated mainly by ratoons, while those judged to be annuals regenerated mainly by seed germination, and intermediate ones by both means. The number of soil-buried seeds was larger in sites inhabited by annuals than in sites of perennials.

The habitats of *O. rufipogon* are usually sunny marshy places which are inundated during the rainy season in which vegetative growth is performed. However, the perennial and annual types seem to prefer different habitats. According to our field study in India and Thailand (Morishima *et al.* 1984a), the habitats of perennials are characterized by deeper water, less disturbance, and lower frequency of annual coexisting species in the community. The habitats of intermediate types are characterized by strong disturbance, low dominance of wild rice, and high species diversity in the community. The perennial populations are found in stable deepwater habitats which retain soil moisture throughout the year. The annual populations are in temporary swamps which are parched in the dry season.

Different conditions needed for reproductive success of the perennial and annual types were also suggested from the fate of wild rice populations that were experimentally introduced into new sites. Plants of annual, perennial, and hybrid populations were planted in several sites in Taiwan (Oka 1984) and in Okinawa, Japan, and were left under natural conditions. Annual populations disappeared within a couple of years in all sites. Perennial populations persisted but declined if subjected to competition with *Leersia hexandra* (Oka 1984). For the annual populations to persist, strong drought and/or disturbance in the dry season would be necessary.

In general, populations living in environments imposing high density-independent mortality will be selected for high reproductive allocation at the cost of their capacity to propagate under crowded conditions. On the other hand, populations living in environments imposing high density-dependent mortality will be selected for allocating more resources to vegetative growth at the cost of their capacity to use unstable habitats. This is predicted by the theory of r- vs. K-selection (Gadgil and Solbrig 1972). In the habitats of wild rice, drought and submergence would be two main natural stresses causing density-independent mortality. Tolerances to both stresses were higher in annual than in perennial types and were intercorrelated. Mortality resulting from water stress probably played a significant role in the differentiation of the annual from perennial type. The annual types would then be selected for higher tolerances to both drought and submergence. The contrasting resource allocation patterns thus associated with a certain set of life history traits in *O. rufipogon* represent intraspecific differentiation into K-(perennial) and r-(annual) strategists (Oka 1976).

In contrast, the African cultivated rice *O. glaberrima* and its wild ancestor *O. barthii* are both completely annual plants which do not survive after seed maturity. They therefore do not show the amplitude of variation in life history traits found in the Asian wild rices.

4.2 Variation in mating system

Rice plants have simultaneous hermaphroditic flowers and are wind pollinated. Cultivars of *O. sativa* as well as *O. glaberrima* are predominantly selfed, their outcrossing rate being usually 1 per cent or less. The common wild rices, on the other hand, are more or less allogamous. Their flowers have a larger number of pollen grains, bigger protruding stigmas, and a longer time period from flower opening to pollen emission than cultivars (Oka and Morishima 1967). These characteristics are expected to promote outbreeding.

Outcrossing rate was estimated by different methods. Direct estimation was based on the frequency of a single marker gene in the progeny (Oka and Chang 1961). Indirect estimation was attempted using a discriminant

function combining the measurements of time interval from flower opening to pollen emission, style and stigma length, and anther length (Oka and Morishima 1967). The ratio of within-line and between-line genetic variance of metric characters (Sakai and Iyama 1957; Sakai and Narise 1960) was also used. The estimates obtained by these different methods were generally comparable to each other (Oka 1988, p. 34). Estimated outcrossing rates varied markedly according to races of *O. rufipogon*. The perennial types showed a higher outcrossing rate (30–50 per cent) than the annuals (5–20 per cent). *O. barthii* (or *O. breviligulata*) had a 5–20 per cent outcrossing rate, and *O. glumaepatula* (the American race of common wild rice) showed a 20–60 per cent outcrossing rate (Oka and Morishima 1967). *O. longistaminata* has partial self-incompatibility and is outcrossed nearly completely (Chu *et al.* 1969b).

Using isozyme variation, Barbier (1989) obtained more reliable estimates of outcrossing rate for three Thailand populations of *O. rufipogon*. Multilocus estimates based on eight loci gave 5 per cent for an annual population and 51 per cent and 56 per cent for two perennial populations. In one of the two perennial populations, multilocus estimates computed separately for 27 mother plants ranged from 11 to 90 per cent, indicating variation of outcrossing rate within populations.

Variation in the mating system has been discussed in the context of pollination cost or sex allocation theory. Charnov (1982) has predicted that the ratio of resource allocation to male vs. female function, $m/(1-m)$, is adjusted at an optimum balance, which maximizes the reproductive potential. In allogamous populations, this ratio will be equal to c , which stands for the relative efficiency of male function in determining reproductive potential. In partly allogamous populations, when the rate of inbreeding depression is around 1/2, the relation can be approximated as: $m/(1-m) = c(1-s)$, where s stands for selfing rate. The ratio of anther weight to seed weight called ‘pollinating allocation’ may be taken as a measure of $m/(1-m)$. Among races of *O. rufipogon*, selfing rate was negatively correlated ($r = -0.84$) with ‘pollinating allocation’ (Oka and Sano 1981). In this case, c was estimated to be 0.19. This can be interpreted as evidence of a trade-off between resource allocation to male (pollen production) and female (seed production) functions.

Factors determining the rate of selfing vs. outbreeding are the level of inbreeding depression (Lande and Schemske 1985), cost of pollination (Charnov 1982), and mating opportunity (Moore and Lewis 1965). It was theoretically demonstrated that if the fitness of selfed progeny is greater than half of that of outcrossed progeny, there is selection for increased selfing rate in the population (Lande and Schemske 1985). In wild rice, there have been no explicit studies of inbreeding depression. Yet, some observations relevant to inbreeding depression have been made. Perennial populations contain pollen-sterile individuals that increase in selfed

progenies and decrease in hybrids (Hinata and Oka 1962). Seedling mortality, if tested in the first-generation plants derived from original seeds collected in the natural habitats, was higher in perennials than in annuals (Morishima and Barbier 1990). These observations suggest that deleterious mutations are accumulated in a heterozygous state in perennial populations. Accordingly, a higher level of inbreeding depression is expected in perennial than in annual plants. This could act as a constraint on the evolution of selfing in perennial populations, but the issue deserves more investigation in the future.

4.3 Association between propagating system and mating system, and its evolutionary significance

Annual wild rices are predominantly inbred and perennials are more or less outbred. Therefore, they show a significant association of propagating system with mating system. As shown in Fig. 2, outcrossing rate and reproductive allocation were negatively correlated ($r = -0.71$) among races of *O. rufipogon*. Various life history traits that are intercorrelated with one another were correlated with reproductive allocation and with pollinating allocation either positively or negatively (Table 3). The association of mating system and propagating system has an important consequence for the genetic structure of wild rice populations. Larger intrapopulational variability and smaller interpopulational variability in

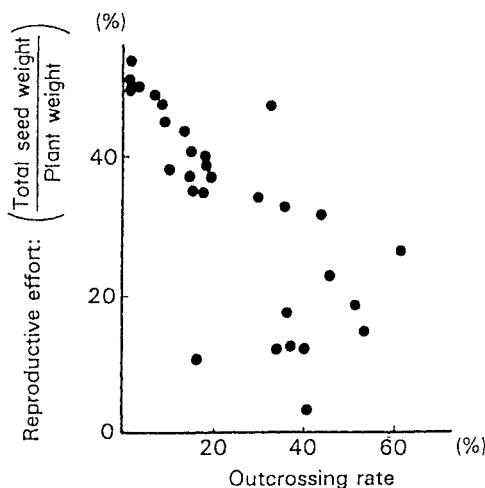


Fig. 2. Relationship between reproductive effort and outcrossing rate among races of Asian common wild rice *O. rufipogon*. Rate of outcrossing was indirectly estimated by a discriminant function combining three measurements (from Morishima and Barbier 1990).

Table 3

Measurements of life history traits in common wild rices and their correlations with reproductive allocation (RA) and pollinating allocation (PA)

Trait	Mean	Range	Correlation with:	
			RA	PA
Seed no./plant	481	5–950	0.94**–0.69**	
Panicle no./tiller no. (%)	69	5–95	0.87**–0.67**	
Seed fertility (%)	73	24–95	0.81**–0.61**	
Selfing rate (%)	71	5–95	0.71**–0.79**	
Single seed weight (mg)	21.6	13–30	0.61**–0.64**	
Seed dormancy index	2.3	1.4–2.6	0.43**–0.11	
Submergence tolerance (%)	29	0–86	0.47**–0.36*	
Drought tolerance index	2.1	0–4	0.38**–0.36*	
Regenerating ability of excised stem segments (index)	1.4	0–3	-0.48** 0.76**	
Tiller angle	56	15–85	-0.49** 0.52**	
Relative stubble weight (%)	16	3–30	-0.55** 0.57**	
Culm length (cm)	91	38–144	-0.60** 0.64**	

* $p < 0.05$, ** $p < 0.01$ ($n = 35$).

(From Oka and Sano 1981.)

the perennials than in the annuals mentioned earlier could be accounted for by the higher rate of asexual propagation and higher outcrossing rate of the perennial populations.

The evolution of selfing associated with sexual reproduction system can be inferred as follows. The primitive type of *O. rufipogon* was probably an outbreeding perennial adapted to deepwater habitats. It would have had a potential to preserve a large amount of genetic variability, including deleterious recessive genes. When such primitive types were subjected to drought stress, seed propagation would become an important means of survival. Seed production may have led to increased self-fertilization for several reasons, decreasing in turn the frequency of such recessive genes and the deleterious effect of selfing. Moreover, selfing would ensure seed set that may be uncertain in an outcrossed population. Thus seed propagation associated with inbreeding would have uncovered deleterious genes that were quickly eliminated.

4.4 Genetic structure and dynamics of natural populations

The genetic structure of natural populations has been considered to play an important role in mediating adaptive differentiation. To investigate

genetic structure and its change in local populations of *O. rufipogon*, we have monitored four perennials including intermediate perennial-annual and four annual populations located 5–20 km apart from each other in the Central Plain of Thailand. In order to compare the potential genetic variability and realized genetic variability, seeds and living plants were sampled from the same population at the seed maturation stage and juvenile stage, respectively, in our permanent study sites. Then, seed-derived and juvenile-derived plants were grown under the same conditions to examine isozymes and characters. Based on isozyme variation assessed at five loci, intrapopulational gene diversity was computed for each population (Barbier 1989). In contrast to the result obtained from seed-derived samples (which was consistent with that observed previously), juvenile-derived samples showed a relatively lower level of intrapopulational gene diversity in both perennial and annual types (Table 4). The intermediate perennial-annual population had the highest gene diversity.

The total genetic variation was partitioned into inter- and intrapopulational components by *F* statistics as shown in Table 4 (Barbier 1989). The overall inbreeding coefficient of an individual (F_{IT}) includes a contribution due to actual non-random mating within a population (F_{IS})

Table 4

Intrapopulational gene diversity (H_s) and partitioning of the total variation into inter- and intrapopulational components by F statistics in four perennial and four annual populations

	Seed sample	Juvenile sample
Perennials		
H_s^a	0.085–0.208	0.068–0.145
F_{IT}^b	0.420	0.042
F_{IS}^c	0.400	0.389
F_{ST}^d	0.033	0.363
Annuals		
H_s	0.198–0.350	0.129–0.356
F_{IT}	0.883	0.758
F_{IS}	0.868	0.792
F_{ST}	0.114	0.139

^a Ranges of H_s values (average gene diversity) found in four populations are given.

^b $F_{IT} = (H_T - H_I)/H_T$.

^c $F_{IS} = (H_S - H_I)/H_S$.

^d $F_{ST} = (H_T - H_S)/H_T$.

H_I = average gene diversity of an individual; H_S = average gene diversity within populations; H_T = average gene diversity for all populations.

and another contribution due to subdivision of the total population (F_{IS}). Both F_{IT} and F_{IS} were higher in the annuals than in the perennials as expected. F_{ST} , a measure of population differentiation, was higher in the annuals than in the perennials in the seed-derived samples as previously observed in other seed-derived population samples. In the juvenile-derived samples, however, F_{ST} was higher in the perennials than in the annuals. This implies that the perennial populations are more differentiated than the annual ones in existing populations.

Larger F_{ST} values in perennials than in annuals observed in the juvenile-derived samples could be explained by differences in gene flow and reproductive system. Gene flow in *O. rufipogon* may be mediated mainly by pollen flow in the perennials, and mainly by seed dispersal in the annuals. A rough estimate of the number of migrants per generation, Nm (N : population size, m : migration rate), was calculated using Wright's formula (1931); $F_{ST} = 1/(1+4Nm)$. The values estimated from seed-derived samples gave 1.94 for the annuals and 7.52 for the perennials. Those from juvenile-derived samples were 1.55 for the annuals and 0.44 for the perennials. Since pollen-mediated gene flow in the perennials becomes efficient only when the plants are recruited by seeds, the perennial populations that usually propagate asexually tend to preserve their respective genetic identity. The smaller F_{ST} found in the annuals would indicate that they are not very isolated from one another genetically despite the reduction in effective gene flow resulting from selfing.

Extinction and recolonization events would also have significant genetic consequences on the structure of local populations. The eight natural populations in Thailand we used in the above observations met different fates during the last decade. We observed one case of a population explosion (colonization into a newly provided vacant site), one case of gradual decrease (probably due to increased competition with coexisting perennial species caused by environmental change) and two cases of abrupt extinction (due to land reclamation). Those populations that experienced drastic changes were all annual types. The annuals are fugitive as expected from their adaptive strategy. Such extinction/recolonization events could decrease population differentiation under certain conditions as theoretically pointed out by McCauley (1991). Repeated extinction/recolonization from a common genetic source might have caused the structural pattern we found in the annual populations in Thailand. The observed population structure may therefore reflect not only gene flow expected on the basis of mating system but also other factors such as reproductive system and extinction/recolonization rates.

4.5 Intrapopulational structure on fine spatial scale

Different patterns of spatial distribution of genetic variation on a fine scale were found in two populations of *O. rufipogon* in Thailand. One was a typically perennial population growing in a ditch near Saraburi. Its habitat seemed relatively stable, and the plants propagated mainly by vegetative sprouting (ratooning). Seed samples were taken on an individual basis along the ditch (approximately 300 m long). The other was an intermediate perennial-annual population growing in a swampy place adjacent to a deepwater rice field in Ayuthaya. Its habitat was strongly disturbed, and plants propagated by both seeds and ratoons. Individual seed samples were collected along five transects set in two rectangular directions.

The genotypes of the maternal parents were determined from progeny arrays based on seven isozyme loci. Degree of similarities among genotypes was quantified in each population using a multivariate technique called 'pattern analysis' (Hayashi 1956), and microgeographical distribution of the scores was examined on a map. In the ditch population in which 39 different genotypes were detected among 60 plants, neighbors tended to have similar genotypes suggesting a family structure. In the intermediate perennial-annual population, which contained 42 different genotypes among 52 plants, no such trend was found, different genotypes being randomly distributed. This difference in microgeographical distribution of neutral genes could be explained by a difference in reproductive behavior of the two populations whose habitats are subjected to different degrees of disturbance.

In habitats showing pronounced environmental heterogeneity, the tendency to intrapopulational differentiation of adaptive traits was observed more clearly (Morishima *et al.* 1984a). Two such examples are shown in Table 5. The first population was found in a small pond (approximately 30 m in diameter) near Bhubaneswar, India. The periphery of the pond was occupied by a short-statured and early-maturing type, while in the deeper center was found a taller and late-maturing type; zonation was clearly recognizable. The second population was found in a roadside ditch in Chiangrai, Thailand. In the habitat of the second population, which appeared to be disturbed by man, the topography of the land was undulating and the populations were fragmented into several patches growing under different water depths. Seeds were sampled in bulk from each subpopulation or patch. Progeny tests in a uniform condition revealed that plants of different subpopulations differed significantly in various life history traits.

As shown in Table 5, plants from the shallower periphery of the Bhubaneswar population exhibited higher reproductive effort, shorter

Table 5
Habitat conditions, life history traits, and genetic parameters observed in parapatric subpopulations at Bhubaneswar and Chiangrai

Item	Bhubaneswar		Chiangrai		
	A (periphery)	B (center)	A	B	C
No. of plants tested	32	39	11	30	22
Water depth (cm)	0–50	50	0–5	0–5	0–10
Grade of disturbance ^a	1	1	2	3	2
Days to flowering	102	118	121	118	123
Anther length (mm)	2.0	2.9	2.7	2.6	3.0
Reproductive allocation (%)	52	20	15	43	20
Average gene diversity (H)	0.053	0.372	0.092	0.082	0.231
Heterozygote frequency	0	0.245	0.120	0	0.230
Fixation index (F)	1.0	0.649	0.510	1.0	0.501
Frequency of <i>Pox-1</i> ¹	1.0	0.510	0.950	1.0	0.770
Genetic distance	(A–B)	0.131	(A–B)	(B–C)	(A–C)
			0.049	0.037	0.066

^a Grade of disturbance: 1, slight; 2, moderate; 3, strong.
 (Reorganized from Morishima *et al.* 1984a).

stature, and earlier flowering than those growing in the deeper center area. Furthermore, the former showed lower gene diversity and higher frequency of a particular isozyme allele (*Pox-1*¹) than the latter. Plants from the strongly disturbed sites of the Chiangrai population were annual type while those from less disturbed sites in deeper water tended toward intermediate perennial–annual types.

When estimated on the basis of gene frequencies at six isozyme loci, the genetic distance between subpopulations was 0.131 for the Bubaneshwar population and 0.037–0.066 for the Chiangrai population. These values were much smaller than those found between allopatric populations collected nearby (0.193–0.361 for Bubaneshwar, and 0.113–0.145 for Chiangrai). Judging from their small genetic distances, subpopulations within a population do not seem to be fully differentiated. Whether those subpopulations have been derived from a common gene pool or they have come from immigrants from other perennial and annual populations is not known. But, it may be inferred that the gene pool of each population, once blended thoroughly, has been subjected to disruptive selection owing to habitat conditions causing differentiation.

The trend of periphery and center subpopulations to become differentiated into the perennial and annual types in adaptation to water condition

may be compared with the situation reported in *Veronica peregrina* in Californian vernal pools by Linhart (1974). In both cases, the pattern of character variations between subpopulations was similar to that found between perennial and annual populations of the Asian common wild rice. This suggests that life history traits are frequently selected as a coadapted set.

5. STERILITY BARRIERS DEVELOPED BETWEEN AND WITHIN TAXA

5.1 Hybrid sterility as an internal barrier

Species are recognized as groups of potentially interbreeding populations. Since presence of reproductive isolation permits diverged populations to coexist without losing their identity, the formation of isolating barriers has been considered to be a basic target for understanding speciation processes. Various kinds of postmating (or internal) barriers are observed in rice hybrids, i.e. F_1 sterility, F_1 inviability, F_1 weakness, and F_2 breakdown (Oka 1988, pp. 181–188). Of them, we will take up F_1 sterility, which is conspicuous not only between species but also within species. It has been extensively studied by a number of workers and a considerable amount of information has been accumulated on its genetic basis. Although our present knowledge is still sketchy, comparisons of various genic factors for F_1 sterility in rice hybrids gives a significant body of information for understanding the evolution of reproductive barriers.

5.2 Genic sterility

Cryptic structural differences in chromosomes has been proposed as a contributor to hybrid sterility observed in crosses without meiotic irregularities (Stebbins 1958; Shastry and Misra 1961). Recent molecular studies have shown that it is difficult to distinguish small chromosomal rearrangements from genic changes since numerous small changes in nucleotide sequences often prevail in the genomes of organisms and such changes might be associated partly with altered expression of certain genes. We will therefore consider that small chromosomal rearrangements may be analysed as genic changes in causing hybrid sterility.

Genes causing hybrid sterility are either gametophytic or sporophytic in action. Four possible genic models for F_1 sterility (Oka 1974a) were proposed assuming one or two loci as follows:

Model	Gene action	No. of loci	Genotype of F_1
A	Gametophytic	1	S/S^a
B	Gametophytic	2	$+_1/s_1, +_2/s_2$
C	Sporophytic	1	S/S^a
D	Sporophytic	2	$S_1/+_1, S_2/+_2$

In the first model, infertility results from allelic interaction at a single locus and a hybrid sterility gene (S) induces abortion of gametes carrying its opposite allele (S^a) in the heterozygote (S/S^a). Both homozygotes (S/S and S^a/S^a) are assumed to be fertile so that infertility only appears following hybridization between them. This type of sterility was first reported in *Nicotiana* (Cameron and Moav 1957) and it is divisible into gamete eliminator and pollen killer according to the manner of gametic elimination (Loegering and Sears 1963; Rick 1966). The second model, the duplicate gametic lethal hypothesis, assumes that recombinant gametes having both s_1 and s_2 , each of which is separately carried by the parents, become inviable in the hybrid ($+_1/s_1, +_2/s_2$). This was first reported in a varietal cross within *O. sativa* (Oka 1953). The genes could affect both male and female gametes or male gametes only. The third and fourth models assume one or two loci causing partial sporophytic sterility owing to complementary interaction. In spite of the occurrence of infertility, all alleles involved are expected to contribute equally to the next generation without any gametic selection. The third model was adopted for the corky hybrid in cotton (Stephens 1946), and the fourth model, the two dominant complementary gene system, has been frequently reported for F_1 inviability in various plants (Grant 1981). In rice hybrids, the F_1 inviability causing a crossing barrier between *O. longistaminata* and its related taxa (Chu and Oka 1970a) as well as cases of F_1 weakness found between varieties of *O. sativa* (Oka 1957a; Amemiya and Akemine 1963; Sato and Hayashi 1983) and between varieties of *O. glaberrima* and *O. barithii* (Chu and Oka 1972) are attributed to this type of gene interaction. Although the third model includes the possibility that the single locus resulted from the formation of a gene complex containing two dominant genes as assumed in the fourth model, these four models can be distinguished by conventional genetic analysis (Sano *et al.* 1979).

Since Kato *et al.* (1928, 1930) observed F_1 sterility barriers between the Indica and Japonica types of Asian rice, a number of studies have been reported on the mechanism of hybrid sterility in varietal crosses of *O. sativa*. As is the case in other organisms, there is general agreement that sterility barriers seem to have resulted from the accumulation of genes with profound effects of hybrid dysfunction as well as of genes with minor effects. The degree of hybrid sterility found within *O. sativa* varies rather continuously depending on the particular parental

combination (Terao and Mizushima 1939; Morinaga and Kuriyama 1958). Fertility is usually measured by the percentage of seed set and pollen stainability, but pollen stainability represents only a part of male sterility and is not necessarily related to embryo sac fertility.

After proposing duplicate gametic lethal genes as the genetic mechanism of F_1 sterility in rice, Oka (1953, 1955, 1956, 1957b) asserted that F_1 sterility relations observed between *O. sativa* varieties could be explained by assuming sets of duplicate gametic lethal genes with different fertilization success of pollen. If cultivars carried different kinds and numbers of such lethal gene sets, the complex features of their F_1 sterility relations could be explained. Later, the presence of duplicate gametic lethal genes was shown by Oka (1974a) using isogenic F_1 -sterile lines. Of five sterility genes carried by different cultivars, only two were identical indicating that various sets of duplicate lethals exist among cultivars. In addition, sterility genes introduced into isogenic lines conferred certation as their pleiotropic effect. Namely, in plants with $s_1 + s_2$, the fertilizing capacity of pollen grains with $+_1 s_2$ was higher than that of pollen grains with $s_1 + 2$ while that of pollen grains with $+_1 +_2$ was much lower. These genes could be evenly transmitted to the next generation only through megasporogenesis (Oka 1974a). Sterility genes were also detected between Japanese cultivars where they appeared to induce abortion in both female and male gametes as well as having pleiotropic effects of semi-dwarfism (Tanisaka *et al.* 1990).

Hybrid sterility resulting from allelic interaction (Model A) was also detected in some varietal crosses within *O. sativa* (Kitamura 1962; Oka 1964b) giving evidence for the presence of gamete eliminator and pollen killer. Further, Ikehashi and Araki (1986) noticed that this kind of sterility gene may be common among breeding materials. In addition, some varieties carry an allele, S_5^n , which is neutral and does not induce any gametic abortion against the other two alleles, S_5^i and S_5^j . The S_5^n gene located on chromosome 6 was first recognized as a gene that diminished seed infertility observed in varietal crosses. The two alleles, S_5^i and S_5^j , were assumed to induce abortion of female gametes although the genetic analysis was incomplete.

Anther indehiscence found in the hybrids between Japanese and Indonesian cultivars appeared to be controlled by dominant complementary genes fitting the fourth Model (Inukai *et al.* 1990). No case fitting the third model has been found in *O. sativa* so far. Thus, sporophytic sterility occurs less frequently than gametophytic sterility within *O. sativa*. However, there are inconsistent opinions as to which kind of genes for F_1 sterility (one locus or two loci system of the gametophytic type) are predominant within *O. sativa*. Since it is not easy to identify different sterility genes observed in different crosses, it is only safe to say, at present, that there exist a variety of genes with gametophytic action that

affect hybrid dysfunction within species. These genes give rise to the complex F₁ sterility relationships actually found in *O. sativa*. They may also serve as genetic material for further development of isolating barriers as observed between species.

Interspecific hybrid sterility was examined between the two cultivated rice species, *O. sativa* and *O. glaberrima* (Sano *et al.* 1979). The two species have the same AA genome and their hybrids show nearly normal meiosis but nearly complete sterility (Morinaga and Kuriyama 1957; Chu *et al.* 1969a). After introducing alien genes for sterility into both species reciprocally through repeated backcrosses, it was found that both species carried two different genes (S_1 and S_2) which fitted Model A. Each gene eliminated its opposite alleles from the other species acting as gamete eliminator in heterozygotes. Another sterility gene, S_3 , was found in the same strain of *O. glaberrima* and it acted as a pollen killer showing a pleiotropic effect of photoperiod sensitivity (Sano 1983). The observed sterility barriers between the two species were only partly explained by the three genes examined, suggesting that the barrier had been established through the accumulation of multiple disharmonious genes.

Dominant complementary genes (Model D) causing anther indehiscence seemed also to be involved in sterility between the two species (Sano 1985). In addition, another alien sterility gene was detected during the backcrossing experiment and appeared to fit Model C which gave sporophytic sterility due to allelic interaction (Sano 1986). However, this was proved to have resulted from a spontaneous occurrence of a translocation during backcrosses since the heterozygote formed a quadrivalent in meiosis. Thus, Model C is explained also by the presence of chromosomal interchanges. If chromosomal arrangements such as small translocations causing infertility prevail, they could have been easily detected by this kind of backcrossing experiment. In other words, small interchanges or chromosomal arrangements, which have been supposed to be a contributor to sterility barriers in various plant taxa, might be ruled out in rice since no such case was detected except for this one mutational event.

The relationship between the two cultivated rice species is much more distant than that between varietal groups within species. The sterility barriers developed between the two rice species may reflect the genetic divergence between their respective wild ancestors. For understanding the evolution of reproductive barriers, knowledge of genetic basis or sterility barriers between cultivated rice species and their wild ancestors would be more informative. However, the wild ancestor *O. rufipogon* generally produces male fertile hybrids when crossed with cultivars (Hinata and Oka 1962; Chu *et al.* 1969a). Accordingly, sterility barriers observed among Asian cultivars (*O. sativa*) might be regarded as the result of their differentiation (Oka 1974a). Another possible explanation

might be that sterility barriers within species reflect the genetic divergence within their wild ancestors.

As a general tendency, the domestication processes in crops seem to have occurred at the infraspecific level as recognized from the fact that cultivated and ancestral wild species possess common primary gene pools. An interesting question arises as to whether or not domestication pressure has given rise to reproductively isolated subpopulations as found between varietal groups in rice (Terao and Mizushima 1939; Morinaga and Kuriyama 1958). Recently, a strain of *O. rufipogon* from Malaysia was found to carry S_6 (near a photoperiod sensitivity gene Se_1 on chromosome 6), which showed seed infertility and fully stainable pollen grains in the heterozygote (S_6/S_6^a). However, the S_6 gene was proved to act as a gamete eliminator since no S_6^a gene was transmitted to the progeny of the heterozygote (Sano 1989b). It also gives rise to a marked segregation distortion of the genes flanking it. The distribution of this sterility gene in the wild ancestor remains to be studied.

Of the genes for F_1 sterility found within and between rice species, those with gametophytic action are widespread. No duplicate gametic lethal genes (Model B) has been detected between species, but this might be due to technical difficulties of detection rather than due to an abundance of allelic sterility at a single locus (Model A). Sterility genes with strong effects tend to be extracted by backcrosses. Of the genes examined, those with the strong effects were gamete eliminators and this kind of gene seemed to be of wide occurrence between rice species. Genes causing F_1 female sterility seem to be less frequent than those causing male sterility. This agrees with the general trend for male sterility to occur more frequently and strongly than female sterility in plant hybrids.

5.3 Cytoplasmic male sterility

Hybrid sterility may also be caused by cytoplasmic differences that could operate as an effective internal barrier. Cytoplasms were extensively surveyed in rice after the success of hybrid rice breeding in China. So far, 95 strains have been found to carry cytoplasms causing male sterility (Shinjyo 1984; Virmani and Shinjyo 1988). Out of 95, 17 were found from *O. sativa* and 78 from its wild relatives. The detection of cytoplasmic effects depends on the presence or absence of a restoring gene(s) involved in the cross. The cytoplasm of *O. sativa* generally induces male sterility when combined with the nucleus of *O. glaberrima* (Yabuno 1977; Sano 1985). The difference in cytoplasms is detected only by comparing the effects of restorers. Restorers had gametophytic or sporophytic effects. The genic system involved in restoration seemed to be complex such as that found in genic sterility barriers (Virmani and

Shinjyo 1988), which makes it difficult to compare different cytoplasms from various strains.

Shinjyo (1984) examined 130 cytoplasms from four wild species by substituting their nucleus with that of a Japonica cultivar, Taichung 65. The cytoplasms from *O. rufipogon* frequently induced male sterility (61 out of 95 strains), while the occurrence of male sterility was rare in *O. glumaepatula* (one out of 25), *O. meridionalis* (none out of six) and *O. longistaminata* (none out of four). Of these 62 strains with male sterile cytoplasms, only eight carried restorer genes of the gametophytic type and the remaining 54 carried those of the sporophytic type. The prevalence of sporophytic restorers contrasts to the prevalence of gametophytic sterility genes in genic hybrid sterility. Most of the restorers examined so far behave as dominant genes and, as a result, sterile segregants occur only in the F₂. This suggests that cytoplasmic sterility does not work as a major mechanism of F₁ sterility. However, it is uncertain why recessive or gametophytic restorers are less frequent than dominant or sporophytic restorers in rice.

When combined with the nucleus of Taichung 65 (*O. sativa*), the cytoplasms of closely related ancestors seem to induce male sterility more frequently than those from distantly related taxa. The divergence in cytoplasmic DNAs was detected with comparisons of ctDNA (Dally and Second 1990) and plasmid-like mtDNAs (Kadowaki *et al.* 1988). It is noteworthy that cytoplasmic sterility is not necessarily associated with the degree of divergence estimated from organellar DNAs. The diversified male sterile cytoplasms in the *sativa*-*rufipogon* complex seem to be detected by the diversified restorers present in the cultivated form. Cytoplasmic male sterility has been associated with alterations in mtDNAs (Kadowaki *et al.* 1986) and/or plasmid-like DNAs detected in mitochondrial fractions (Yamaguchi and Kakiuchi 1983; Nawa *et al.* 1987), but the molecular mechanism of male sterility is unknown in rice.

5.4 Formation of profound barriers through gene complexes

An important question in evolution is how a species develops a profound reproductive barrier characteristic of a fully established species. It has been suggested that the genomes of rice have an internal mechanism which helps alien genes increase in hybrid populations and probably leads to the development of F₁ sterility relationships among derived populations (Oka 1974a). The duplicate gametophytic lethal genes, as mentioned before, often have a certation effect, which increases the fertilizing capacity of pollen grains with an alien gene. If such certation effects are common, sterility barriers may easily be established by accumulating different sets of duplicate lethals in isolated populations. Such duplicated lethals may be common because of the genetic redundancy frequently

observed in plant genomes. As a result, partially isolated subpopulations could emerge from an ancestral fertile population.

Another evolutionary change in sterility genes was proposed from the geographical distribution of a gamete eliminator found in tomato by Rick (1971). Ge^c and Ge^p were assumed to have been derived from the neutral allele, Ge^n , and the same story might be adopted in the case of S_5^n in rice. However, in these cases, there is no need to involve an internal mechanism to accelerate the fixation of the newly derived gene. If gamete eliminator (S) emerged from the opposite allele (S^a) as a mutation, it is expected to increase rapidly in populations as the result of no contribution of the opposite allele to the progeny of the heterozygotes. Compared with sporophytic sterility, an advantage of gametophytic sterility may be that it minimizes its evolutionary cost. Thus, pollen killers would automatically increase in frequency if they result in no reduction in seed fertility (Sano 1983).

The strongest sterility gene detected in rice was gamete eliminator. Gametic eliminators seem to be of wide occurrence between distantly related taxa (Sano 1983; Scoles and Kibirige-Sebunya 1983). Additional genes that modify their expression also exist between and within species. Such modifiers make the degree of sterility fluctuate depending on species and strains used for crosses (Loegering and Sears 1963). It was shown that gamete eliminator, S_1 , in rice could be made from accumulation of a pollen killer and its modifiers when pollen killer and the modifiers are linked (Sano 1990). S_1 essentially acted as gamete eliminator in an Indica cultivar but as a pollen killer in a Japonica cultivar. Intergrades between gamete eliminator and pollen killer were observed depending on the modifiers involved in the linked region during the procedure of backcrosses. The segment containing S_1 also carried a modifier responsible for determining the degree of female abortion.

Furthermore, a marked suppression of crossing-over took place around S_1 suggesting the presence of an inversion. This is consistent with an assumption by Oka (1968) that chromosomal rearrangements might be involved between the two species as revealed by preferential pairing of chromosomes in the tetraploid hybrid. This will make the segment containing S_1 and its modifiers behave as a gene complex.

Thus, the formation of gene complexes may give rise to the establishment of profound barriers. Depending on the accumulation of different genes involved in the gene complex, they can respond to different species and strains in different ways and show the complex features of compatibility relationships actually found. Differential transmission of alien chromosomes often found in wheat (Endo and Tsunewaki 1975; Finch *et al.* 1984) is well explained by assuming that alien chromosomes carry sterility genes similar to gamete eliminator or pollen killer. The alien chromosomes of wheat are much differentiated from their homoeologous chromosomes.

In these cases, the manner of gametic elimination varies from species to species and often shows extreme abnormalities such as chromosomal breakages, suggesting that alien chromosomes accumulate additional modifiers and/or other disharmonious genes. These considerations lead us to think that pronounced sterility barriers have been established in a gradual process with many steps rather than a macromutational event.

Of particular interest in rice is a frequent association between sterility genes and photoperiod sensitivity genes which determine its adaptive range in different localities. S_1 , S_3 , S_5 and S_6 were all linked loosely or tightly with photoperiod sensitivity genes. Therefore, gametophytic selection is able to influence the traits in the sporophytic generation and vice versa, and the fixation of these sterility genes will be accelerated by premating isolation owing to the difference in their blooming time. The formation of such gene complexes could result in a more rapid spread of sterility genes in a population than would be anticipated on the basis of its selective differential. This serves as an additional internal mechanism for the development of reproductive isolation.

6. INDICA-JAPONICA DIFFERENTIATION OF RICE CULTIVARS

6.1 Occurrence of the two types

O. sativa cultivars generally show a tendency to be differentiated into two types, which were named subspecies *indica* and *japonica* by Kato *et al.* (1928). These types have been recognized as Hsien and Keng by the Chinese people since the Han dynasty (about 1800 BP). The two types can be clearly differentiated by examining the association pattern of certain diagnostic characters, such as $KClO_3$ resistance, cold tolerance of seedlings, apiculus hair length, and phenol reaction of lemma and palea (Oka 1958). These characters are strongly correlated when an arbitrary sample of rice varieties is examined, as Japonicas have stronger $KClO_3$ resistance, cold tolerance, and longer apiculus hairs than Indicas, and many of Indica and Japonica varieties show positive and negative phenol reaction, respectively. When the sample varieties are classified into the two groups, however, these characters are almost uncorrelated within each group.

Drought tolerance of seedlings is also useful for distinguishing between the Indica and Japonica types, but has not been used often since the test is laborious. The principal component analysis of correlations between these and six other characters (some being not diagnostic) showed that some 100 cultivars from different Asian countries were clearly divisible into the two types (Morishima and Oka 1981). Such multivariate analysis

involving no external criteria serves as evidence for the presence of the two types as natural varietal groups.

The Indica and Japonica types differ in many characters and genes, but show overlapping variations and there is no single criterion by which the two types can be distinguished with certainty. Essentially, the occurrence of the two types represents a variation trend of rice cultivars. Some cultivars have a recombined or an intermediate state of diagnostic characters. A discriminant function combining measurements of KClO_3 and other diagnostic traits gives a low probability of misclassification, yet a few varieties remain intermediate or atypical (Oka 1988, pp. 144–145). Such a trend of varietal differentiation is not found in the African rice, *O. glaberrima* (Morishima *et al.* 1962).

Phenol reaction is controlled by a dominant gene, *Ph* (chromosome 4), which is carried by about 90 per cent of Indica varieties. Apiculus hair length is controlled by a dominant gene (*Aph*, chromosome 6) and some modifiers (Sato 1985; Sato *et al.* 1987), but this character is not expressed in glabrous varieties (having *gl-1*, chromosome 5), which are rather frequent among tropical upland Japonicas. KClO_3 resistance and cold tolerance seem to be controlled by a few genes and their interactions, which remain unknown.

People tend to generalize that Indicas have a long grain and Japonicas a short or round grain, but the classification by length-width ratio of grains is unreliable (Sato 1991). Also, many rice workers like to assume that the Indica and Japonica types are separated by F_1 sterility. But this is a misbelief. Some Indica–Japonica hybrids are fertile, while other crosses within Indica and within Japonica may show F_1 sterility. Varietal classification by F_1 -pollen fertility relationships results in a high probability of misclassification (Morishima and Oka 1981).

6.2 Variation in isozymes and DNA

At a number of isozyme loci, allele frequencies differ between the Indica and Japonica types, although there is no locus whose alleles characterize the two types without exception. The situation is similar to that observed for phenol reaction.

Glaszmann (1987) observed more than 1600 native varieties from different Asian countries for 15 isozyme loci and analyzed the data using a multivariate technique (factor analysis of correspondence). The results showed that 95 per cent of the varieties fell into six groups, each having similar genotypes, while the remaining 5 per cent scattered over intermediate positions. Of the six groups, group I (53 per cent) corresponded to the Indica type classified by diagnostic characters, and group VI (27 per cent) to the Japonica group. Groups II to IV, rather close to the Indica type, each tended to represent a special ecotype, i.e.

II (7.3 per cent) Aus and some Boro, III (0.4 per cent) photoperiod-insensitive deepwater rice, IV (0.7 per cent) Rayada or long-duration deepwater rice, and V (6.3 per cent) diverse varieties including Basmati or a high quality rice with special flavour. Groups III and IV were endemic to Bangladesh.

This study involved no external criterion and can be compared with the principal component analysis of character variations mentioned above. It has also shown that a greater part of rice cultivars are classifiable into the Indica and Japonica groups leaving some atypical ones as intermediates. Accordingly, a particular association pattern is found among diagnostic characters and genes for isozymes and other phenotypic traits that are distributed differently between the Indica and Japonica types (Sato *et al.* 1990a).

Morishima and Gadrinab (1987) have evaluated the likeness of a strain to the Japonica or Indica type in isozymes as follows. When a strain has allele i at locus j , the respective frequencies $P_{ij(\text{Indica})}$ and $P_{ij(\text{Japonica})}$ obtained from a set of standard Indica and Japonica varieties give an index of Japonica (or Indica) likeness as $P_{ij(\text{Jap})}/(P_{ij(\text{Ind})} + P_{ij(\text{Jap})})$. When a strain is examined for n loci, its general likeness to Japonica (or Indica) is shown by the means of indices for n loci. The survey by this method showed that the Indica and Japonica types are distinguished well by isozymes at several diagnostic loci (Table 6). Their wild progenitor, the Asian common wild rice (*O. rufipogon*), is distributed over a wide range from Indica to Japonica types, strains with intermediate index values being predominant.

Recently, in addition to isozymes, varietal differences at the DNA level have also been reported. Wang and Tanksley (1989) analyzed RFLPs of

Table 6

Distribution of isozyme scores showing likeness to the Japonica type (X_{Jap}) in Asian common wild rices and cultivars

Plant group	Isozyme score					No. of strains
	0.1	0.3	0.5	0.7	0.9	
Wild rice	2	7	26	12	4	51
Indica	11	17	3	1	—	32
Japonica	—	—	2	2	28	32

$X_{\text{Jap}} = \frac{1}{n} \sum P_{ij(\text{Jap})}/[P_{ij(\text{Ind})} + P_{ij(\text{Jap})}]$, where P_{ij} is the frequency of allele i at locus j and n is the number of loci examined. Seven loci (*Acp-1*, *Cat-1*, *Est-2*, *Est-9*, *Pgi-1*, *Pgi-2*, and *Pox-2*) were examined, which show different allele frequencies between the Indica and Japonica types.

70 cultivars of *O. sativa* using five restriction endonucleases and 10 single-copy probes. About 80 per cent of the cultivars assayed could be uniquely distinguished from one another, and a dendrogram constructed from the data produced varietal clusters showing a good agreement with the six isozyme groups given by Glaszmann (1987). Tanaka *et al.* (1989) analyzed 124 native cultivars from various Asian countries using 38 RFLP probes, and by multivariate analysis detected two major varietal groups corresponding to the Indica and Japonica types, respectively. Sano and Sano (1990) surveyed 76 cultivars with regard to the spacer length of rDNA. They found that Indica, Japonica and 'Javanica' types tended to show different spacer length types. Kadokawa *et al.* (1988) surveyed polymorphism of mitochondrial plasmid-like DNAs using about 100 cultivars, and found that Indica and Japonica types tended to differ in banding patterns. Dally and Second (1990) examined ctDNA of 74 rice cultivars with various restriction endonucleases, and detected 10 different restriction patterns. In their study, Indica and Japonica types generally showed different restriction patterns.

When the Asian common wild rice, Indica and Japonica cultivar groups were compared with regard to average gene diversity, $H = 1/n \sum_j (1 - \sum_i x_{ij}^2)$, for isozymes (11 loci, 40 alleles), the wild rices showed the highest diversity, followed by Indica cultivars, and Japonicas with lowest diversity (Table 7). In rDNA (Sano and Sano 1990) and ctDNA (Dally and Second 1990), the order of diversity represented by H' (diversity index) was Wild > Indica \geq Japonica. In contrast, when genes for the coloration of apiculus and stigma were examined, the order of diversity was reversed: Wild < Indica < Japonica. Probably, genes for

Table 7
Average gene diversity (H) estimated in Asian common wild rice and Indica and Japonica cultivars

Plant group	Isozyme genes ^a		Coloration genes ^b	
	Strains	H	Strains	H
Wild rice	56	0.476	49	0.052
Indica	82	0.338	64	0.211
Japonica	85	0.145	66	0.384

^a 11 loci, 40 alleles (*Acp-1*, *Amp-3*, *Cat-1*, *Est-2*, *Est-9*, *Pgd-2*, *Pgi-1*, *Pgi-2*, *Pox-1*, *Pox-2*, and *Sdh-1*).

^b Seven loci, 15 alleles for apiculus and stigma coloration (*C*, *A*, *P*, *Ps-1*, *Ps-2*, *Ps-3*, and *I-Ps-3*).

$H = \frac{1}{n} \sum_j (1 - \sum_i x_{ij}^2)$, where x_{ij} = frequency of gene i at locus j , and n = number of loci examined.

morphological traits would also show the same trend as coloration genes if their frequencies in different groups were known.

These observations suggest that gene diversity has either decreased or increased with the domestication of wild progenitors and differentiation of domesticates between the Indica and Japonica types. The Indicas are left midway between the wild and Japonica groups in both decreasing and increasing trends.

6.3 Possible role of upland culture in the evolution of Japonica type

The Japonica cultivars are distributed from the tropic to northern temperate zones and tend to be differentiated into tropical and temperate varietal groups (Sato 1987), while Indicas are in the tropic and subtropic zones. Japonicas generally have higher cold tolerance than Indicas. In Yunnan, China, the areas below 1400 m are planted mainly to Indicas and those above 1800 m (up to 2200 m) to Japonicas, the areas in between being planted to a mixture of both. A survey of an altitudinal cline of isozyme variation also showed a similar trend for rice cultivars in Nepal (Oka 1988, p. 155; Sano, R. *et al.* 1985).

On the other hand, in tropical and subtropical regions, Japonicas are often grown as upland rice at low altitudes (Sato 1986). The same trend is also found in China. A survey of *O. sativa* native varieties in Africa with regard to isozymes also revealed that those grown in upland fields tended to have genes characterizing the Japonica type (Kochko 1987). This supports the hypothesis by Wang *et al.* (1984) that the Japonica type evolved in upland culture of incipient domesticates.

When an Indica–Japonica hybrid population is grown without deliberate selection, genes derived from the Indica parent generally increase against those from the Japonica parent (Oka 1988, pp. 172–174). In this regard, when an F_2 population ($\text{Ac}130 \times \text{Ac}221$; $\text{Ac}130$, Taiwan lowland Indica; $\text{Ac}221$, Philippine upland Japonica) was divided into two blocks, one grown in upland and the other in lowland conditions, the frequencies of genes derived from the Japonica parent were significantly higher in the upland than in the lowland F_5 population (Sato 1990). This may serve as circumstantial evidence for the hypothesis of evolution of the Japonica type in upland culture.

6.4 Domestication and differentiation

It has been discussed whether the Indica and Japonica types are monophyletic or diphyletic, in other words whether the differentiation has occurred after or before the start of domestication of wild progenitors. According to the diphyletic hypothesis, there should be wild rices differentiated into the two types or at least destined to become the Indica

and Japonica types when domesticated. In contrast, the monophyletic hypothesis assumes that the two types share the same wild progenitor.

The wild and cultivated rices differ in seed shedding rate and several other characters and have different adaptive syndromes, as mentioned earlier. Intermediate wild-cultivated plants were obtained from the Jeypore Tract, India, some being cultivated and others growing wild. Combining measurements of several characters distinguishing wild from cultivated plants, discriminant scores were computed, and were compared with scores given by another discriminant formula for classifying Indica and Japonica types (Oka and Chang 1962; Oka 1974b). The intermediate wild-cultivated plants scattered according to these discriminant scores together with wild and cultivated control strains showed that those close to the wild rice were not differentiated into the Indica and Japonica types, yet as they approached the status of cultivars they became gradually differentiated into the two types (Fig. 3). This suggests that Indica-Japonica

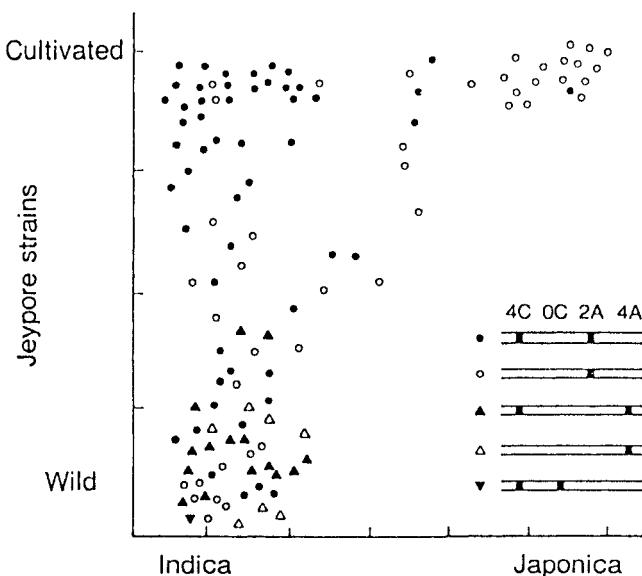


Fig. 3. Wild controls, intermediate wild-cultivated strains from Jeypore Tract, India, and cultivated controls scattered according to the scores given by two discriminant formulae, one (abscissa) for classifying the Indica and Japonica types of *O. sativa* cultivars, and the other (ordinate) for wild and cultivated forms. Alleles for peroxidase isozyme at *Pox-1* and *Pox-2* are also shown: 4C, *Pox-2^{4C}*; OC, *Pox-1^{OC}*; 2A, *Pox-1^{2A}*; and 4A, *Pox-1^{4A}*. The discriminant formula for Indica-Japonica differentiation was based on $KClO_3$ resistance, cold tolerance, apiculus hair length, and phenol reaction, and that for wild and cultivated forms was based on seed shedding rate, single grain weight, spikelet number per panicle, rachis number per panicle, and seed dormancy index (from Oka 1974b).

differentiation has proceeded as the plants are domesticated. Indeed, KClO_3 resistance, cold tolerance, and apiculus hair length were uncorrelated among the Jeypore strains growing wild and showed weak but significant correlations as they approached cultivars.

An Indian wild strain (intermediate perennial-annual type) was crossed with a typical Indica and a typical Japonica cultivar, respectively, and the hybrid populations were grown separately in pedigrees without any selection (Oka and Morishima 1982). Seeds for raising the next generation were taken from bagged plants. Their F_8 lines, representing F_7 plants taken at random, were examined for characters distinguishing between wild and cultivated forms and between Indica and Japonica types. Among those approaching a cultivated form, a few Japonica-like lines were found in the wild \times Indica population, and a few Indica-like lines in the wild \times Japonica population. The wild parent was resistant to KClO_3 like the Japonica parent, and non-tolerance to low temperatures like the Indica parent, but the F_7 plants showed a full range of KClO_3 resistance and cold tolerance. The result of this experiment suggests the potential of wild rices to evolve both the Indica and Japonica types, even though the genetic basis of the difference between the two remains without ample explanation. These experimental results are in support of the monophyletic hypothesis of the Indica and Japonica types.

The diphyletic hypothesis of the two types was proposed first by Chou (1948) who found a Japonica-like weedy deepwater rice in a lake in Anhui Province, China (now extinct). Second (1982), finding that some Chinese wild-rice strains had Japonica-like zymograms, also put forward this hypothesis. The common wild rices do not seem to be clearly differentiated into the two types, but some Chinese strains are Japonica-like in KClO_3 and cold resistances (Morishima 1986; Morishima and Gadrinab 1987). It was also found that some Chinese wild strains were of Japonica-specific type in rDNA (Sano *et al.* 1989) as well as in ctDNA (Dally and Second 1990). As already mentioned, Indicas and Japonicas tend to differ not only in nuclear DNA but also in organellar DNA. These results may be regarded as supporting evidence for the diphyletic hypothesis.

Weedy rices tend to be differentiated into the two types. Our study, though it is far from an extensive survey, showed that with regard to isozymes, weedy rices found in India are Indica-like, those in China and Korea are Japonica-like, and those in Nepal and Brazil are intermediate. In diagnostic characters, however, they varied from typical Indica to intermediate Indica-Japonica type (Tang and Morishima 1988). This is probably because there is gene flow from coexisting cultivars.

Archaeological excavations in Zhejiang Province, China, have brought to light Neolithic rice grains of about 7000 BP. Judging from their variation in grain shape, You (1986) considers that the ancient rice grains

are a mixture of many Hsien (Indica) and fewer Keng (Japonica) types. In later ages (4000–5000 BP), Japonica-like grains increase to half or more. Even though visual judgment is unreliable, we recognize that the incipient domesticates in the lower Yangtze basin were differentiated into the two types. The spikelet length and width of the ancient rice grains show large variances and tend to be negatively associated. Similar variation patterns are found in some land-race populations from the hilly area of tropical Asia containing Indica-like and Japonica-like plants.

Neolithic rice grains similarly as old as 7000 BP were excavated in India (Maharaga site, Utter Pradesh) and in Thailand (Non-Nok-Tha; Oka 1988, pp. 129, 134), but variations within samples were not reported. It has been shown that the plant opal (plant-derived silica body), which stays preserved in soil for a long time, shows some difference in shape between the Indica and Japonica types (Sato *et al.* 1989). Studies of plant opals may shed more light on the Indica–Japonica differentiation of incipient domesticates. The findings of similarly old rice remains in China, Thailand, and India suggest the diffused origins of rice domestication over an area, as proposed by Harlan (1971) in the context of his ‘non-center’ theory.

6.5 Putative internal mechanisms of Indica–Japonica differentiation

We have some circumstantial evidence for internal mechanisms of differentiation involved in the genome of *O. sativa*. In distant crosses, in addition to segregation distortion, a tendency for parental gene combinations to increase against recombinations has been observed (Oka 1988, pp. 174–178). New evidence for this was obtained in the F₅ of Ac130 (Taiwan Indica) × Ac221 (Philippine Japonica) grown in bulk. Recombinations between genes at two independent loci, *wx*/*+* (for glutinous endosperm) and *Rc*/*rc* (for red pericarp coloration) were significantly fewer than the frequencies expected from random combination, and parental combinations were more numerous than the expected frequencies.

The F₂ and F₅ populations from another cross, Ac419 (Indica, PTB10, from India) × T65 (Japonica, Taichung 65, from Taiwan) were observed for the association trends of genes for some isozymes and marker traits and a few Indica–Japonica diagnostic characters (Fig. 4; Sato *et al.* 1990a). Out of 66 possible combinations between 12 genes and characters, nine were found in F₂, which were largely due to gene linkage (e.g. *Est-2*, *Pgi-2*, *c* and *Aph* belong to chromosome 6). In F₅, seven associations between independent genes and characters that were not found in F₂ were recovered, all showing the same directions as found among cultivars, of which 51 of 66 combinations showed non-random association reflecting their Indica–Japonica differentiation. Since the hybrid populations were raised by the single-seed descent scheme, the recovery of non-random

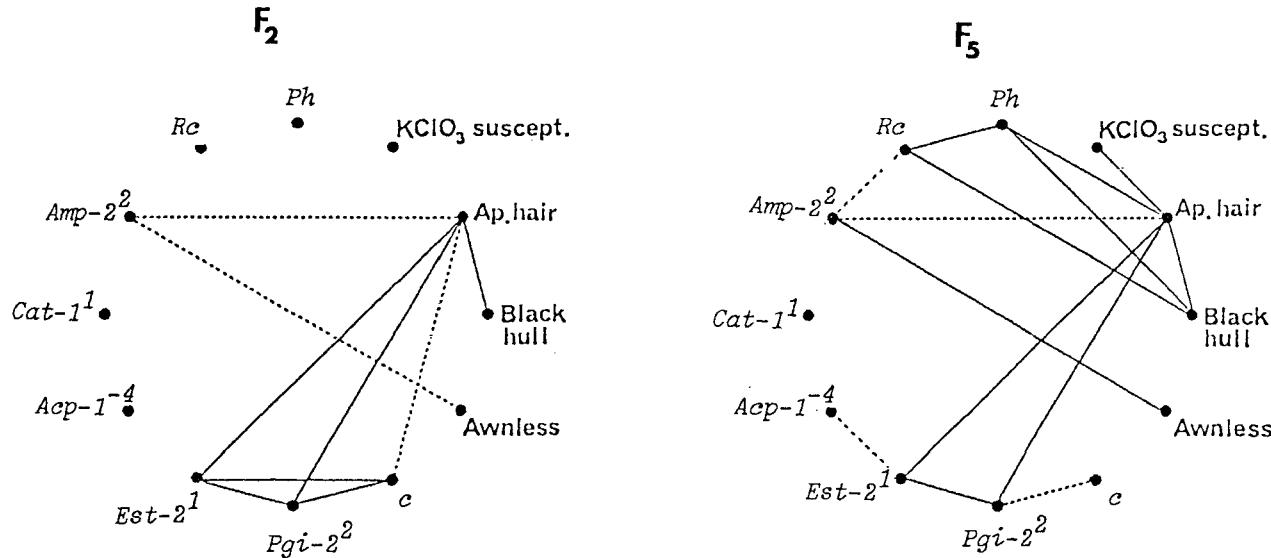


Fig. 4. Association trends found among Indica-Japonica diagnostic characters, genes for isozymes and some marker traits in the F_2 and F_5 populations from a cross, T65 (Japonica, Taiwan) \times 419 (Indica, P. T. B. 10, India). Solid and dotted lines show significant association at 1 per cent and 5 per cent levels, respectively. The association trends in F_2 are due to linkage of relevant genes. Those newly added in F_5 seem to be due to gametic selection. Among cultivars from different Asian countries, most of these genes and characters are associated, showing the same direction as found in F_5 (from Sato *et al.* 1990a).

associations in F_5 is most probably due to gametic selection caused by differential fertilizing ability of gametes with different genotypes.

In crosses between isogenic F_1 -sterile lines, pollen grains having an alien sterility gene were found to have much higher fertilizing capacity than those having the original gene (Oka 1974a). This certification helps alien genes to increase in hybrid populations and to develop F_1 -sterility relationships among *O. sativa* varieties. In addition, the advantage of alien pollen in mixed pollination, called 'alien pollen primacy,' was found between Taichung 65 (Japonica) and IR36 (Indica) (Sato 1988). A trend to preferential fertilization of female gametes was also detected. In the Ac130 (Indica) \times Ac221 (Japonica) cross, when heterozygotes for *gl-1* (for glabrous character, from Ac221) were backcrossed in different directions, ovules having $+^{gl}$ were fertilized with a frequency twice as high as those having *gl* by the same pollen having *gl* (Sato *et al.* 1990b). In this case, pollen grains having $+^{gl}$ and *gl* showed no certification effect. This implies that female gametes with different genes have unequal chances of being fertilized by the same pollen grains. This also results in gametic selection.

These observations indicate that gametic selection prevails in distant crosses of rice. Such selection may result in gametic disequilibrium as particular gene combinations may be favored. In some cases, the implication of the observed trend of selection is not clear, but in other cases in which parental gene combinations are favored, gametic selection is directional and may help hybrid populations to become differentiated into the parental types. Gametic disequilibrium is known in various organisms (Hedrick *et al.* 1978), but a situation favoring certain gene combinations directionally such as is found in rice seems to be uncommon. Probably, the whole mechanisms is complex and has yet to be clarified, but some of the findings so far reported suggest that there may be internal mechanisms promoting Indica-Japonica differentiation.

When Indica and Japonica varieties grow sympatrically, they will hybridize with a certain frequency that will tend to blend genic differences. But the evidence presented above suggests that the progeny may become differentiated again into parent-like types. The isozyme loci that are neutral would be linked with fitness-related loci, and their disequilibrium results from selection caused by coadaptation of independent genes. The rate of change in gene frequency would be low, but the rate of hybridization would also be low. In certain conditions, the blurring effect of hybridization and recovering effect of selection may be in balance and will sustain genetic diversity of the populations.

7. CONCLUSIONS

There are two cultivated rice species, one domesticated in Asia and the other in Africa. Their wild progenitors are an Asian wild rice showing varying degrees of perenniability, and an African annual wild rice, respectively. Domestication proceeded within each gene pool independently. The variation patterns developed are more complex in the Asian gene pool than in the African gene pool, and indicate the importance of life history and mating system of ancestral taxa in determining genetic architecture and evolutionary outcome under domestication. We have discussed different aspects of genetic diversity in the cultivated and wild species and the genetic bases of sterility barriers developed between and within those taxa.

The Asian gene pool has three major axes of differentiation, namely, differentiation of annual types from perennials in wild populations, evolution of domesticates from wild plants, and differentiation of domesticates into the Indica and Japonica types. Variation studies at phenotypic and molecular levels indicate that evolutionary changes along the former two axes are due to adaptive processes. Intermediate perennial-annual wild populations were considered most likely to be the progenitor of incipient domesticates. The mechanisms of Indica-Japonica differentiation are more problematic. These two varietal groups, differing at a subspecies level, could have been produced by fortuitous evolutionary events combined with adaptive genic changes. Speciation appears to have been the results of historical and selectional factors combined with inner genomic mechanisms. For full understanding of the origin of the two types, we will need much more information ranging from archaeological to molecular data.

Genetic differences between rice taxa have been moulded through gradual processes that have accumulated numerous genic changes. We have become inclined to consider that the relevant genes are organized rather than randomly dispersed in the genome. Differentiation of taxa is always accompanied by evolution of multilocus systems. Experimental results have led us to consider that coadaptive interaction between genes is an important factor in generating such non-random multilocus systems. Interlocked adaptive sets of genes might have been selected through environmental agents and internal mechanisms. Increased selfing accompanying evolution could also help non-random association to persist. Furthermore, gene complexes or gene blocks in which functionally related genes are linked attract our interest. These might have played a positive role in forming coadapted sets of genes as well as in developing effective reproductive barriers.

In addition to the multilocus variation, allelic diversity occurring at each locus is also of interest. Domestication has either increased or

decreased allelic diversity in cultivars through adaptive and neutral processes. Recently developed molecular techniques enable us to reveal a variety of mutational changes underlying allelic diversity. Molecular genetic studies in rice are still in their infancy, and the evolutionary significance of the phenomena revealed is little known. However, accumulation of knowledge on genome structure and gene regulation will give us a clearer perspective on rice evolution.

REFERENCES

- Amemiya, A. and Akemine, H. (1963). Biochemical genetic studies on the root growth inhibiting complementary lethal genes of rice plants. *Bull. Natl. Inst. Agric. Sci. Jpn.* **D10**, 139–226.
- Appels, R. and Honeycutt, R. L. (1986). rDNA: evolution over a billion years. In *DNA systematics*, (ed. S. K. Dutta), pp. 81–135. CRC Press, Boca Raton, Florida.
- Barbier, P. (1989). Genetic variation and ecotypic differentiation in the wild rice species *Oryza rufipogon*. II. Influence of the mating system and life-history traits on the genetic structure of populations. *Jpn. J. Genet.* **64**, 273–85.
- (1990). The annual–perennial differentiation in wild rice: Quantitative and molecular genetic studies. PhD thesis, Nagoya University.
- , Morishima, H. and Ishihama, A. (1991). Phylogenetic relationships of annual and perennial wild rice: Probing by direct DNA sequence. *Theor. Appl. Genet.* **81**, 693–702.
- Cameron, D. R. and Moav, R. (1957). Inheritance in *Nicotiana tabacum*. XXVII. Pollen killer, an alien genetic locus inducing abortion of microspores not carrying it. *Genetics* **42**, 326–35.
- Chang, T. T. (1976). The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. *Euphytica* **25**, 435–41.
- Charnov, E. L. (1982). *The theory of sex allocation*. Princeton University Press, Princeton.
- Chatterjee, D. (1951). Note on the origin and distribution of wild and cultivated rices, *Ind. J. Genet. Plant Breed.* **11**, 18–22.
- Chou, S. L. (1948). China is the place of origin of rice. *J. Rice Soc. China* **7(5)**, 53–4 (in Chinese).
- Chu, Y. E. and Oka, H. I. (1970a). The genetic basis of crossing barriers between *Oryza perennis* subsp. *barthii* and its related taxa. *Evolution* **24**, 135–44.
- and Oka, H. I. (1970b). Introgression across isolating barriers in wild and cultivated *Oryza* species. *Evolution* **24**, 344–55.
- and Oka, H. I. (1972). The distribution and effects of genes causing F_1 weakness in *Oryza breviligulata* and *O. glaberrima*. *Genetics* **79**, 163–73.
- , Morishima, H. and Oka, H. I. (1969a). Reproductive barriers distributed in cultivated rice species and their wild relatives. *Jpn. J. Genet.* **44**, 225–9.
- , Morishima, H. and Oka, H. I. (1969b). Partial self-incompatibility in *Oryza perennis* subsp. *barthii*. *Jpn. J. Genet.* **44**, 225–9.

- Cordesse, F., Second, G. and Delseny, M. (1990). Ribosomal gene spacer length variability in cultivated and wild rice species. *Theor. Appl. Genet.* **79**, 81–8.
- Dally, A. M. and Second, G. (1990). Chloroplast DNA diversity in wild and cultivated species of rice (Genus *Oryza*, section *Oryza*). Cladistic-mutation and genetic-distance analysis. *Theor. Appl. Genet.* **80**, 209–22.
- Eiguchi, M. and Sano, Y. (1990). A gene complex responsible for seed shattering and panicle spreading found in common wild rices. *Rice Genet. Newslett.* **7**, 105–7.
- Endo, T. and Tsunewaki, K. (1975). Sterility of common wheat with *Aegilops triuncialis* cytoplasm. *J. Hered.* **66**, 13–8.
- Finch, T. R., Miller, T. E. and Bennett, M. D. (1984). 'Cuckoo' *Aegilops* addition chromosome in wheat ensures its transmission by causing chromosome breaks in meiospores lacking it. *Chromosoma* **90**, 84–8.
- Gadgil, M. D. and Solbrig, O. T. (1972). The concept of r- and K-selection: Evidence from wild flowers and some theoretical considerations. *Am. Nat.* **106**, 14–31.
- Ghesquiere, A. (1986). Evolution of *Oryza longistaminata*. In *Rice genetics*, pp. 15–25. IRRI, Manila.
- Glaszmann, J. C. (1987). Isozymes and classification of Asian rice varieties. *Theor. Appl. Genet.* **74**, 21–30.
- Grant, V. (1981). *Plant speciation*. Columbia University Press, New York.
- Hakoda, H., Inooye, J. and Morishima, H. (1990). Isozyme diversity found among Asian deepwater rices. *Rice Genet. Newslett.* **7**, 91–3.
- Harlan, J. R. (1965). The possible role of weed races in the evolution of cultivated plants. *Euphytica* **14**, 173–6.
- (1966). Plant introduction and biosystematics. In *Plant breeding* (ed. K. J. Frey), pp. 55–83. Iowa State University Press, Ames.
- (1971). Agricultural origins: Centers and noncenters. *Science* **174**, 468–74.
- and De Wet, J. M. J. (1971). Toward a rational classification of cultivated plants. *Taxon* **20**, 509–17.
- , De Wet, J. M. J. and Price, E. G. (1973). Comparative evolution of cereals. *Evolution* **27**, 311–25.
- Hayashi, C. (1956). Theory and examples of quantification. II. *Proc. Inst. Statis. Math., Japan*, **42**(2), 19–30 (in Japanese).
- Hedrick, P. W., Jain, S. K. and Holden, L. (1978). Multilocus systems in evolution. *Evol. Biol.* **11**, 101–84.
- Hill, H. G. and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**, 226–31.
- Hinata, K. and Oka, H. I. (1962). A survey of hybrid sterility relationships in the Asian forms of *Oryza perennis* and *O. sativa*. *Jpn. J. Genet.* **37**, 314–28.
- Ho, K. C. and Li, H. W. (1965). Cytogenetical studies of *Oryza sativa* L. and its related species, 9. Studies on meiosis of the first backcross generation of (*O. sativa* L. × *O. officinalis* Wall.) × *O. sativa* L. *Bot. Bull. Acad. Sinica* **6**(1), 39–47.
- Ichikawa, H., Hirai, A. and Katayama, T. (1986). Genetic analyses of *Oryza* species by molecular markers for chloroplast genomes. *Theor. Appl. Genet.* **72**, 353–8.

- Ikehashi, H. and Araki, H. (1986). Genetics of F_1 sterility in remote crosses of rice. In *Rice genetics*, pp. 119–30. IRRI, Manila.
- Inukai, T., Maekawa, M. and Shinbashi, N. (1990). Seed sterility in F_1 s of the crosses between Silewah, Indonesian native variety and rice varieties in Hokkaido. *Jpn. J. Breed.* **40**, (suppl. 2), 394–5 (in Japanese).
- Ishii, T., Terachi, T. and Tsunewaki, K. (1988). Restriction endonuclease analysis of chloroplast DNA from A-genome diploid species of rice. *Jpn. J. Genet.* **63**, 523–36.
- Jena, K. K. and Khush, G. S. (1990). Introgression of genes from *Oryza officinalis* Wall. ex Watt to cultivated rice, *O. sativa* L. *Theor. Appl. Genet.* **80**, 737–45.
- Kadowaki, K., Ishige, T., Suzuki, S., Harada, K. and Shinjyo, C. (1986). Differences in the characteristics of mitochondrial DNA between normal and male sterile cytoplasms of Japonica rice. *Jpn. J. Breed.* **36**, 333–9.
- Kadowaki, K., Yazaki, K., Osumi, T., Harada, K., Katsuta, M. and Nakagahra, M. (1988). Distribution of mitochondrial plasmid-like DNA in cultivated rice (*Oryza sativa* L.) and its relationship with varietal groups. *Theor. Appl. Genet.* **76**, 809–14.
- Katayama, T. and Onizuka, W. (1979). Intersectional F_1 plants from *Oryza sativa* \times *O. ridleyi* and *O. sativa* \times *O. meyeriana*. *Jpn. J. Genet.* **54**, 43–6.
- Kato, S., Kosaka, H. and Hara, S. (1928). On the affinity of rice varieties as shown by fertility of hybrid plants. *Bull. Sci., Fac. Agric. Kyushu Univ.* **3**, 132–47.
- Kato, S., Kosaka, H., Maruyama, Y. and Takiguchi, Y. (1930). On the affinity of cultivated varieties of rice plants, *Oryza sativa* L. *J. Dept. Agric. Kyushu Univ.* **2**, 241–76.
- Kitamura, E. (1962). Genetic studies on sterility observed in hybrids between distantly related varieties of rice, *Oryza sativa* L. *Bull. Chugoku Agric. Exp. Station A8*, 141–205 (in Japanese).
- Kochko, A. de. (1987). Isozyme variability of traditional rice in Africa. *Theor. Appl. Genet.* **73**, 675–82.
- Lande, R. and Schemske, D. W. (1985). The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution* **39**, 24–40.
- Li, H. W. (1964). Studies on genetic and cytogenetic evidence for species relationships in the Republic of China. In *Rice genetics and cytogenetics* (ed. IRRI), pp. 118–43. Elsevier, Amsterdam.
- , Yang, K. S. and Ho, K. C. (1964). Cytogenetical studies on *Oryza sativa* L. and its related species, 7. Non-synchronization of mitosis and cytogenesis in relation to the formation of diploid gametes in the hybrids of *O. sativa* L. and *O. officinalis* Wall. *Bot. Bull. Acad. Sinica* **5**(2), 142–53.
- Linhart, V. B. (1974). Intra-population differentiation in annual plants. I. *Veronica peregrina* L. raised under non-competitive conditions. *Evolution* **28**, 232–43.
- Loegering, W. G. and Sears, E. R. (1963). Distorted inheritance of stemrust resistance of timstein wheat caused by a pollen killing gene. *Can. J. Genet. Cytol.* **5**, 67–72.
- McCauley, D. E. (1991). Genetic consequences of local population extinction and recolonization. *Trends Ecol. Evol.* **6**(1), 5–8.
- Moore, D. M. and Lewis, H. (1965). The evolution of self-pollination in *Clarkia xantiana*. *Evolution* **19**, 104–14.

- Morinaga, T. and Kuriyama, H. (1957). Cytological studies on *Oryza sativa* L. and *O. glaberrima* Steud. *Jpn. J. Breed.* **7**, 57–65.
- (1958). Intermediate type of rice in the subcontinent of India and Java. *Jpn. J. Breed.* **7**, 253–9.
- Morishima, H. (1969). Phenetic similarity and phylogenetic relationships among strains of *Oryza perennis*, estimated by methods of numerical taxonomy. *Evolution* **23**, 429–43.
- (1986). Wild progenitors of cultivated rice and their population dynamics. In *Rice genetics* (ed. IRRI), pp. 3–14. IRRI, Manila.
- (1989). Intra-populational genetic diversity in landrace of rice. In *Breeding research: The key to the survival of the earth*, pp. 159–62. Org. Commit. 6th Intl. Congr. SABRAO, Tokyo.
- (1991). Association between *Pox-1* variation and seed productivity potential in wild rice. In *Rice genetics II*. IRRI, Manila, in press.
- and Barbier, P. (1990). Mating system and genetic structure of natural populations in wild rice *Oryza rufipogon*. *Plant Species Biol.* **5**, 31–9.
- and Gadrinab, L. U. (1987). Are the Asian common wild-rices differentiated into the Indica and Japonica types? In *Crop exploration and utilization of genetic resources* (ed. S. C. Hsieh), pp. 11–20. Taichung Distr. Agric. Improv. Station, Changhua, Taiwan.
- and Oka, H. I. (1970). A survey of genetic variations in the populations of wild *Oryza* species and their cultivated relatives. *Jpn. J. Genet.* **45**, 371–85.
- and Oka, H. I. (1981). Phylogenetic differentiation of cultivated rice, 22. Numerical evaluation of the Indica–Japonica differentiation. *Jpn. J. Breed.* **31**, 402–13.
- , Hinata, K. and Oka, H. I. (1962). Comparison between two cultivated rice species, *Oryza sativa* L. and *O. glaberrima* Steud. *Jpn. J. Breed.* **12**, 153–65.
- , Sano, Y. and Oka, H. I. (1984a). Differentiation of perennial and annual types due to habitat conditions in the wild rice *Oryza perennis*. *Plant Syst. Evol.* **144**, 119–35.
- , Shimamoto, Y., Sano, Y. and Sato, Y. I. (1984b). *Observations on wild and cultivated rices in Thailand for ecological-genetic study: Report of study-tour in 1983*. Natl. Inst. Genet. Jpn., Mishima, Japan.
- Nakagahra, M. (1978). The differentiation, classification and center of genetic diversity of cultivated rice (*Oryza sativa* L.) by isozyme analysis. *Trop. Agr. Res. Ser.* **11**, 77–82.
- , Nagamine, T. and Okuno, J. (1986). Spontaneous occurrences of low amylose genes and geographical distribution of amylose contents in Asian rice. *Rice Genet. Newslett.* **3**, 46–8.
- Nawa, S., Sano, Y., Yamada, M. and Fujii, T. (1987). Cloning of the plasmids in cytoplasmic male sterile rice and changes of organization of mitochondrial and nuclear DNA in cytoplasmic reversion. *Jpn. J. Genet.* **62**, 301–14.
- Nayar, N. M. (1973). Origin and cytogenetics of rice. *Adv. Genet.* **17**, 153–292.
- Oka, H. I. (1953). Phylogenetic differentiation of cultivated rice. 8. Gene analysis of intervarietal hybrid sterility and certation due to certain combinations of gametic-development genes in rice. *Jpn. J. Breed.* **3**, 23–30.
- (1955). Phylogenetic differentiation of cultivated rice. 11. Changes of gene frequency in hybrid populations of rice. *Jpn. J. Breed.* **5**, 59–64.

- (1956). Phylogenetic differentiation in cultivated rice. 12. Polygenic nature of 'gametic-development-genes' controlling intervarietal hybrid sterility in rice. *Jpn. J. Breed.* **6**, 51–5.
- (1957a). Phylogenetic differentiation of cultivated rice, 15. Complementary lethal genes in rice. *Jpn. J. Genet.* **32**, 83–7.
- (1957b). Genic analysis for the sterility of hybrids between distantly related varieties in rice. *J. Genet.* **55**, 397–409.
- (1958). Intervarietal variation and classification of cultivated rice. *Ind. J. Genet. Plant Breed.* **18**, 79–89.
- (1964a). Pattern of interspecific relationships and evolutionary dynamics in *Oryza*. In *Rice genetics and cytogenetics* (ed. IRRI), pp. 71–90. Elsevier, Amsterdam.
- (1964b). Considerations on the genetic basis of intervarietal sterility in *Oryza sativa*. In *Rice genetics and cytogenetics* (ed. IRRI), pp. 158–74. Elsevier, Amsterdam.
- (1968). Preferential pairing of chromosomes in a tetraploid between *Oryza glaberrima* and *O. sativa*. *Can. J. Genet. Cytol.* **10**, 527–35.
- (1974a). Analysis of genes controlling F₁ sterility in rice by the use of isogenic lines. *Genetics* **77**, 521–34.
- (1974b). Experimental studies on the origin of cultivated rice. *Genetics* **78**, 475–86.
- (1976). Mortality and adaptive mechanisms of *Oryza perennis* strains. *Evolution* **30**, 380–92.
- (1984). Secondary succession of weed communities in lowland habitats of Taiwan in relation to the introduction of wild-rice (*Oryza perennis*) populations. *Vegetatio* **56**, 177–87.
- (1988). *Origin of cultivated rice*. Japan Sci. Soc. Press/Elsevier, Tokyo/Amsterdam.
- and Chang, W. T. (1959). The impact of cultivation on populations of wild rice, *Oryza sativa* f. *spontanea*. *Phyton* **13**, 105–17.
- and Chang, W. T. (1961). Hybrid swarms between wild and cultivated rice species, *Oryza perennis* and *O. sativa*. *Evolution* **15**, 418–30.
- and Chang, W. T. (1962). Rice varieties intermediate between wild and cultivated forms and the origin of the Japonica type. *Bot. Bull. Acad. Sinica* **3**, 109–31.
- and Morishima, H. (1967). Variation in the breeding systems of a wild rice, *Oryza perennis*. *Evolution* **21**, 249–58.
- and Morishima, H. (1971). The dynamics of plant domestication: Cultivation experiments with *Oryza perennis* and its hybrid with *O. sativa*. *Evolution* **25**, 356–64.
- and Morishima, H. (1982). Phylogenetic differentiation of cultivated rice, 23. Potentiality of wild progenitors to evolve the Indica and Japonica types of rice cultivars. *Euphytica* **31**, 41–50.
- and Sano, Y. (1981). Differentiation of *Oryza perennis* populations in adaptive strategy. *Proc. XIV Intl. Congr. Genet.* (ed. Y. P. Altukhov), pp. 68–85. Mir Publications, Moscow.
- Pental, D. and Barnes, S. R. (1985). Interrelationship of cultivated rices *Oryza*

- sativa* and *O. glaberrima* with wild *O. perennis* complex. *Theor. Appl. Genet.* **70**, 185–91.
- Rick, C. M. (1966). Abortion of male and female gametes in tomato determined by allelic interaction. *Genetics* **53**, 85–96.
- (1971). The tomato *Ge* locus: linkage relations and geographic distribution of alleles. *Genetics* **67**, 75–85.
- Roschevitz, R. J. (1931). A contribution to the knowledge of rice. *Bull. Appl. Bot. Genet. Plant Breed.* **17**, 3–133 (Russian/English).
- Sakai, K. I. and Iyama, S. (1957). Statistical studies on the breeding behavior of wild populations of *Setaria pumila*. *Ann. Rep. Natl. Inst. Genet., Jpn.* **7**, 53–5.
- and Narise, T. (1960). Further note on natural crossing in wild rice. *Ann. Rep. Natl. Inst. Genet., Jpn.* **10**, 65.
- Sampath, S. and Rao, M. B. V. N. (1951). Interrelationships between species in genus *Oryza*. *Ind. J. Genet. Plant Breed.* **11**, 14–7.
- and Govindaswami, S. (1958). Wild rices of Orissa and their relationship to the cultivated varieties. *Rice News Teller* **6**(3), 17–20.
- Sano, R., Konishi, T. and Morishima, H. (1985). An altitudinal cline of isozyme variation found in land-race populations of Nepalese rice. *Rice Genet. Newslett.* **2**, 51–2.
- Sano, Y. (1983). A new gene controlling sterility in F₁ hybrids of two cultivated rice species. *J. Hered.* **74**, 435–9.
- (1984). Differential regulation of waxy gene expression in rice endosperm. *Theor. Appl. Genet.* **68**, 467–73.
- (1985). Interspecific cytoplasm substitutions of an Indica strain of *Oryza sativa* L. and *O. glaberrima* Steud. *Euphytica* **34**, 587–92.
- (1986). Sterility barriers between *Oryza sativa* and *O. glaberrima*. In *Rice genetics* (ed. IRRI), pp. 109–18. IRRI, Manila.
- (1989a). The direction of pollen flow between two co-occurring rice species, *Oryza sativa* and *O. glaberrima*. *Heredity* **63**, 353–7.
- (1989b). Gamete eliminator detected in the wild progenitor of *Oryza sativa*. *Rice Genet. Newslett.* **6**, 93–4.
- (1990). Genic nature of gamete eliminator in rice. *Genetics* **125**, 183–91.
- and Morishima, H. (1982). Variation in resource allocation and adaptive strategy of a wild rice, *Oryza perennis*. *Moench. Bot. Gaz.* **143**, 518–23.
- and Sano, R. (1990). Variation of the intergenic spacer region of ribosomal DNA in cultivated and wild rice species. *Genome* **33**, 209–18.
- , Chu, Y. E. and Oka, H. I. (1979). Genetic studies of speciation in cultivated rice. I. Genic analysis for the F₁ sterility between *O. sativa* L. and *O. glaberrima* Steud. *Jpn. J. Genet.* **54**, 121–32.
- , Chu, Y. E. and Oka, H. I. (1980a). Genetic studies of speciation in cultivated rice, 2. Character variations in backcross derivatives between *Oryza sativa* and *O. glaberrima*: M-V linkage and key characters. *Jpn. J. Genet.* **55**, 19–39.
- , Morishima, H. and Oka, H. I. (1980b). Intermediate perennial-annual populations of *Oryza perennis* found in Thailand and their evolutionary significance. *Bot. Mag. Tokyo* **93**, 291–305.
- , Katsumata, M. and Amano, E. (1985). Correlations between the amount of amylose and *Wx* protein in rice endosperm. *SABRAO J.* **17**, 121–33.

- , Katsumata, M. and Okuno, K. (1986). Genetic studies of speciation in cultivated rice. 1. Genic analysis for the F_1 sterility between *Oryza sativa* L. and *O. glaberrima* Steud. *Jpn. J. Genet.* **54**, 121–32.
- , Yi, H. X., Shao, Q. Q. and Iyama, S. (1989). Ribosomal DNA spacer-length variations in a wild rice population from Dongxiang, China. In *Breeding research: The key to the survival of the earth*, pp. 493–6. Org. Commit. 6th Intl. Congr. SABRAO, Tokyo.
- , Hirano, H. and Nishimura, M. (1991). Evolutionary significances of differential regulation at the *wx* locus of rice. In *Rice genetics II*. IRRI, Manila, in press.
- Sato, Y. I. (1985). Genetic control of apiculus hair length. *Rice Genet. Newslett.* **2**, 72–4.
- (1986). The Indica–Japonica differentiation of rice cultivars in Thailand and its neighboring countries. In *New frontiers in breeding research* (Proc. 5th Intern. Congr. SABRAO, ed. B. Napompeth and S. Subhadraavandhu), pp. 185–93. Kasetsart University, Bangkok.
- (1987). Character association patterns among Japonica rice varieties of Asian origin. In *Crop exploration and utilization of genetic resources* (ed. S. C. Hsieh), pp. 21–9. Taichung Distr. Agric. Improv. Station, Changhua, Taiwan.
- (1988). ‘Alien pollen primacy’ detected in rice. *Rice Genet. Newslett.* **5**, 80–1.
- (1990). Increasing trend of *japonica*-derived genes in hybrid populations grown under upland conditions. *Rice Genet. Newslett.* **7**, 94–5.
- (1991). Variation in spikelet shape of the *indica* and *japonica* rice cultivars of Asian origin. *Jpn. J. Breed.* **41**, 121–34.
- and Hayashi, K. (1983). Distribution of the complementary genes causing F_1 weakness in the common rice and its wild relatives, I. *L-2-a* gene in Asian native cultivars. *Jpn. J. Genet.* **58**, 411–8.
- , Ishikawa, R. and Morishima, H. (1987). Linkage analysis of gene *Aph* for apiculus hair length. *Rice Genet. Newslett.* **4**, 74–5.
- , Fujiwara, H., Udatstu, T. and Morishima, H. (1989). Plant opal analysis as a method of distinguishing between subspecies *indica* and *japonica* of *O. sativa*. *Rice Genet. Newslett.* **6**, 67–9.
- , Ishikawa, R. and Morishima, H. (1990a). Non-random association of genes and characters found in *indica* \times *japonica* hybrids of rice. *Heredity* **65**, 75–9.
- , Sato, T. and Morishima, H. (1990b). Changing segregation ratios for *gl-1* due to unequal changes of fertilization of female gametes. *Rice Genet. Newslett.* **7**, 97–8.
- Scoles, G. T. and Kibirge-Sebunya, I. N. (1983). Preferential abortion of gametes in wheat induced by an *Agropyron* chromosome. *Can. J. Genet. Cytol.* **25**, 1–6.
- Second, G. (1982). Origin of the genic diversity of cultivated rice (*Oryza spp.*): Study of the polymorphism scored at 40 isozyme loci. *Jpn. J. Genet.* **57**, 25–57.
- (1985). Evolutionary relationships in the *Sativa* group of *Oryza* based on isozyme data. *Genet. Sel. Evol.* **17**, 89–114.
- Shastry, S. V. S. and Misra, R. N. (1961). Pachytene analysis in *Oryza*. II. Sterility in *japonica*-*indica* hybrids. *Chromosome* **12**, 248–71.
- Shinjyo, C. (1984). Cytoplasmic male sterility and fertility restoration in rice

- having genome A. In *Biology of rice* (ed. S. Tsunoda and N. Takahashi), pp. 321–38. Japan Sci. Soc. Press, Tokyo/Elsevier, Amsterdam.
- Stebbins, G. L. (1958). The inviability, weakness and sterility of interspecific hybrids. *Adv. Genet.* **9**, 147–215.
- Stephens, S. G. (1946). The genetics of 'corky.' I. The New World alleles and their possible role as an interspecific isolating mechanism. *J. Genet.* **47**, 150–61.
- Takahashi, M. (1982). Gene analysis and its related problems – Genetical studies on rice plant, LXXX-. *J. Fac. Agr. Hokkaido Univ.* **61**(1), 91–142.
- Tanaka, T., Kawase, M., Saito, A., Yano, M., Kishimoto, N., Saito, K., Yoshimura, A., Takeda, H., Nagamine, T. and Nakagahra, M. (1989). Intraspecific variation on restriction fragment length polymorphism of nuclear DNA in rice, *Oryza sativa* L. In *Breeding research: The key to the survival of the earth*. pp. 599–602. Org. Commit. 6th Intl. Congr. SABRAO, Tokyo.
- Tang, L. H. and Morishima, H. (1988). Characteristics of weed rice strains. *Rice Genet. Newslett.* **5**, 70–2.
- Tanisaka, T., Tomita, M. and Yamagata, H. (1990). Gene analysis for the semidwarfism of two mutant strains. Hokuriku 100 and Kanto 79, induced from a rice variety Koshihikari. Studies on the utility of artificial mutations in plant breeding XVIII. *Jpn. J. Breed.* **40**, 103–17.
- Terao, H. and Mizushima, U. (1939). Some considerations on the classification of *O. sativa* L. *Jpn. J. Bot.* **10**, 213–58.
- Vaughan, D. A. (1989). *The genus Oryza* L.: Current status of taxonomy, IRPS **138**. IRRI, Manila.
- Vavilov, N. I. (1951). Phytogeographic basis of plant breeding. The origin, variation, immunity and breeding of cultivated plants. *Chron. Bot. (Waltham)* **13**, 1–366.
- Virmani, S. S. and Shinjyo, C. (1988). Current status of analysis and symbols for male-sterile cytoplasms and fertility-restoring genes. *Rice Genet. Newslett.* **5**, 9–15.
- Wang, X. K., Cheng, K. S., Lu, Y. X., Luo, J., Huang, N. W. and Liu, G. R. (1984). Studies on indigenous rice in Yunnan and their utilization, 3. Glabrous rice in Yunnan. *Acta Univ. Agric. Pekinensis* **10**(4), 333–44 (in Chinese).
- Wang, Z. Y. and Tanksley, S. D. (1989). Restriction fragment length polymorphism in *Oryza sativa* L. *Genome* **32**, 1113–8.
- Whyte, R. O. (1972). The Gramineae, wild and cultivated, of monsoonal and equatorial Asia. *Asian Perspect.* **25**, 127–51.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.
- Yabuno, T. (1977). Genetic studies on the interspecific cytoplasm substitution lines of japonica varieties of *Oryza sativa* L. and *O. glaberrima* Steud. *Euphytica* **26**, 451–63.
- Yamaguchi, H. and Kakiuchi, H. (1983). Electrophoretic analysis of mitochondrial DNA from normal and male sterile cytoplasms in rice. *Jpn. J. Genet.* **58**, 607–11.
- Yasui, H. and Iwata, N. (1991). Production of monosomic alien addition lines of *Oryza sativa* having a single chromosome of *O. punctata*. In *Rice genetics II*. IRRI, Manila, in press.
- You, X. L. (1986). Origin of rice culture around Lake Taihu, its spread and development. *Chinese Agric. History* **1**, 8–19 (in Chinese).

The theoretical population genetics of autopolyploidy

JAMES D. BEVER and FRANÇOIS FELBER

1. INTRODUCTION

Polyplody has played an important role in the evolution of both animals and plants. Polyplody is present in insects, amphibians, reptiles, and fishes (Sexton 1980). In plants, polyplody has been noted in each of the major systematic divisions: algae, fungi, bryophytes, pteridophytes, gymnosperms, and angiosperms (Averett 1980). The frequency of polyplody varies considerably between groups, but is particularly high in pteridophytes and angiosperms. Estimates of polyplloid taxa in angiosperms range from 30 to 35 per cent (Stebbins 1971), to 40 per cent (White 1952), 47 per cent (Grant 1981), and even 50 per cent (Müntzig 1936; Darlington 1937), depending on the method of calculation.

The theoretical population genetics of polyploids, particularly autopolyploidy, differs qualitatively from that of diploids. The diversity of potential chromosomal pairing arrangements (called meiotic figures) allows for unique segregation patterns in autopolyploids. These potential pairing arrangements also alter our expectations for chromosomal and gametic recombination. Autopolyploid segregation patterns alter our expectations of the consequences of different breeding systems. Furthermore, allelic selection in polyploid populations is potentially much more complex than in diploid populations. These unique aspects of polyploid population genetics are rarely included in discussions of microevolutionary process.

While many subjects pertaining to polyploidy have been reviewed, including cytology (Jackson 1984), cytogeography (Favarger 1967, 1984; Stebbins 1985, Stebbins and Dawe 1987), systematics (Löve 1960; Lewis 1967), evolution (de Wet 1971; Jackson 1976; Dobzhansky *et al.* 1977; Levin 1983), and even the quantitative genetics of autopolyploidy (Rowe and Hill 1984a; Wricke and Weber 1986), the theoretical population genetics of polyploidy has been frequently overlooked. The goal of this paper is to review work on the population genetics of polyploids and to emphasize novel features of gene frequency dynamics in such populations.

2. SEGREGATION OF POLYPLOIDS

Polyploids are commonly categorized as auto- or allopolyploids depending on presumed origin (Kihara and Ohno 1926). These terms commonly refer, respectively, to chromosome doubling of the same genome and to increase of chromosome number through hybridization and subsequent chromosome doubling. Segregation of allopolyploids is similar to that of nonhomologous pairs of chromosomes in diploids. Segregation of autopolyploids is qualitatively more complex than that of diploids because more than two homologous chromosomes can pair at meiosis. Because deviation from autopolyploid behavior may not be exclusively caused by genomic differentiation, but may also result from the action of specific genes controlling pairing behavior (Waines 1976, Jackson 1982), classifications based purely on putative origin have been criticized. Jackson and Casey (1982) proposed a classification based on pairing behavior at meiosis, with autopolyploids forming multivalents at meiosis. Because genetic peculiarities of polyploids are largely the consequence of multivalent formation at meiosis, this paper focuses mainly on autopolyploids as defined by Jackson and Casey (1982). We shall refer to autopolyploid when we write polyploid, unless otherwise specified.

Segregation patterns at a given polyploid locus can vary between two extremes – random chromosome segregation (RCeS) or random chromatid segregation (RCdS). Under RCeS, the chromosomes in a gamete may originate from any random combination of homologous chromosomes. In this case, the two sister chromatids of a chromosome never sort into the same gamete. Segregation ratios under RCeS have been given for autotetraploids by Müller (1914), Haldane (1930), Mather (1935, 1936), Geiringer (1944, 1948), and Li (1955). Other ploidy levels have also been considered: triploidy (Mather 1935), hexaploidy (Haldane 1930), octoploidy (Lawrence 1929; Haldane 1930), 10-, 12- and 16-ploidy (Haldane 1930).

Chromatids also may behave independently during meiosis and distribute randomly into gametes (RCdS). Consequently, two sister chromatids may migrate into the same gamete. This phenomenon, called double reduction, could never happen in diploids. Segregation patterns of tetraploids under this scheme were described by Haldane (1930), Mather (1935, 1936), Fisher and Mather (1943), Geiringer (1949), and Li (1955). Segregation of triploids (Mather 1935, 1936), hexaploids (Fisher and Mather 1943; Geiringer 1949), and octoploids (Geiringer 1949) have also been considered.

Double reduction is a phenomenon unique to meiosis in polyploids. At anaphase I, chromatids located on a chromosome may migrate either to the same pole (reductional separation) or to different poles (equational separation). The type of separation depends on the number and the type

of cross-over located between the centromere and the locus under consideration. Segregation of two loci A and B of an autotetraploid in the case of quadrivalent formation results in three possible outcomes or paths (Fig. 1). In path X, no cross-over occurs between the centromere and locus A. First division is reductional and double reduction never happens. In the other two paths one cross-over is assumed to occur between the centromere and locus B which undergoes equational separation. Assuming that the four homologues segregate randomly, either chromosomes 1 and 4 and their respective homologues, 2 and 3, may migrate to the same cell or chromosomes 1 and 2, and 3 and 4 may migrate to the same cell. In the first case, alleles located on sister chromatids reach different cells and double reduction never occurs (path Y). In the second class, however, double reduction may occur in half the cases when chromatids segregate randomly (path Z).

For a given locus, if the proportion of equational separation is e , and if the frequency of equational separation leading to double reduction is

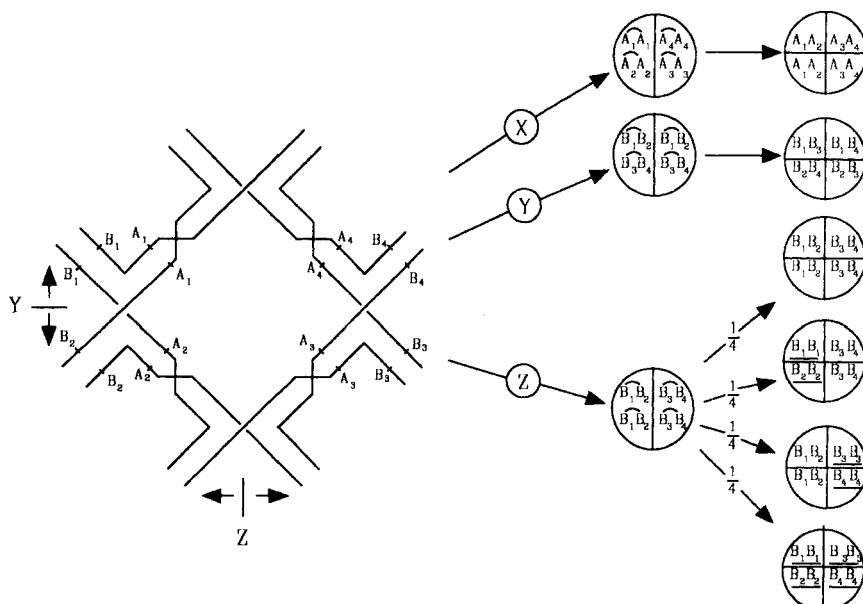


Fig. 1. Segregation of two loci, A and B, of an autotetraploid in the case of quadrivalent formation (from Mather 1936). X describes the segregation of the locus A characterized by the absence of a cross-over between the centromere and the locus. First division in this case is reductional. One cross-over occurs between the centromere and locus B. Separation is always equational (Y and Z) and may lead to double reduction (Z). Gametes having undergone double reduction are underscored.

a , then the product ae describes the proportion of divisions leading to double reduction. This product was called the index of separation, α , by Mather (1936). It is twice the proportion of gametes that have undergone double reduction. It should be noted that Fisher and Mather (1943) also adopted the symbol α for the actual proportion of gametes that have undergone double reduction. Seyffert (1959, 1960) adopted this new terminology, but Demarly (1963) did not. For simplicity, α will refer later in this paper to the original definition of the index of separation (Mather 1936). General formulations with a frequency α of double reduction were given for tetraploids by Mather (1936), Fisher (1949), Seyffert (1960) and Demarly (1963) (Table 1). The equations reduce to RCeS if $\alpha = 0$ and to RCdS if $\alpha = 2/7$.

The segregation of most polyploid loci is likely to be intermediate between RCeS and RCdS. The proportion of double reduction is related to the presence of cross-overs between the centromere and the locus – a gene located in the vicinity of the centromere undergoes mostly chromosome segregation (low values of α) and chromatid segregation increases with distance (higher values of α) (Li 1955).

The prediction of the segregation ratio is mathematically complex. A branch of mathematics called ‘genetic algebra,’ describing cross-generation segregation expectations, has been developed (Gonshor 1965, 1973; Holgate 1966; Heuch 1978; Wörz-Busekros 1978; Fortini and Barakat 1980).

2.1 Meiotic configurations

Prediction of segregation patterns from meiotic configurations is equally complex. Double reduction can occur only if multivalents are formed.

Table 1

Segregation patterns of the different possible genotypes of zygotes for tetraploids (from Seyffert 1960, modified). α is the index of separation, which is twice the proportion of double reduction

Parental genotype		12 (13) (14)	22 (33) (44)	23 (24) (34)
1111	1	—	—	—
1112	$(4+\alpha)/8$	$(2-\alpha)/4$	$\alpha/8$	—
1122	$(1+\alpha)/6$	$2(2-\alpha)/6$	$(1+\alpha)/6$	—
1123	$(1+\alpha)/6$	$(2-\alpha)/6$	$\alpha/8$	$(2-\alpha)/12$
1234	$\alpha/8$	$(2-\alpha)/12$	$\alpha/8$	$(2-\alpha)/12$

Consequently, observations of a low proportion of multivalents indicate a low proportion of double reduction. A large literature has accumulated that develops theoretical expectations of polyploid meiotic configurations in order to assess evidence for autopolyploidy or to examine genomic relationships. These expectations are based on assumptions of chromosome association, chiasmata formation, and affinities between genomes, and include models for autopolyploids, for allopolyploids, and models that use optimization techniques that do not make a priori assumptions about polyploid type.

Early autotetraploid models considered that two-thirds of quadrivalent formation and one-third of bivalent formation are expected if each of the four homologues pair in two points and if pairing occurs randomly. Two bivalents or one quadrivalent are formed for each chromosome set. The same number of bivalents and of quadrivalents in a cell are present under these conditions (Hughes-Schrader 1943; John and Henderson 1962; Sved 1966; Jackson and Casey 1980). Meiosis was considered in more recent models in two steps: chromosome association and chiasmata formation. Meiotic figures may be predicted depending on assumptions about the distribution of chiasmata among chromosome arms. The main models for autopolyploids are those of Jackson and Casey (1982), Jackson and Hauber (1982) and Sybenga (1975, 1988).

Models for allopolyploids were conceived mainly for genome analysis. The probability of chiasma formation was considered either as equal for both arms (Driscoll *et al.*, 1979, 1980; Alonso and Kimber 1981; Kimber and Alonso 1981; Espinasse and Kimber 1981; McGuire and Dvorak 1981, 1982; Sybenga 1988) or the probability of pairing of the long arm was differentiated from that of the short arm (Kimber and Hulse 1978). Moreover, Kimber *et al.* (1981) proposed several indexes of similarity for genome analysis of wheat using telocentric chromosomes.

Contrasting with the other models, Crane and Sleper (1989a, b), using an optimization technique, developed a theoretical description of meiotic configuration of triploids and tetraploids, without any a priori assumption of genomic affinity, based on size of the chromosome arms and probability of chiasma formation among pairwise genomes.

Only two papers have described relationships between meiotic configuration and gene segregation. The model of Sarvela (1958) applied to the progeny of allopolyploids backcrossed to a parent. A linear relationship was expected between the number of multivalents and the segregation ratio of recessives. Indeed, a better fit to empirical data was found with a quadratic equation by Phillips (1964). Gallais (1977) decomposed the index of separation, α , into (1) the probability of having a first equational division, (2) the probability of quadrivalent formation, (3) the probability of non-disjunction, and (4) the probability of having two sister chromatids

in the same gamete in a double reduction, which is equal to $\frac{1}{2}$ if no selection occurs.

Meiotic configuration data alone do not allow the prediction of gene segregation because a single meiotic figure may be generated by a variable number of chiasmata. Moreover, if one is interested in the genes underlying a particular trait, then their location on the chromosomes, their linkage relationships, and the probability of crossing over must be known in order to predict a correct segregation. This information is very difficult to gather empirically and this may explain the lack of consideration of meiotic configurations in most theoretical predictions of segregation of polyploids.

3. RANDOM MATING EQUILIBRIUM

Haldane (1927, 1930) derived the frequencies of genotypes for a polyploid population at random mating equilibrium (RME). The RME frequencies depend on the type of allelic segregation (Haldane 1930; Geiringer 1949). For random chromosomal segregation (RCeS) of autotetraploids, these values are p^4 , $4p^3q$, $6p^2q^2$, $4pq^3$, and q^4 for the respective frequencies of the genotypes, A^4 , A^3a , A^2a^2 , Aa^3 , and a^4 , where p and q are frequencies of alleles A and a, respectively. These frequencies are derived from the binomial expansion of $(p+q)^{2m}$, where m is the haploid ploidy level. The RME frequencies of genotypes in populations with multiple alleles can be similarly derived through the expansion of $(p+q+r+\text{etc.})^{2m}$. However, the RME frequencies for random chromatid segregation in polyploids are not as easily derived. For random chromatid segregation (RCdS) in tetraploids the RME frequencies are x^2 , $2xy$, $2(xz+2y^2)$, $2yz$, and z^2 , where $x=(4/5)p^2+(1/5)p$, $y=(4/5)pq$, and $z=(4/5)q^2+(1/5)q$ (Geiringer 1949; Li 1955). Geiringer (1949) and Demarly (1963) developed explicit formulae for RME under segregation patterns intermediate between RCeS and RCdS.

Parsons (1959a) and Bennett (1968) illustrate that RCdS is effectively a system of inbreeding such that at RME the population has an inbreeding coefficient of 0.067. Hence, RME is characterized by a higher proportion of homozygotes under RCdS than under RCeS, owing to the increase of homozygote gamete production when double reduction occurs (Demarly 1963).

In contrast to diploids, polyploids do not reach RME in one generation (Haldane 1927, 1930; Geiringer 1948, 1949; Seyffert 1960; Demarly 1963). RME is approached asymptotically. For random chromosomal segregation in tetraploids, Haldane (1930) and Li (1955) describe a gametic disequilibrium measure:

$$d = y^2 - xz = y - p \times q,$$

where x , $2y$, and z equal the frequency of the gametes AA, Aa, and aa. Similar measures are given by Seyffert (1959) and Demarly (1963) in the general case of a mix of RCeS and RCdS. The quantity d diminishes by two-thirds with each generation of random mating under RCeS and five-sevenths with each generation of random mating under RCdS. Thus, random chromatid segregation increases the advance to equilibrium.

4. INBREEDING

The expected result of complete sib mating in polyploids is the same as that of diploids – homozygous lines in proportion to the original allele frequencies. However, inbreeding in polyploids differs from that of diploids in the rate of approach to equilibrium and in the equilibria reached under intermediate levels of inbreeding. Inbreeding in polyploids differs qualitatively from inbreeding in diploids in that it is not removed by a single generation of random mating. Quantitative comparisons of the effects of inbreeding amongst ploidy levels is difficult because the measures of heterozygosity in polyploids is more complex than in diploids. Diploids have two types of zygotes: homozygotes and heterozygotes. Tetraploids, however, have five types of zygotes (listed in Table 1) including three intermediate levels of homozygosity.

Selfed diploids and selfed tetraploids differ in their rate of approach to homozygosity. Diploids attain a given proportion of full homozygotes 2.9-fold faster than chromatid-segregating tetraploids and 3.8-fold faster than chromosomal-segregating tetraploids (Parsons 1959a). Both Haldane (1930) and Parsons (1959a) emphasize the difficulty in obtaining homozygous lines from polyploid populations. However, inbreeding in polyploids does not just create more homozygotes – it also creates more partial heterozygotes. This suggests that a measure of population homozygosity more general than the proportion of homozygotes is needed for polyploids and below we discuss several approaches that have been suggested.

Two methods have been used to analyze inbreeding: path analysis, developed by Wright; and the generation matrix method, first used by Bartlett and Haldane (1934). Wright (1938) used path analysis to develop expressions for the average correlation of random pairs of genes within zygotes, or the inbreeding coefficient, in polyploid populations under RCeS. Malecot (1948) identified F as the probability that two genes taken at random from an individual are identical by descent. In a tetraploid population:

$$F = f(iii) + 1/2 \times f(iiij) + 1/3 \times f(ijj) + 1/6 \times f(ijk),$$

where $f(iii)$ is the frequency of zygotes with four alleles identical by

descent, etc. The inbreeding coefficient, F , is a function of both the correlation of genes from the same gamete and that between different gametes. Using the inbreeding coefficient, Wright estimated the change in the proportion of heterogeneous pairs of alleles, H ($H=1-F$). In diploids, H is the proportion of heterozygotes. For tetraploids, the heterozygotes $ijjj$ and $iijj$ have H values of $2/3$ and $1/2$, respectively. Wright, using path analysis, demonstrated this measure of heterozygosity declines faster with selfing in diploids than in polyploids. Kempthorne (1957) extended the analysis to allow for the occurrence of double reduction. Table 2 compares the rates of decline in H for different mating systems. Note that the relative difference between diploids and polyploids is less under milder forms of inbreeding.

Bartlett and Haldane (1934) developed the generation matrix method to analyze sib-mating in tetraploids. Fisher (1949) extended this analysis by including other systems of inbreeding in tetraploids and hexaploids. The generation matrix method derives an expression that describes the rate of change in the principal components of the inbreeding matrix. In all cases, the first principal component of the inbreeding matrix was an estimate of Wright's H , verifying that this measure is indeed an important one. As expected, the values describing the rate of change of the first principal components are identical to those derived from path analysis (Table 2). However, inbreeding in polyploids yields more than one principal component. In the case of selfing at a digenic tetraploid locus,

Table 2
Rate of decline of heterozygosity for different mating systems

Ploidy	Segregation	Selfing		Sib-mating		Parent-offspring mating	
		Abs. ^a	Rel. ^b	Abs.	Rel.	Abs.	Rel.
Diploid		0.5	1	0.809	1	0.809	1
Tetraploid	RCeS	0.83 ^{c,e}	3.8	0.9236 ^{d,e}	2.67	0.929 ^{f,g}	2.87
Tetraploid	RCdS	0.786 ^{f,g}	2.87	0.903 ^g	2.08	0.91 ^g	2.25
Hexaploid	RCeS	0.9 ^{c,e}	6.58	0.952 ^{d,e}	4.35	0.957 ^{f,g}	4.82
Octoploid	RCeS	0.928 ^{c,e}	9.32	0.965 ^{e,h}	6.05	0.969 ^g	6.73

^a Absolute rate of loss of heterozygosity ($1-F$) per generation.

^b Generations to an equivalent loss of heterozygosity ($1-F$) as in diploids.

Work first performed by:

^c Haldane (1930);

^d Bartlett and Haldane (1934) using generation matrices;

^e Wright (1938) using path analysis;

^f Fisher (1947) using generation matrices;

^g Kempthorne (1957) using path analysis;

^h Parsons (1959a) using generation matrices.

the second principal component changes at a rate equal to, and the third principal component changes at a rate faster than, the rate of change of H in a selfing diploid (Fisher 1949; Bennett 1976).

Clearly the inbreeding coefficient does not measure all consequences of inbreeding in polyploids (Demarly 1963; Busbice and Wilsie 1966; Gallais 1967, 1974). Using the path analysis framework, the inbreeding of a tetraploid will increase the probability of two alleles being identical by descent, as well as increasing the probability that any three alleles and any four alleles taken at random from an individual will be identical by descent. Measures for these probabilities, similar to those developed for a pair of diploid zygotes by Gillois (1966) and Cockerham (1971), have been developed by Gallais (1967). Which covariance measure – as well as which principal component of the inbreeding matrix – is most critical for estimating the fitness consequences of inbreeding depends on the selective regime.

Haldane (1930) derived the equilibrial proportions of homozygotes in a mixed selfing–outcrossing population of tetraploids at a chromosomal segregating locus. Bennett (1968) explored more thoroughly a mixed mating population of tetraploids. Bennett derived the equilibrial values of F at various frequencies of double reduction. Figure 2 depicts the equilibrial inbreeding values of diploids, RCeS tetraploids, and RCdS tetraploids at different proportions of selfing. It is interesting to note that at low frequencies of selfing, the equilibrium inbreeding coefficient of a tetraploid may be greater than that of a diploid. These populations also differ in their rate of approach to equilibrium (Bennett 1968; Glendinning 1989). Double reduction increases the approach to equilibrium while selfing reduces it. The rate of approach is always faster in diploids than tetraploids.

Glendinning (1989), using a Monte Carlo simulation of a mixed mating tetraploid population, monitored the frequency distributions of the five zygote types described in Table 1. Glendinning demonstrated that selection against inbred individuals (inbreeding depression) reduces the equilibrial inbreeding coefficient and quickens the populations approach to that equilibrium.

4.1 Analysis of gene frequencies

Wright (1938) developed expressions for the inbreeding coefficients of finite populations of polyploids with and without self-fertilization. He also suggests an approximation for the sampling variance of a random mating polyploid population. This effort could serve as the theoretical foundations for efforts to partition variation in gene frequencies within individuals, within populations, and between populations for polyploids as developed for diploids by Cockerham and Weir (Cockerham 1969,

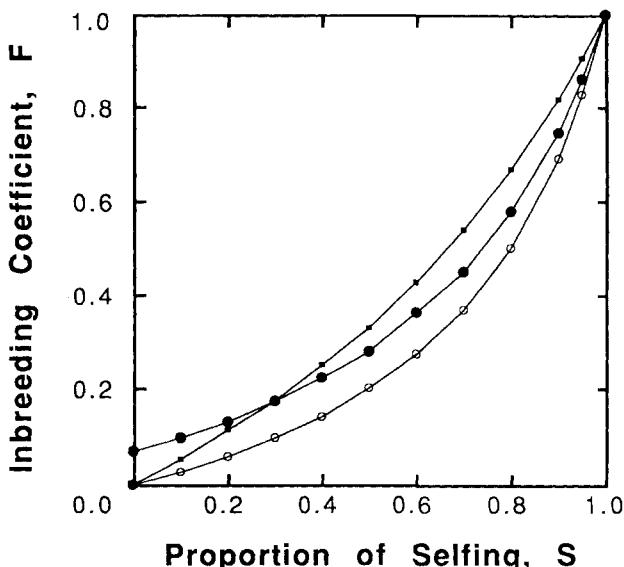


Fig. 2. The probability of two alleles being identical by descent, F , in the equilibrium population resulting from proportions s of selfing and $1-s$ of outcrossing of diploids (solid squares), RCeS tetraploids (open circles), and RCdS tetraploids (solid circles) (from Bennett 1968).

1973; Weir and Cockerham 1984). However, Wright does not develop specific expressions for the sampling variance and he does not include the possibility of double reduction. These developments, as well as work paralleling that of Cockerham and Weir, are needed before data on the distribution of gene frequencies in populations of polyploids can be properly evaluated.

5. LINKAGE BETWEEN LOCI

The determination of recombination frequencies is much more difficult in polyploids than in diploids. In diploids, a single test cross with a multiply homozygous recessive is sufficient to categorize the two modes of gamete formation: recombinant and non-recombinant. In polyploids, a gamete may contain one or more dominant alleles. Furthermore, even if there is no dominance, an Ab/aB and an AB/ab gamete cannot be separated without a second test cross. Fisher (1947) introduced the theory necessary to determine the recombination frequency in polyploids. He enumerated the modes of gamete formation for tetra- and hexaploids. There are, for example, 11 modes of gamete formation for two loci in a

tetraploid organism. To determine the recombination frequency in polyploids, the frequencies of these different modes of gamete formation can be estimated and then weighted appropriately. Fisher (1947) developed a maximum likelihood method of estimating the modes of formation from the genotypic frequencies resulting from the second test cross. Using this method, it is possible to estimate the recombination frequencies in a tetraploid between two loci, each with two alleles.

Recombination between two loci is expected to occur at a higher rate in polyploids than diploids. Assuming a constant cross-over rate per chromosome, Sved (1964) demonstrated that, unless they solely form bivalents, tetraploids have a higher recombination rate than diploids. Tetraploids, in fact, have a maximum recombination rate of $3/4$ compared with $1/2$ for diploids. Sved (1964) noted that none of the models he considered could explain the recombination rate of 0.282 ± 0.009 in diploid maize and 0.556 ± 0.062 in autotetraploid maize found by Oram (1959, from Sved 1964). Sved suggested that a higher cross-over rate per chromosome in tetraploids is necessary to explain these results.

The cross-over per chromosome rate may be expected to be higher in tetraploids than diploids if we assume a constant cross-over per chromatid rate. Cross-overs, presumably, can be initiated between any two chromatids; however, recombination results only from cross-overs between non-sister chromatids and not from those between sister chromatids. Cross-overs will be initiated between non-sister chromatids with a frequency of $2/3$ in diploids and $6/7$ in a random model of chromatid exchange for tetraploids (Model III of Sved 1964). Thus, under a random chromatid exchange model, the cross-over rate between chromosomes will be much higher in tetraploids than in diploids.

In polyploids, recombination within chromosomes and recombination resulting from reassortment of chromosomes into gametes must be distinguished. While the chromosomal recombination rate, the consequence of cross-overs, may be greater than $1/2$ in polyploids, the maximum probability of alleles at different loci segregating into different gametes is $1/2$. In fact, double reduction will reduce the proportion of recombinant gametes to below $1/2$.

Bennett (1954) developed expressions for the decay of linkage disequilibrium for two loci in random mating tetraploids and hexaploids. Crow (1954) derived the identical results for tetraploids using a simpler procedure. As in diploids, the linkage equilibrium frequency of gamete type AB equals the product of the two allelic frequencies. However, in polyploids, the rate of decay of linkage disequilibrium under random mating depends on the frequency of recombination, as in diploids, and also on the frequency of double reduction. The decay of linkage disequilibrium under random mating is a second order equation, compared with the first order decay function of diploids. Consequently, the decay

of linkage disequilibrium in a random mating polyploid population is a function not only of the level of linkage disequilibrium in the present generation but also on the level of linkage disequilibrium in the previous generation.

6. SELECTION

Little effort has been given to theoretical analyses of selection in populations of polyploids relative to that of diploids. This may be due to the fact that selection in polyploids is much more difficult to analyze theoretically than in diploids. The relative selective values of diploid genotypes can be described with two terms: the selective difference between homozygotes, $2s$; and the deviation of the heterozygote from the average of the two homozygotes, h . A similarly complete description of the selective values in tetraploids would require four terms for two alleles (Rowe and Hill 1984b). In light of the many possible selective regimes, the published work focuses on a few special cases.

Wright (1938) derived an expression for gene frequency change in response to selection on an additive allele. He then generalized this to an expression for the local response to average dosage effects. Haldane (1927) derived an expression for gene frequency change under selection for a dominant allele. From this, he concluded that the response to selection would be similar in diploids and tetraploids at low frequencies of the dominant allele but would be much slower in tetraploids at higher allelic frequencies.

The equilibrium gamete frequencies under various types of allelic overdominance were investigated by Parsons (1959b) and Li (1967). Focusing, for simplicity, on a two-allele RCeS locus under symmetrical overdominant selection, Parsons (1959b) demonstrated that in contrast to diploids, multiple internal equilibria can arise from simple selection patterns. Five equilibrial points are possible: two trivial ($p=0, q=0$), one center ($p=q=0.5$), and two side ($0 < p$ or $q < 0.5$). Table 3 gives the criteria for the existence and stability of these equilibrial points. Selfing and double reduction may move the population from one equilibrium point to another (Wöhrmann 1970a). If selfing rates are genetically determined, a higher selfing rate of a genotype has similar consequences as selection for that genotype (Wöhrmann 1970b).

Li (1967), more thoroughly exploring Parsons (1959b) examples, demonstrated several other unique aspects of polyploid population genetics. Firstly, selection on one locus generates gametic disequilibrium (Wright 1938; Li 1967; Hill 1971). Under conditions of internal equilibrium, selection generates permanent gametic disequilibrium. Li noted the similarities between the generation of this disequilibrium and the

Table 3

Conditions for equilibria under symmetric overdominant selection where a, b, and l are the fitnesses of the two homozygotes, the two digenic simplexes, and the digenic duplex, respectively. (From Parsons 1959b)

Fitness values		Equilibria		
a	b	Trivial	Asymmetric	Symmetric
<1/3	<1	u	—	s
>1/3, >b, <1	<1	s	u	s
>1/3, <b, <1	<1	u	—	s
<1	>1	u	—	s
>1	<a	s	—	u
>1	>a	u	s	u

u = unstable, s = stable.

disequilibrium generated by selection between loci in diploids. Secondly, in contrast to diploids, equilibrium gamete frequencies are not invariant with respect to linear transformations of the fitnesses. Finally, Li demonstrated that the average fitness of the population does not necessarily increase in response to selection. Hence, the technique for examining equilibria by maximizing the function for average fitness may not give meaningful results for tetraploid populations under overdominant selection.

The most complete analysis of the response of a polyploid population to selection is that of Hill (1971). Assuming a chromosomally segregating allele in a population of tetraploids at random mating equilibrium, Hill derived equations for the response to selection for an additive allele, a dominant allele, a recessive allele, a diploid dominant allele, and various types of overdominance. Hill concluded that the response to selection for an additive, completely dominant, or completely recessive allele was always greater in diploid than tetraploid populations. In tetraploid populations, the response to selection is not as dependent on selection intensity in selection for dominants as in selection for recessive or additive alleles. Direct comparisons of diploids and tetraploids in their response to overdominance is not possible because of the many possible forms of overdominance in tetraploids.

As mentioned above, selection in polyploid populations generates gametic disequilibrium (Wright 1938; Hill 1971). All of the theoretical work described above assumes the populations are in RME. Because this will not be true after the first generation, the response to selection after the first generation is not easily described. Rowe (1982) demonstrated that at a given gene frequency, gametic disequilibrium may increase or

decrease the response to directional selection from that expected at RME. The response to directional selection will be greater than expected if the frequency of the AA gamete, $f(AA)$, is greater than p^2 and, conversely, less than expected if $f(AA)$ is less than p^2 , where p equals the frequency of A, the favored allele.

Another consequence of the disequilibrium caused by selection is that after selection stops, the population mean may change as it approaches RME (Hill 1971). For directional selection, the population mean will decrease unless the allelic expression is additive, in which case the population mean does not change. This change in the mean will not be large if the population mean is high. For overdominance, the population mean may change in either direction as it approaches RME. Hill points out that because selection is not likely to remove all genetic variation in polyploid populations, breeders need to be conscious of this change in the phenotype following relaxation of selection.

7. ESTABLISHMENT OF A POLYPLOID AND GENE FLOW ACROSS PLOIDY LEVELS

Diploids can generate polyploids through polyspermy, zygotic chromosome doubling, meristematic chromosome doubling, and gametic non-reduction (de Wet 1980; Grant 1981, p. 309). Conversely, tetraploids can generate diploids through viable unfertilized polyhaploid gametes (de Wet 1980). Gene flow may also occur through viable triploid individuals (e.g. Vardi 1974, Lumaret 1988b). Mendiburu and Peloquin (1976) detailed examples and the genetic consequences of some of these processes.

Polyploids originate in diploid populations. Their survival is tightly linked to the effect of gene flow. It may contribute to the origin of new polyploids and have favorable or unfavorable genetic consequences on the establishment of the polyploid. Investigations of the conditions for the establishment of a polyploid are essential for an evolutionary analysis, yet only three theoretical papers have addressed this question.

Levin (1975) proposed the minority cytotype exclusion model. Intercytotype pollinations were supposed to be ineffective and intracytotype pollinations were assumed to produce viable offspring. Assuming random mating within and between cytotypes, the intercytotype pollination for one cytotype was equal to the frequency of the other cytotype. Consequently, the lower the frequency of the cytotype was, the higher the proportion of intercytotype (ineffectual) pollination it had. Thus for each generation the minority cytotype produced proportionally less offspring than the majority cytotype, leading to the extinction of the minority cytotype. Self-compatibility reduced the amount of intercytotype

pollinations by multiplying its value by $(1 - s)$ where s corresponds to the level of autogamy. Consequently, the extinction rate of the minority cytotype slowed down. The proportion of immigrating seeds necessary to compensate for the decreased fitness of the minority cytotype was small for low frequency of the minority cytotype and increased if the frequency of the cytotype approached 0.25–0.30 and decreased again as it attained 0.5. Selfing reduced the amount of migrants necessary to maintain equilibrium. Immigration via pollen dispersal necessitated a higher level of pollen flow than for seeds, with increasing values as the frequencies of the cytotype decreased. Levin concluded that a polyploid might not get established unless polyploid seed is dispersed to a suitable area for diploids, which by chance is uninhabited, or unless niche differentiation occurred.

Fowler and Levin (1984) proposed a system of two differential equations based on the Lotka-Volterra model to test the effect of niche differentiation on the establishment of a tetraploid, assuming triploids are lethal. In a first model, intercytotype pollinations were supposed to be independent of niche overlap. An extra term was added to the equations which expressed the negative effect of intercytotype pollinations on population growth. This term, which was frequency dependent, took a negative value if the cytotype was below the minority limit and a positive one if it was higher. In a second model, pollination was set proportional to niche overlap by introducing a frequency-dependent term in the competition coefficient of the equations. Both models allowed coexistence of the cytotypes under a range of values, some for the establishment of each of the cytotypes and one in which the initial frequencies of the cytotypes determine the outcome. The establishment of a polyploid was likely only in a small population, owing to chance effects.

Felber (1991) extended the previous models by allowing spontaneous polyploids to arise by the fusion of $2n$ gametes of the diploid. The future of a pure diploid population or of a mixed diploid-tetraploid population was investigated when the diploid cytotype produced $2n$ gametes. Diploids and tetraploids could have different relative fertility and viability. Triploids were lethal. Recursion equations of the frequency of the cytotypes were solved analytically or simulations were performed for mathematically complex cases. When the diploid produced no $2n$ gametes, an unstable equilibrium occurred that depended only on the relative fertility (f) and viability (v) of the cytotypes (f^2v/f^2v+1). If each of these components of fitness equalled 1, then an unstable equilibrium existed when the cytotypes were in equal proportions. Under these conditions, the minority cytotype was excluded, confirming the results of Levin (1975).

If the diploid cytotype produced a fixed proportion of $2n$ gametes, then a stable equilibrium was reached if the starting point was a pure diploid population. If the starting proportion of diploids was higher than

the unstable equilibrium, the population evolved towards the stable equilibrium, but if it was lower, the tetraploid population established. As the diploid produced more $2n$ gametes, a lower initial frequency of tetraploids was necessary for their establishment. For a pure diploid starting population and if fertility and viability were equal for both cytotypes, stable equilibrium was reached for a frequency of $2n$ gametes lower than 17.16 per cent. The tetraploid population was established with higher values. If the fertility and the viability of the tetraploid varied, this limit fell and was 10 per cent when one parameter was twice that of the diploid. It dropped to 6 per cent if both of them were double that of the diploid.

The theoretical approach of the establishment of polyploid taxa is still an open area of research. Some extensions of the previous models have to be made in order to test the influence of viable triploids. The effects of breeding system, population size, and population subdivision should be studied more thoroughly.

Gene flow across reproductive barriers has been variously hypothesized and demonstrated to have ecological significance (Anderson 1953; Lewontin and Birch 1966; de Wet 1968; Barrett 1983; Syvanen 1986). These observations/hypotheses present very difficult empirical and theoretical questions. How important has gene flow across reproductive barriers been in evolution? Such gene flow seems to be particularly important in the initiation of ecological invasions (Anderson 1953; Lewontin and Birch 1966). A related question is whether gene flow across reproductive barriers is maintained by selection. Although we expect partial reproductive barriers to be reinforced by individual selection, do the occasional benefits of gene flow prevent the completion of reproductive isolation? The diploid–polyploid system provides an excellent model system for theoretical work on this question.

8. SELECTION TOWARD ALLOPOLYPLOID INHERITANCE

The expected fate of established autoploid populations has been a matter of considerable debate. Once in the polyploid state, selection could favor auto- or allopolyploid inheritance. When do we expect populations to evolve toward allopolyploid inheritance? We note that there is some confusion in the literature in discussions of allopolyploidization versus diploidization. We refer to allopolyploidization as the conversion to allopolyploid inheritance. Alternatively, diploidization is understood to involve the silencing and differentiation of function of homologous loci. While diploidization is undoubtedly an important evolutionary process (see Werth and Windham 1991), it will not be addressed in this review.

For a trait to respond to selection, it must be variable and heritable. The type of variation and inheritance that would allow response to selection for allopolyploid inheritance depends on the mechanism of chromosome pairing. Several models of chromosome pairing have been suggested. A mechanical model suggests that structural differences in chromosomes, such as inversions or deletions, are responsible for the preferential pairing of similar chromosomes (Sybenga 1969). Allopolyploidization would then involve selection for increased divergence between chromosomes, which would lead to more consistent bivalent formation. Hence, this model assumes that adjustments to allopolyploid inheritance involves selection for chromosomal variants.

Alternatively, Waines (1976) suggests that some plant species are 'preadapted' for allopolyploidization. He argues that since plant species commonly hybridize, there is selection for discernment of small differences among chromosomes even in diploid populations. This selection may favor genes at the Ph-like (pairing homologous) locus, which enables fine distinctions to be made. The hypothesis of two alleles at one Ph-like locus seems to explain the available data for graminoid species (Jackson 1982). Under this model allopolyploidization may involve selection for more stringent chromosome-pairing genes. Response to selection for allopolyploid inheritance would be expected to be quicker if the population did have variation at a Ph-like locus.

Selection is generally expected to favor allopolyploid inheritance. The allopolyploid meiotic system is believed to reduce the likelihood of unequal chromosome separation leading to a reduced frequency of aneuploidy and therefore to increased fertility compared with complete autopolyploid inheritance. However, Gottschalk (1985) suggested that reduced fertility in polyploids is not necessarily caused by meiotic irregularities. The response of polyploids to selection for fertility does not necessarily correlate with a decrease in multivalent formation (reviewed in Jackson and Casey 1980; Gottschalk 1985). Furthermore, many fertile species form multivalents at meiosis.

9. QUANTITATIVE GENETICS OF POLYPLOID POPULATIONS

While the development of quantitative genetics of polyploids has trailed that of diploids, attention has been given to quantitative genetics of polyploids because of the economic importance of autopolyploid crops, mainly alfalfa and potato. Theoretical work on the applied quantitative genetics of polyploids has been reviewed by Rowe and Hill (1984a) and Wricke and Weber (1986). Here we briefly review the general results of quantitative genetic work on polyploids.

Kempthorne (1955, 1957) provided a general treatment of the quantitative genetics of polyploid populations assuming a single locus under random chromosome segregation. Though Kempthorne treated the general case of a single locus of any polyploid, here we just review work on tetraploids. The genotypic value of a tetraploid individual, $A_i A_j A_k A_l$, can be divided into the population mean (μ), four genic effects (α), six digenic effects (β), four trigenic effects (γ), and a quadrigenic effect (δ). The genotypic variance, σ_G^2 , can be similarly partitioned into additive (σ_A^2), digenic (σ_D^2), trigenic (σ_T^2) and quadrigenic (σ_F^2) components.

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_T^2 + \sigma_F^2,$$

where,

$$1/4 \sigma_A^2 = \sum_i p_i \alpha_i^2,$$

$$1/6 \sigma_D^2 = \sum_{ij} p_i p_j \beta_{ij}^2,$$

$$1/4 \sigma_T^2 = \sum_{ijk} p_i p_j p_k \gamma_{ijk}^2,$$

and

$$\sigma_F^2 = \sum_{ijkl} p_i p_j p_k p_l \delta_{ijkl}^2.$$

The genetic covariance between individual X and Y is equal to $4\rho_1\sigma_A^2 + 6\rho_2\sigma_D^2 + 4\rho_3\sigma_T^2 + \rho_4\sigma_F^2$, where ρ_1 , ρ_2 , ρ_3 , and ρ_4 are the probabilities that one, two, three, or four genes, respectively, chosen at random from individuals X and Y are identical by descent. Assuming a population in random mating equilibrium, the genotypic covariance of relatives are given in Table 4. Note that as the parent-offspring covariance in diploids

Table 4

Genetic covariances of autotetraploid relatives (from Kempthorne 1957). Covariance
 $= K\sigma_A^2 + L\sigma_D^2 + M\sigma_T^2 + N\sigma_F^2$.

Relationship	K	L	M	N
Identical twin	1	1	1	1
Parent-offspring	1/2	1/6	0	0
Parent-grandoffspring	1/4	1/36	0	0
Full-sib	1/2	2/9	1/12	1/36
Uncle-nephew	1/4	2/216	0	0
Half-sib	1/4	1/4	0	0
Double first cousin	1/4	70/1296	2/216	64/4176

includes a portion of the additive \times additive variance component, the parent-offspring covariance includes a portion of the digenic variance component, even without inbreeding. Thus, in polyploids, the additive and the digenic components should be estimated to assess the heritable fraction of the variance. Gallais (1974) developed expressions for the covariance of relatives in a non-equilibrial random mating population and the covariance of inbred relatives.

Li (1957) explored the special case of two alleles in a tetraploid population using the method of successive linear regression. He demonstrated that under the special conditions of symmetric overdominance and $p=q=1/2$, the additive and trigenic components of variance will be zero, leaving

$$\sigma_G^2 = \sigma_D^2 + \sigma_F^2.$$

Li also demonstrated that the genetic variance of a diallelic tetraploid locus is increased by 20 per cent by random chromatid segregation. The additive variance component is similarly inflated by 20 per cent. Li was unable to derive expressions for the other three variance components under random chromatid segregation.

Gallais (1975) compared the response to selection of a tetraploid population with that of a diploid population. He noted that digenic effects in tetraploid and additive \times additive effects in diploids cause curvilinear responses to selection. The mean of such populations would decline on relaxation of selection and asymptotically approach an equilibrial population mean determined by the additive effects. That the population mean always recedes after relaxation of selection is a different result from that obtained from a population genetic analysis by Hill (1971). Gallais (1975) interpreted this discrepancy as being due to the assumption of many genes of small effects in his model. Gallais noted that the decline of the mean is more rapid in a tetraploid population than in a diploid population.

Ehlke and Hill (1987) compared the quantitative genetics of allo- and autotetraploids using direct product matrix algebra. Similarly to Gallais (1975), Ehlke and Hill noted that the additive, digenic, trigenic, and quadrigenic components of variance of autotetraploids are equivalent to additive, dominance + additive \times additive, additive \times dominance, and dominance \times dominance components of variance in allotetraploids, respectively. Assuming identical gene functions and gene frequencies and RCeS, the population mean and the total genetic variance of the allo- and autotetraploid populations were equal. However, the parent-offspring covariance and variance among siblings of allotetraploids is greater than that of autotetraploids. Because of this, Ehlke and Hill predicted that directional selection would proceed more quickly in allotetraploids than

in autotetraploids. Whether this prediction is robust to assumptions of other types of polyploid segregation has not been explored.

10. DISCUSSION

Theoretical work on polyploidy started in the 1930s, with analysis of gene segregation and population genetics, but these aspects were later neglected. Analysis of meiotic configuration and research in quantitative genetics of polyploids began in the late 1950s and received more regular attention, probably because of their agronomic significance.

The chronology of work on population genetics of polyploids is striking in comparison to equivalent work on diploids. Published work on population genetics of polyploids as well as that of population genetics as a whole has increased exponentially over time (Fig. 3). However, especially since the 1940s, work on polyploids has not kept pace with that on population genetics as a whole. We note that the three 'founders' of population genetics, Haldane, Wright, and Fisher, each devoted some energy to developing a theory for polyploid populations. Their efforts were followed by several other authors – Li, Bennett, and Parsons. Several early texts on population genetics even included chapters on polyploidy (Li 1955; Kempthorne 1957). However, since this early work, polyploidy has not received much attention by theoretical population geneticists. Recently published texts or reviews of population genetics have not included sections on polyploidy (Lewontin 1967; Roughgarden 1979; Wallace 1981; Hartl and Clark 1989).

Several factors are possibly responsible for the neglect of polyploid population genetics. Ironically, one factor that may have discouraged work on polyploidy is the very success of the 'Synthesis' of evolution and Mendelian genetics in the 1940s. Polyploidization as an evolutionary process is consistent with the view of evolution held by the 'mutationist school' of the beginning of this century (reviewed by Mayr and Provine 1980). According to it, new species are created by macromutational events. Conversely, polyploidization is not consistent with the much praised synthesis of genetics and evolution (reviewed in Mayr and Provine 1980), which generalized evolution as resulting from natural selection acting on continually varying, heritable traits. Antonovics (1987) suggested that the enthusiasm for the Synthesis may have discouraged work on polyploidy because of this inconsistency.

Alternatively, we might attribute the neglect of the study of polyploidy as a symptom of the historical disarticulation of botany and botanists from evolution (discussed in Mayr 1980; Stebbins 1980a). Evolutionary theorists were not familiar with properties unique to plants and botanists were not familiar with the tools and conclusions of the theorists.

A third factor that may account for the decline of interest in polyploidy

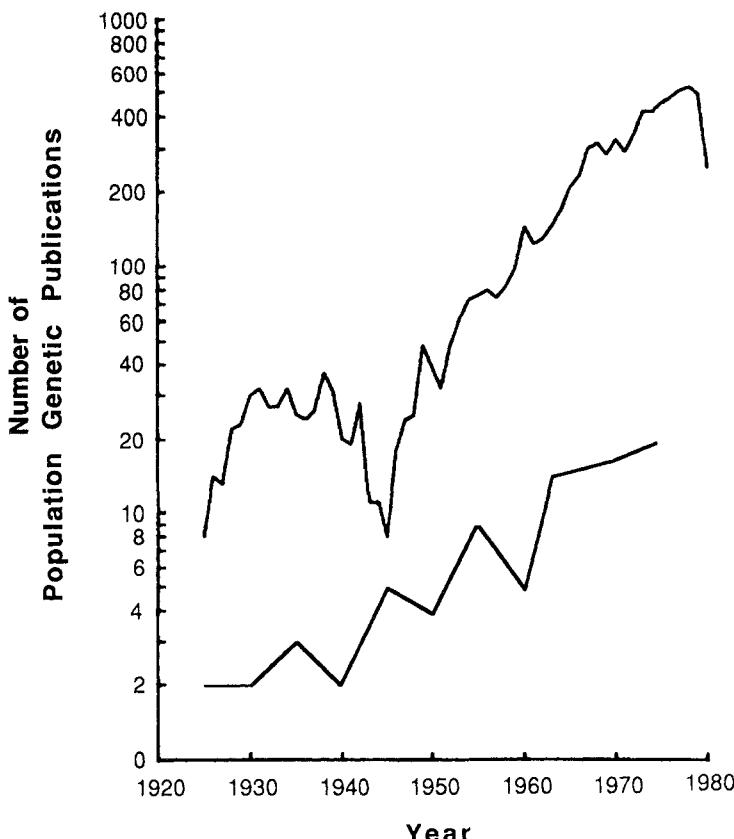


Fig. 3. Chronology of publications on theoretical polypliod population genetics (lower line) compared with those of theoretical population genetics as a whole (upper line) (from Felsenstein 1981). The abrupt drop in the upper line is an artefact of the termination of the citation collection (Felsenstein, pers. comm.) Publications on polypliod population genetics include all of those listed in Felsenstein under 'autotetraploid,' 'autopolyploid,' 'polyploid,' and 'tetraploid' as well as those referenced in the population genetics section of this review. The 5-year total of publications on polyploids is graphed.

was the overshadowing of the field of cytogenetics by the accomplishments in the fields of enzymology and molecular genetics. Thus, theoretical population geneticists were challenged by the quantity of allelic variation found through isozymes and molecular genetics, rather than by issues of variation in chromosome number.

10.1 Hypothesis of genetic mechanisms for polyploid success

Polyplodization is an important evolutionary process. While the life history and breeding system correlates of successful polyploidization is

much researched (reviewed in Mac Key 1970; Stebbins 1971; Jackson 1976; Ehrendorfer 1980), the genetic mechanism that accounts for the success of polyploids has received less rigorous attention. Polyploidy is acknowledged to be both a creative source of phenotypic variation in plant evolution (reviewed by Levin 1983; Lumaret 1988a) and a conservative force in maintaining variation (Stebbins 1950; Briggs and Walters 1984). We note three independent hypotheses for the genetic mechanism through which polyploidy, *per se*, is adaptive. While we did not originate these hypotheses, they have not previously been juxtaposed against each other.

1. The obvious hypothesis is that polyploidy allows new phenotypes through the increased dosages of functional genes. Under this hypothesis polyploidy may be adaptive so long as the optimal dosage is greater than two at a number of loci. This 'dosage-effect hypothesis' is implicit in much discussion of polyploidy. While the hypothesis is undoubtedly theoretically feasible, it has not been tested.
2. Alternatively, the success of polyploids may result from the reduced expression of deleterious recessives (Briggs and Walters 1984). The frequency of homozygotes in a polyploid is proportionally lower than that of a diploid. However, this lower expression will allow a higher equilibrium frequency of deleterious alleles in tetraploids owing to the balance of mutation and selection. Hence the feasibility of this 'deleterious recessive hypothesis' needs to be theoretically evaluated.
3. A popular hypothesis is that polyploidy is adaptive because it allows a high allelic diversity, which allows the expression of greater overdominance than in diploids (Mac Key 1970; Stebbins 1980b, 1985; Bingham 1980). Under this 'allelic diversity hypothesis,' an individual would be able to express genes with different functional optimas in different environments. We note that the allelic diversity hypothesis requires more than two functionally different alleles, while the previous two hypotheses require only two allelic states – functional and non-functional. Theoretical models confirm that allelic overdominance may favor polyploids (Bever, unpublished data). However, validating empirically that allelic diversity is the mechanism that allows the maintenance of polyploid populations is an extremely difficult task. In fact, allelic overdominance in diploids has rarely been demonstrated (Hughes and Nei 1988).

In addition polyploid establishment might occur by chance alone. This does not seem likely because of severe selection against odd ploidy levels. Polyploidy might be adaptive because of effects of multiple chromosomes themselves rather than the alleles. Under this hypothesis chromosomes have phenotypic effects independent of the genes of which they are composed.

Except for the allelic diversity hypothesis, no theoretical work has focused on these genetic hypotheses for polyploid success. These three hypotheses will be examined with respect to two different aspects: heterosis and inbreeding, followed by allopolyploidization.

10.1.1 Heterosis and inbreeding

While these three hypotheses are functionally distinct, they are not mutually exclusive, and empirically evaluating their relative importance is difficult. Overdominance is required for the allelic diversity hypothesis, yet the dosage effect and recessive-deleterious hypotheses can also account for the appearance of overdominance. For example, selection for optimal dosage could create overdominance. This would be the case if the optimum dosage for a given protein were two copies – the diploid equivalent (Fig. 4). Evidence for selection for intermediate dosages has

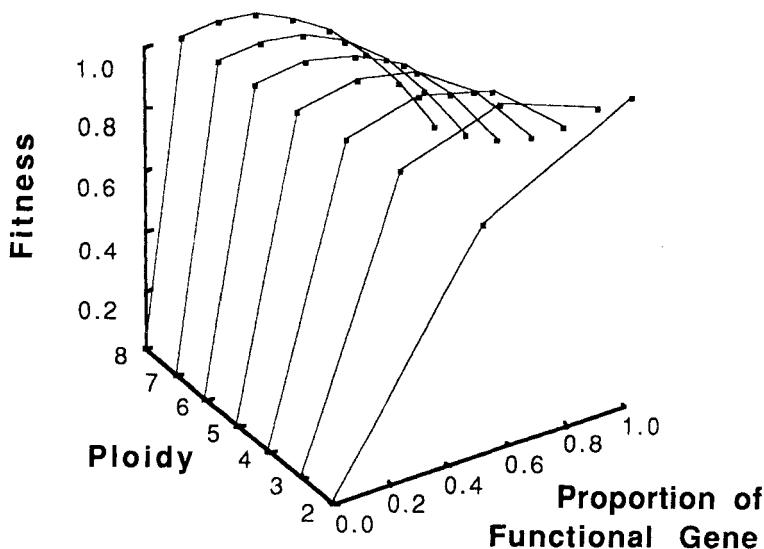


Fig. 4. This hypothetical fitness set illustrates that overdominance in polyploids resulting from the optimal level of functional genes does not necessarily confer an advantage to polyploids. The relative fitness of genotypes of diploids ($2N$) through octoploids ($8N$) are depicted for the following fitness function in which two functioning genes is optimal.

For all $f = 0$, $W = 0$.

For all $0 > f > 1$, $W = 1 - (2/p - f)^2$.

Where p equals the ploidy level and f equals the frequency of the functional allele ($f = (0..p)/p$).

been reviewed by Allendorf (1979). From the example presented in Fig. 4, we note that overdominance, *per se*, in the polyploids does not give them a selective advantage over diploids. Overdominance from selection for the optimal dosage of a functional protein and overdominance through selection for the expression of allelic diversity are functionally different, yet empirically difficult to distinguish.

The recessive-deleterious hypothesis does not predict overdominance at a single locus, but the appearance of overdominance is likely in a quantitative investigation because of linkage between recessive-deleterious alleles at one locus with operational alleles at other loci. The appearance of overdominance in diploid maize, for example, is likely to be due to the masking of recessive deleterious alleles (Moll *et al.* 1964).

Several workers have tried to test the allelic diversity hypothesis using quantitative investigations of inbreeding (reviewed in Bingham 1980). Specifically, they analyze the relative rate of accumulation of 'inbreeding depression' in diploids compared with that of polyploids. Inbreeding depression is the decline of a fitness character as a result of inbreeding. Inbreeding depression can be caused by the loss of overdominance (due to allelic diversity or gene dosage) or to the expression of deleterious recessives.

Expectations of how rapidly inbreeding depression accumulates in polyploids relative to diploids depends on assumptions about the selective regime. Because the inbreeding coefficient, F , increases much more slowly in polyploids than in diploids, some authors expected the effects of inbreeding to accumulate correspondingly slowly. However, the effects of inbreeding in the natural tetraploids, alfalfa and wheat grass, is more rapid than might be predicted by the increase in F . In fact, Tysdal *et al.* (1942) used the rapid accumulation of inbreeding depression in alfalfa as the basis for mistakenly identifying alfalfa as amphidiploid. Busbice and Wilsie (1966) argued that this rapid accumulation of inbreeding depression in tetraploids is due to the importance of multiple alleles. Multiple allelic combinations are reduced much more rapidly than F increases. This argument is the basis for much of the work undertaken to prove/disprove the allelic diversity hypothesis using the rates of accumulation of the effects of inbreeding.

Bennett (1976), however, pointed out that inbreeding depression may accumulate quickly in polyploids even at a digenic locus. The second and third principal components of a selfing tetraploid at a digenic locus change as fast or faster than F in diploids. The second and third principal components are likely to be important when q , the frequency of the functioning allele, is not equal to 0.5. Thus, rapid accumulation of inbreeding depression is possible under the hypothesis of selection for optimal gene expression.

Furthermore, Bennett (1976) pointed out that tetraploid populations

are expected to accumulate a higher frequency of recessive deleterious alleles than diploids owing to selection-mutation balance. Hence, a more rapid accumulation of inbreeding depression in natural tetraploids may also be consistent with the recessive-deleterious hypothesis. In fact, selection-mutation balance easily explains the observation that the synthesized tetraploids of maize or rye do not accumulate inbreeding depression as rapidly as natural tetraploids. The frequency of deleterious recessive alleles in diploids, and hence the newly synthesized tetraploids, may be much lower than that in natural tetraploids.

Parenthetically, Doyle (1986) suggested an independent process that may generate greater inbreeding depression in polyploids than in diploids. Doyle notes that irregularities at meiosis in polyploids generate aneuploidy with some regularity. Doyle modeled the expected frequency of aneuploids and the average fitness generated under four fitnesses regimes in a random mating and a selfing tetraploid population. Selfing generates a greater frequency of aneuploids and a correspondingly lower average fitness of the population. However, over most of the parameter space explored by Doyle, including the biologically reasonable values, the inbreeding depression resulting from aneuploidy was not large. Hence, Doyle demonstrated that, while differences in the frequency of aneuploidy may be a component of inbreeding depression in polyploids, it is unlikely to be a major component.

A substantial quantity of theoretical and empirical work has been performed on quantitative genetic effects of inbreeding in tetraploid populations. In the most comprehensive review to date, Bingham (1980) argued that the allelic diversity hypothesis is supported by this evidence. An adequate review of this work is beyond the scope of this paper. Here, we simply note that we cannot yet distinguish between these three alternative explanations of polyploid advantage or polyploid heterosis/inbreeding depression. Linkage of loci makes differentiation of these hypotheses extremely difficult. We also note that the distinction between these three hypotheses is directly relevant to Bingham's main thesis – maximizing allelic diversity will maximize yield. The plant breeder cannot be sure that the tedious procedures that maximize allelic diversity (Bingham 1980) would be expected to give better results than other selection procedures.

We emphasize that the current theoretical work on population genetics does not suggest that we will be able to differentiate the relative importance of the 'allelic diversity,' the 'deleterious recessive,' and the 'dosage effect' hypothesis through an analysis of inbreeding depression alone. The empirical and theoretical quantitative genetic work on inbreeding depression needs to be complemented by a range of theoretical population genetic work. Firstly, we need to analyze the magnitude of inbreeding depression resulting from different selective regimes at one

locus. As in diploids, the main categories to compare are overdominance, including that caused by allelic diversity and gene-dosage effects, and mutation-selection balance. In contrast to diploids, a wide diversity of models would exist within each of these categories. Secondly, the behavior of multiple locus systems under these different selection regimes might be interesting. Specifically, under what conditions are linkage groups maintained in a system similar to that of Franklin and Lewontin (1970).

10.1.2 *Allopolyploidization*

We can usefully address the issue of selection toward diploid inheritance in the context of the three genetic hypotheses for success of polyploids. If polyploids are successful because optimal dosages of genes are greater than 2, then allopolyploid gene inheritance may reduce the variance in gene dosages and thereby increase the populations average fitness. For example, if at one autopolyploid locus three functional genes are optimal, an allotetraploid population could remain homozygous for functional alleles at one locus and heterozygous at the second locus, limiting the range of dosages between 2 and 4 in the allopolyploid, compared with between 0 and 4 in the autopolyploid population.

Similarly, if polyploids are successful because of maximal allelic diversity, allopolyploid inheritance may likely stabilize this allelic diversity. For example, two-allelic diversity can be guaranteed by fixing different alleles at the two loci. However, it is likely that the favored inheritance pattern would depend on the specific selection regime. This issue requires a more thorough theoretical investigation. Alternatively, if the success of polyploids results from masking deleterious recessives, then allopolyploid inheritance would not be favored.

Much of the work on allopolyploidization focuses on the possibility that crop improvement in autotetraploids might occur more quickly if allotetraploid populations were created (Sybenga 1969). Ehlke and Hill (1987) do indeed predict that allopolyploid populations would respond to directional selection more quickly than an equivalent autopolyploid population. But whether allopolyploids are generally more successful than autopolyploids will likely depend on the genetic basis of the autopolyploid's establishment.

11. CONCLUSIONS

Polyplody presents challenging theoretical questions. Although much descriptive work has been undertaken on polyploids, basic questions about the origin, maintenance, and evolution of polyploid populations are still unaddressed. This failure partly stems from the lack of attention to theoretical analyses of these important evolutionary processes.

ACKNOWLEDGMENTS

We are indebted to the encouragement of Janis Antonovics, without whom this study would not have been completed. We thank Janis Antonovics and Dennis Rowe for critical reading of the manuscript. F. F. was supported by Grant 31-25736.88 from the Swiss National Foundation for Scientific Research through the duration of this study. J. B. acknowledges the support of NIH training grant 5T32GMO7754.

REFERENCES

- Allendorf, F. W. (1979). Rapid loss of duplicate gene expression by natural selection. *Heredity* **43**, 247–58.
- Alonso, L. C. and Kimber, G. (1981). The analysis of meiosis in hybrids. II. Triploid hybrids. *Can. J. Genet. Cytol.* **23**, 221–34.
- Anderson, E. (1953). Introgressive hybridization. *Biol. Rev.* **28**, 280–307.
- Antonovics, J. (1987). The evolutionary dys-synthesis, which bottles for which wine? *Am. Nat.* **129**, 321–31.
- Averett, J. E. (1980). Polyploidy in plant taxa, summary. In *Polyplody, biological relevance* (ed. W. H. Lewis), pp. 269–73. Plenum Press, New York and London.
- Barrett, S. C. H. (1983). Crop mimicry in weeds. *Econ. Bot.* **37**, 255–82.
- Bartlett, M. S. and Haldane, J. B. S. (1934). The theory of inbreeding in autotetraploids. *J. Genet.* **29**, 175–80.
- Bennett, J. H. (1954). Panmixia with tetrasomic and hexasomic inheritance. *Genetics* **39**, 150–58.
- (1968). Mixed self- and cross-fertilization in a tetrasomic species. *Biometrics* **24**, 485–500.
- (1976). Expectations for inbreeding depression on self-fertilization of tetraploids. *Biometrics* **32**, 449–52.
- Bingham, E. T. (1980). Maximizing heterozygosity in autopolyploids. In *Polyplody, biological relevance* (ed. W. H. Lewis), pp. 471–90. Plenum Press, New York.
- Briggs, D. and Walters, S. M. (1984). *Plant variation and evolution*. Cambridge University Press, Cambridge.
- Busbice, T. H. and Wilsie, C. P. (1966). Inbreeding depression and heterosis in autotetraploids with application to *Medicago sativa*. *Euphytica* **15**, 52–67.
- Cockerham, C. C. (1969). Variance of gene frequencies. *Evolution* **23**, 72–84.
- (1971). Higher order probability functions of identity of alleles by descent. *Genetics* **69**, 235–46.
- (1973). Analyses of gene frequencies. *Genetics* **74**, 679–700.
- Crane, C. F. and Sleper, D. A. (1989a). A model of meiotic chromosome association in triploids. *Genome* **32**, 82–98.
- (1989b). A model of meiotic chromosome association in tetraploids. *Genome* **32**, 691–707.

- Crow, J. F. (1954). Random mating with linkage in polysomics. *Am. Nat.* **88**, 431–34.
- Darlington, G. D. (1937). *Recent advances in cytology*. Churchill Livingstone, London.
- Demarly, Y. (1963). Génétique des tétraploïdes et amélioration des plantes. *Ann. Amélior. Plantes* **13**, 307–400.
- de Wet, J. M. J. (1968). Diploid-tetraploid-haploid cycles and the origin of variability in *Dichanthium* agamospecies. *Evolution* **22**, 394–7.
- (1971). Polyploidy and evolution in plants. *Taxon* **20**, 29–35.
- (1980). Origins of polyploids. In *Polyplodiy, biological relevance* (ed. W. L. Lewis), pp. 3–15. Plenum Press, New York.
- Dobzhansky, T., Ayala, F. J., Stebbins, G. L. and Valentine, J. W. (1977). *Evolution*. Freeman and Co, San Francisco.
- Doyle, G. G. (1986). Aneuploidy and inbreeding depression in random mating and self-fertilizing autotetraploid populations. *Theor. Appl. Genet.* **72**, 799–806.
- Driscoll, C. J., Bielig, L. M. and Darvey, N. L. (1979). An analysis of frequencies of chromosome configurations in wheat and wheat hybrids. *Genetics* **91**, 755–67.
- Driscoll, C. J., Gordon, G. H. and Kimber, G. (1980). Mathematics of chromosome pairing. *Genetics* **95**, 159–69.
- Ehlke, N. J. and Hill, R. R. (1987). Quantitative genetics of allotetraploid and autotetraploid populations. *Genome* **30**, 63–9.
- Ehrendorfer, F. (1980). Polyploidy and distribution. In *Polyplodiy, biological relevance* (ed. W.H. Lewis), pp. 45–60. Plenum Press, New York and London.
- Espinasse, A. and Kimber, G. (1981). The analysis of meiosis in hybrids. IV. Pentaploid hybrids. *Can. J. Genet. Cytol.* **23**, 627–38.
- Favarger, C. (1967). Cytologie et distribution des plantes. *Biol. Rev. Cambridge Philos. Soc.* **42**, 163–206.
- (1984). Cytogeography and biosystematics. In *Plant biosystematics* (ed. W. F. Grant), pp. 453–76. Academic Press, Canada.
- Felber, F. (1991). Establishment of a tetraploid cytotype in a diploid population: Effect of relative fitness of the cytotypes. *J. Evol. Biol.* **4**, 195–207.
- Felsenstein, J. (1981). *Bibliography of theoretical population genetics*. Dowden, Hutchinson and Ross, Stroudsberg.
- Fisher, R. A. (1947). The theory of linkage in polysomic inheritance. *Phil. Trans. R. Soc. Lond. B.* **233**, 55–87.
- (1949). *The theory of inbreeding*. Academic Press, New York.
- and Mather, K. (1943). The inheritance of style length in *Typhrum salicaria*. *Ann. Eugen.* **12**, 1–23.
- Fortini, P. and Barakat, R. (1980). Genetic algebras for tetraploidy with several loci. *J. Math. Biology* **9**, 297–304.
- Fowler, N. L. and Levin, D. A. (1984). Ecological constraints on the establishment of a novel polyploid in competition with its diploid progenitor. *Am. Nat.* **124**, 703–11.
- Franklin, I. and Lewontin, R. C. (1970). Is the gene the unit of selection. *Genetics* **65**, 707–34.
- Gallais, A. (1967). Moyennes des populations tétraploïdes. *Ann. Amélior. Plant.* **17**, 215–27.

- (1974). Covariances between arbitrary relatives in autotetraploids with panmictic disequilibrium. *Genetics* **76**, 587–600.
- (1975). Selection with truncation in autotetraploids – comparisons with diploids. *Theor. Appl. Genet.* **46**, 387–94.
- (1977). Contribution à l'étude théorique et expérimentale de l'hétérosis chez une plante allogame autotétraploïde. Thèse de doctorat, Université d'Orsay, Paris XI.
- Geiringer, H. (1944). On the probability of theory of linkage in Mendelian heredity. *Ann. Math. Statistics* **15**, 25–57.
- (1948). Contribution to the hereditary theory of multivalents. *J. Math. Phys.* **1**, 246–78.
- (1949). Chromatid segregation of tetraploids and hexaploids. *Genetics* **34**, 665–84.
- Gillois, M. (1966). Note sur la variance et la covariance génotypiques entre apparentés. *Ann. Inst. Henri Poincar. Section B.* **2**, 349–52.
- Glendinning, D. R. (1989). Some aspects of autotetraploid population dynamics. *Theor. Appl. Genet.* **78**, 233–42.
- Gonshor, H. (1965). Special train algebras arising in genetics II. *Proc. Edinburgh Math. Soc.* **14**, 333–8.
- (1973). Contribution to genetic algebras II. *Proc. Edinburgh Math. Soc.* **18**, 273–79.
- Gottschalk, W. (1985). Polyploidy and its role in the evolution of higher plants. In *Advances in chromosome and cell genetics* (ed. A. K. Sharma and A. Sharma), pp. 1–43. Oxford et Ibh Publishing Co., New Delhi, Bombay, Calcutta.
- Grant, V. (1981). *Plant speciation*, 2nd edn. Columbia University Press, New York.
- Haldane, J. B. S. (1927). A mathematical theory of natural selection and artificial selection. Part 3. *Proc. Cambridge Phil. Soc.* **23**, 363–72.
- (1930). Theoretical genetics of autopolyploids. *J. Genet.* **22**, 359–72.
- Hartl, D. L. and Clark, A. G. (1989). *Principles of population genetics*. Sinauer Assoc., Sunderland, MA.
- Heuch, I. (1978). The genetic algebra for polyploidy with an arbitrary amount of double reduction. *J. Math. Biol.* **6**, 343–52.
- Hill, R. R. (1971). Selection in autotetraploids. *Theor. Appl. Genet.* **41**, 181–6.
- Holgate, P. (1966). Genetic algebras associated with polyploidy. *Proc. Edinburgh Math. Soc.* **15**, 1–9.
- Hughes, A. L. and Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167–70.
- Hughes-Schrader, S. (1943). Meiosis without chiasmata in diploid and tetraploid spermatocytes of the mantid *Callimantis antillarum* Saussure. *J. Morphol.* **73**, 111–41.
- Jackson, R. C. (1976). Evolution and systematic significance of polyploidy. *Ann. Rev. Ecol. Syst.* **7**, 209–34.
- (1982). Polyploidy and diploidy: new perspectives on chromosome pairing and its evolutionary implications. *Amer. J. Bot.* **69**, 1512–23.

- (1984). Chromosome pairing in species and hybrids. In *Plant biosystematics* (ed. W. F. Grant), pp. 67–86. Academic Press, Canada.
- and Casey, J. (1980). Cytogenetics of polyploids. In *Polyplody, biological relevance* (ed. W. H. Lewis), pp. 17–44. Plenum Press, New York and London.
- and Casey, J. (1982). Cytogenetic analyses of autopolyploids: models and methods for triploids to octoploids. *Am. J. Bot.* **69** (4), 487–501.
- and Hauber, D. P. (1982). Autotriploid and autotetraploid cytogenetic analyses: correction coefficient for proposed binomial models. *Am. J. Bot.* **69** (4), 644–6.
- John, B. and Henderson, S. A. (1962). Asynapsis and polyploidy in *Schistocerca paranensis*. *Chromosoma* **13**, 111–47.
- Kempthorne, O. (1955). The correlation between relatives in a simple autotetraploid population. *Genetics* **40**, 168–74.
- (1957). *An introduction to genetic statistics*. Wiley and Sons, New York.
- Kihara, H. and Ohno, T. (1926). Chromosomenzahlen und systematische Gruppierung der Rumex-Arten. *Zeitschr. f. Zellforsch. Mikro Anat.* **4**, 475–81.
- Kimber, G. and Alonso, L. C. (1981). The analysis of meiosis in hybrids. III. Tetraploids hybrids. *Can. J. Genet. Cytol.* **23**, 235–54.
- Kimber, G. and Hulse, M. M. (1978). The analysis of chromosome pairing in hybrids and the evolution of wheat. *Proc. 5th Int. Wheat Genet. Symp. New Delhi*, pp. 63–72. Indian Society of Genetics and Plant Breeding, New Delhi.
- Kimber, G., Alonso, L. C. and Sallee, P. J. (1981). The analysis of meiosis in hybrids. I. Aneuploid hybrids. *Can. J. Genet. Cytol.* **23**, 209–19.
- Lawrence, W. J. C. (1929). The genetic and cytology of Dahlia species. *J. Genet.* **21**, 125–59.
- Levin, D. A. (1975). Minority cytotype exclusion in local plant populations. *Taxon* **24**, 35–43.
- (1983). Polyploidy and novelty in flowering plants. *Am. Nat.* **122** (1), 1–25.
- Lewis, H. (1967). The taxonomic significance of autopolyploidy. *Taxon* **16**, 267–71.
- Lewontin, R. C. (1967). Population genetics. *Ann. Rev. Genet.* **1**, 37–70.
- and Birch, L. C. (1966). Hybridization as a source of variation for adaptation to new environments. *Evolution* **20**, 315–36.
- Li, C. C. (1955). *Population genetics*. University of Chicago Press, Chicago.
- (1957). The genetic variance of autotetraploids with two alleles. *Genetics* **42**, 583–92.
- (1967). Genetic equilibrium under selection. *Biometrics* **23**, 397–484.
- Löve, A. (1960). Taxonomy and chromosomes – a reiteration. *Feddes Repertorium* **63**, 192–203.
- Lumaret, R. (1988a). Adaptive strategies and ploidy levels. *Acta Oecologia* **9**, 83–93.
- (1988b). Cytology, genetics, and evolution in the genus *Dactylis*. *CRC Critical Rev. Plant Sci.* **7**, 55–91.
- Malecot, G. (1948). *Les Mathématiques de l'Hérédité*. Masson et Cie, Paris.
- Mather, K. (1935). Reductive and equational separation of the chromosomes in bivalents and multivalents. *J. Genet.* **30**, 53–78.
- (1936). Segregation and linkage in autotetraploids. *J. Genet.* **32**, 287–314.

- Mac Key, J. (1970). Significance of mating systems for chromosomes and gametes in polyploids. *Hereditas* **66**, 165–76.
- Mayr, E. (1980). Introduction to botany. In *The Evolutionary Synthesis* (ed. E. Mayr and W. Provine). Harvard University Press, Cambridge, MA.
- and Provine, W. (eds) (1980). *The Evolutionary Synthesis*. Harvard University Press, Cambridge, MA.
- McGuire, P. E. and Dvorak, J. (1981). Relationship between the number of homologous chiasma frequency and the frequency of telosome pairing in wheat. *Genetics* **98**, 589–96.
- Mendiburu, A. O. and Peloquin, S. J. (1976). Sexual polyploidization and depolyploidization, some terminology and definitions. *Theor. Appl. Genet.* **48**, 137–43.
- Moll, R. H., Lindsey, M. F. and Robinson, H.F. (1964). Estimates of genetic variances and level of dominance in maize. *Genetics* **49**, 411–23.
- Müller, H. J. (1914). A new mode of segregation in Gregory's tetraploid Primulas. *Am. Nat.* **48**, 508–12.
- Müntzig, A. (1936). The evolutionary significance of autopolyploidy. *Hereditas* **21**, 263–78.
- Oram, R. N. (1959). Genetics of an autotetraploid maize. PhD Thesis, University of Adelaide.
- Parsons, P. A. (1959a). Some problems in inbreeding and random mating in tetrasomics. *Agron. J.* **51**, 465–7.
- (1959b). Equilibria in auto-tetraploids under natural selection for a simplified model of viabilities. *Biometrics* **15**, 20–9.
- Phillips, L. L. (1964). Segregation in raw allopolyploids of *Gossypium*. V. Multivalent formation in new world \times Asiatic and new world \times wild American hexaploids. *Am. J. Bot.* **51**, 324–9.
- Roughgarden, J. (1979). *Theory of population genetics and evolutionary ecology, an introduction*. Macmillan, New York.
- Rowe, D. E. (1982). Effect of gametic disequilibrium on selection in an autotetraploid population. *Theor. Appl. Genet.* **64**, 69–74.
- and Hill, R. R. (1984a). Theoretical improvement of autotetraploid crops: interpopulation and intrapopulation selection. *USDA Technical Bull. No. 1689*, 32pp.
- and Hill, R. R. (1984b). Effect of gametic disequilibrium on means and genetic variances of autotetraploid synthetic varieties. *Theor. Appl. Genet.* **68**, 69–74.
- Sarvela, P. A. (1958). Multivalent formation and genetic segregation in some allopolyploid *Gossypium* hybrids. *Genetics* **43**, 601–19.
- Sexton, O. J. (1980). Polyploidy in animal evolution: summary. In *Polyplody, biological relevance* (ed. W. H. Lewis), pp. 379–81. Plenum Press, New York and London.
- Seyffert, W. (1959). Theoretische Untersuchungen über die Zusammensetzung tetrasomer Populationen. II. Selbstbefruchtung. *Z. Vererbungsl.* **90**, 356–74.
- (1960). Theoretische Untersuchungen über die Zusammensetzung tetrasomer Populationen. I. Panmixie. *Biometrische Z.* **2**, 1–44.
- Stebbins, G. L. (1950). *Variation and evolution in plants*. Columbia University Press, New York.

- (1971). *Chromosomal evolution in higher plants*. Addison-Wesley, Reading, MA.
- (1980a). Botany and the synthetic theory of evolution. In *The Evolutionary Synthesis* (ed. E. Mayr and W. Provine), pp. 139–52. Harvard University Press, Cambridge, MA.
- (1980b). Polyploidy in plants, unsolved problems and prospects. In *Polyplodiy, biological relevance* (ed. W. H. Lewis), pp. 495–520. Plenum Press, New York.
- (1985). Polyploidy, hybridization and the invasion of new habitats. *Ann. Missouri Bot. Gard.* **72**, 824–32.
- and Dawe, J. C. (1987). Polyploidy and distribution in the European flora: a reappraisal. *Bot. Jahrb. Syst.* **108**, 343–54.
- Sved, J. A. (1964). The relationship between diploid and tetraploid recombination frequencies. *Heredity* **19**, 585–96.
- (1966). Telomere attachment of chromosomes. Some genetical and cytological consequences. *Genetics* **53**, 747–56.
- Sybenga, J. (1969). Allopolyploidization of autoploids I. possibilities and limitations. *Euphytica* **18**, 355–71.
- (1975). The quantitative analysis of chromosome pairing and chiasma formation based on the relative frequencies of M I configurations. VII. Autotetraploids. *Chromosoma* **50**, 211–22.
- (1988). Mathematical models for estimating preferential pairing and recombination in triploids hybrids. *Genome* **30**, 745–57.
- Syvanen, M. (1986). Cross-species gene transfer, a major factor in evolution? *Trends in Genetics* **2**, 63–6.
- Tysdal, H. M., Kiesselbach, T. A. and Westover, H. L. (1942). Alfalfa breeding. *Univ. Nebraska, Agric. Exp. St. Bull.* **124**, pp. 46.
- Vardi, A. (1974). Introgression from tetraploid durum wheat to diploid *Aegilops longissima* and *Aegilops speltoides*. *Heredity* **32**, 171–81.
- Waines, J. G. (1976). A model for the origin of diploidizing mechanisms in polyploid species. *Am. Nat.* **110**, 415–30.
- Wallace, B. (1981). *Basic population genetics*. Columbia University Press, New York.
- Weir, B. S. and Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–70.
- Werth, C. R. and Windham, M. D. (1991). A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. *Am. Nat.* **137**, 515–26.
- White, M. J. D. (1952). *The chromosomes*, 2nd edn. Methuen, London.
- Wöhrmann, K. (1970a). Untersuchungen zur Wechselwirkung von Selektion und Selbstungsrate auf das genetisch Gleichgewicht unter besonderer Berücksichtigung tetraploider Populationen. I. *Theor. Appl. Genet.* **40**, 267–9.
- (1970b). Untersuchungen zur Wechselwirkung von Selektion und Selbstungsrate auf das genetisch Gleichgewicht unter besonderer Berücksichtigung tetraploider Populationen. III. *Theor. Appl. Genet.* **40**, 296–9.
- Wörz-Busekros, A. (1978). Polyploidy with an arbitrary mixture of chromosome and chromatid segregation. *J. Math. Biol.* **6**, 353–65.

- Wricke, G. and Weber, W. E. (1986). *Quantitative genetics and selection in plant breeding*. Walter de Gruyter and Co., Berlin.
- Wright, S. (1938). The distribution of gene frequencies in populations of polyploids. *Proc. Natl. Acad. Sci. USA* **24**, 372–77.

This page intentionally left blank

A causal analysis of stages in allopatric speciation

WARREN D. ALLMON

1. INTRODUCTION

Every instance of allopatric speciation must involve at least three stages. First, a daughter population must form, that is, it must become separated, at least in some degree, from the parent population. Second, the daughter population must persist in this separated state long enough to become a discrete new species. Third, the daughter population must become differentiated from the parent population during this period of separation. Breaking the speciation process down into these three stages may allow for more explicit identification and analysis of causal factors involved. Examination of previous studies of speciation indicates that each of these stages has been considered before, but they have never been united into a comprehensive view. Such a view has a wide explanatory potential, yet this potential has gone largely unrealized because speciation usually has been seen as a single, albeit complex, process. In this paper I attempt to demonstrate the heuristic value and usefulness of dividing speciation into these three stages, specify potential causal factors underlying each stage, and explore some of the broader evolutionary implications of such an analysis.

This three stage formation–persistence–differentiation framework (Fig. 1) is the same for different types of allopatric speciation, whether separation is by dispersal and isolation of a small daughter population at the periphery of the geographic range of the parent ('peripatry' of Mayr 1982) or by a vicariance event dividing the parent into relatively large daughter populations ('dumbbell' speciation of Stebbins 1969, 'vicariance speciation' of Wiley 1981 or 'speciation by subdivision' of Bush 1975). It obviously does not apply to truly instantaneous speciation by polyploidy. Its application to other non-allopatric modes of speciation is problematic. In most models of sympatric speciation, daughter populations are not geographically separated prior to differentiation; differentiation is rather the initiator of speciation and must precede physical separation (e.g. Tauber and Tauber 1989). While such differentiation is going on, however, it is still the case that the differentiating subpopulations must persist long enough for reproductive isolation to be attained (cf. Wilson 1989, p. 382).

The timing of differentiation may also vary in allopatric speciation.

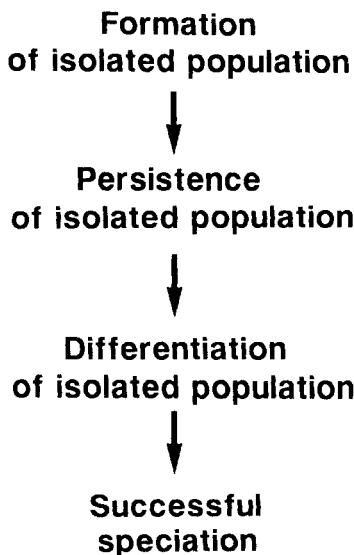


Fig. 1. Formation–persistence–differentiation framework for viewing the stages of the speciation process.

Mayr (1963), among others, has argued that geographic speciation often occurs by separation of already partly differentiated populations at the margins of a species' geographic range or at the end of clines. Indeed, as discussed further below, the timing and mechanism, as well as rate and degree, of differentiation are major variables in the speciation process. Differentiation is often discussed in studies of speciation, but is seldom put in context with other elements of the process. In this paper, I defer to discussions available elsewhere for details of the processes of differentiation, and emphasize instead the relationship of the stages of the speciation process to each other.

In the following discussion, I use a definition of an isolated population modified from those of Mayr (1963, p. 366) and Stanley (1979, p. 195): a population or group of populations prevented from free gene exchange with other populations of the same species, and having the potential to become a new species. This corresponds roughly to the more ecological definition of Taylor (1990, p. 429), who defines a population as ‘the unit within which occur interactions – reproduction, population regulation, predation – and within which most movement is confined.’

Metapopulations (Levins 1969, 1970) are ensembles of local populations connected by occasionally dispersing individuals; both populations and metapopulations have finite lifetimes, that is, expected times to extinction (Hanski and Gilpin 1991). The processes of population formation, persistence, and differentiation occur on what Hanski and Gilpin call the

'metapopulation scale . . . at which individuals infrequently move from one place (population) to another, typically across habitat types which are not suitable for their feeding and breeding activities, and often with substantial risk of failing to locate another suitable habitat patch in which to settle' (Hanski and Gilpin 1991, p. 7).

2. ISOLATE FORMATION

Several authors have suggested that many more daughter populations form than go on to become successful descendant species. Mayr (1963, p. 367), for example, states that '[w]e know, as yet, little about the frequency of genuine isolates in various groups of animals,' but nevertheless argues that '[m]ost species bud off peripheral isolates at regular intervals . . . [and n]early all of them either reestablish contact with the parental species or else die out' (1963, p. 554). Mayr estimates that 'peripheral isolates are produced 50 or 100 or 500 times as frequently as new species' (1963, p. 513). Stanley (1979, p. 175) similarly states that '[s]ome unknown percentage of the myriads of isolates issuing from any species must technically attain the status of a distinct species without actually being recognized as such because they are snuffed out before expanding to become fully established.' Stanley (1978) has labeled these ephemeral isolates 'aborted species.'

Because it is difficult or impossible to document and follow the history of every isolated population of a species, this argument has been supported chiefly by circumstantial evidence. Mayr (1963, pp. 513–4), for example, notes that four families of birds for which he tabulated data have many more isolates than subspecies, and many more weakly than strongly differentiated isolates. Comparing the list of species that seem to have originated in a single postulated Pleistocene forest refuge in Amazonia with the list of species that could potentially have been isolated in that refuge, Mayr (1969, p. 16) finds the number of new derived species to be relatively small. He explains this in part by the extinction of a large proportion of the isolated incipient species.

Direct field studies of the fates of populations are few, but also seem to suggest that many isolates become extinct for every one that becomes a successful species. Several marine snail species that have been studied, for example, are broken into many small, short-lived local populations that are repeatedly re-established by recolonization (Spight 1974; Quinn *et al.* 1989). Cain and Cook (1989) followed eight replicate populations of the land snail *Cepaea* in enclosures over almost 20 years; seven of the populations became extinct and the eighth almost did. Studies of the fates of local populations within metapopulations similarly suggest high rates of population extinction (e.g. Ebenhard 1991; Harrison 1991; and references therein).

Other evidence, however, suggests that rates of isolate formation may vary considerably, depending on a variety of factors. These factors can be divided usefully into those intrinsic and extrinsic to the organism.

2.1 Extrinsic mechanisms

Extrinsic mechanisms of isolate formation are divisible into those that act to divide parental populations by vicariance events, and those that provide habitat patches beyond the parental range for such division to occur by dispersal.

Factors causing vicariance may be either biotic or abiotic. The most commonly discussed abiotic vicariance process is the origin of a 'geographic barrier,' such as a river, lava flow, ocean basin, or mountain range within the range of the parent. Mayr (1963) has implied that such extrinsic factors may be the most important single factor governing the occurrence of speciation. Cracraft (1982, 1985) argues even more strongly for extrinsic control, contending that speciation rates are controlled primarily by large-scale changes in lithospheric complexity, leading to the isolation of larger numbers of daughter populations.

The most important biotic cause of vicariance may be predation. Menge and Sutherland (1976, p. 352) suggest that 'severe predation, by creating genetically disconnected allopatric populations of a species, can presumably lead eventually to speciation . . .' Stanley (1986) has suggested that high rates of speciation in non-siphonate Neogene bivalves may be due in part to high rates of predation fragmenting their populations.

If at least some isolation of incipient species takes place by dispersal to habitat patches outside the parental range, then there must be such patches of tolerable habitat to which dispersal can occur. This is a feature of the extrinsic environment, not the dispersing organisms themselves. To take a trivial example, if there were no Hawaiian islands, there would be no endemic Hawaiian *Drosophila* species. More interesting are less obvious cases such as the occurrence of sand and rock patches on the bottoms of African Great Lakes that may provide habitats for incipient cichlid fish species (e.g. Fryer and Iles 1972), as well as questions such as to what degree is the great diversity of shallow benthic invertebrates in the Indo-West Pacific today due to the presence of thousands of small isolated islands in this region (e.g. Kay 1984; Vermeij 1987b).

To the degree that physical subdivision of habitat can be assessed independently of the distribution of organisms in that habitat, it may offer an opportunity to quantify formation of isolates (= local populations; e.g. Harrison 1991). Using this logic for late Cenozoic ostracodes, Cronin and Ikeya (1991; see also Cronin 1985, 1987) identified environmental changes that could have isolated populations ('opportunities for speciation'), and compared them with the number of new species known

to have arisen during the same time interval. The opportunities consist of the rising of the Isthmus of Panama, which separated Atlantic and Pacific populations by a single vicariance event, and more numerous climatic fluctuations that conceivably could have allowed range expansions and contractions, resulting in fragmentation. During climatic fluctuations, no more than about 3 per cent of the opportunities resulted in new species, while across the Isthmus only 22.4 per cent of pre-existing species gave rise to daughter species.

2.2 Intrinsic mechanisms

Intrinsic mechanisms of isolate formation are those features of organisms themselves that make it more likely that populations will become isolated, either by vicariance or dispersal. Chief among these is dispersal ability. Mayr (1963, p. 569) has claimed that '[a]ny factor that reduces dispersal may facilitate speciation.' Yet this would only be true if all speciation were via vicariance and dispersal and acted only to retard speciation by enhancing gene flow (e.g. Ehrlich and Raven 1969). If at least some speciation occurs by isolation of dispersed propagules, dispersal may enhance probability of speciation.

There is thus some intermediate level of dispersal that maximizes probability of isolation: high enough to allow for colonization, but not so high that reproductive isolation is prevented. Janzen (1977), for example, attributes high species diversity of insects to the fortuitous effects of size on their dispersal ability: 'insects are not so small that they get carried everywhere . . . However, they are small enough such that much of the small-scale heterogeneity on the earth's surface will serve as a barrier to population flow, thereby maximizing the rate of differentiating new populations.' A similar situation has been discussed for birds by Ehrlich (1986).

Stanley (1991) has proposed a general principle to describe how dispersal affects speciation: 'For species that are characterized by stable, relatively continuous geographic distributions, effective dispersal will retard rate of speciation by opposing the formation of isolates. . . . on the other hand, for species characterized by patchy, unstable populations, effective dispersal will promote speciation by generating isolates . . .'

Population structure (i.e. the arrangement and composition of populations of a species in space) has been suggested to be of potential importance in probability of isolate formation (as well as in population subdivision without complete geographic isolation; cf. Wilson *et al.* 1975; Bush 1981). Dorit (1991), for example, has suggested that mouthbrooding in cichlid fishes in East African rift lakes might be responsible for a highly fragmented deme structure that might contribute to the prolific speciation in these animals. Vrba (1987), however, concludes that

population structure does not contribute to differential speciation rates in African antelopes.

The degree of specialization of a species to its environment ('ecological amplitude' of Wilson 1961, Bush 1975, and Antonovics 1976) may also contribute to probability of isolate formation. Habitats of more stenotopic species tend to be patchier, or at least more susceptible to fragmentation by environmental changes (e.g. Vrba 1987). Incidence of isolate formation, however, will be a product of both physical structure of the habitat and ecological amplitude of the parent species. In frogs of the species-rich Neotropical genus *Eleutherodactylus*, for example, both life history pattern and the nature of the environment may contribute to enhanced isolate formation, but so may the nature of the environment. Since they undergo terrestrial development, species of *Eleutherodactylus* are more dependent on the high humidity of continuous forest than are frogs that deposit their eggs directly in water (Lynch 1979b, p. 204). They thus tend to be less vagile and more sensitive to environmental barriers than are frogs of other genera (Lynch 1979a). Geography may also contribute to diversity in this group, however, since most of its species occur in the islands of the West Indies and in the highly subdivided foothills of the northern Andes (Lynch 1976).

Sharing some of the same forests with eleutherodactylines, hummingbirds show a very different pattern (Bleiweiss 1991). One subfamily (Phaethorninae) has low vagility, is ecologically specialized, and contains only 35 species. The other subfamily (Trochilinae) has high vagility, is ecologically generalized, and contains 295 species. Bleiweiss explains this difference by interaction of intrinsic and extrinsic factors favoring isolate formation and persistence. Trochilines have more species, he argues, because high vagility allows access to new habitats (enhancing isolate formation), and generalized ecologies permit use of novel resources encountered there (enhancing isolate persistence). In regions with complex topography, like the Andes or West Indies, where hummingbirds are most diverse, Bleiweiss suggests that high vagility is of particular importance since the physical environment provides barriers to gene flow among populations.

3. ISOLATE PERSISTENCE

Once they form, isolated populations face four possible fates (cf. Mayr 1969, p. 13). (1) They can disappear by merging with the parent population. (2) They can become extinct by the death of their component individuals. (3) They can persist in isolation without speciating. (4) They can persist and differentiate to become new descendant species. Different factors will affect the probabilities of each of these outcomes.

Disappearance of an isolate by merging with another (usually the parental) population can only occur before establishment of mechanisms preventing interbreeding, and so will be favored by delayed differentiation. It can also be favored by transience or incomplete development of extrinsic isolating barriers (see discussion of differentiation, below).

To the extent that isolated populations are vulnerable to extinction by the death of all of their constituent individuals, factors conducive to survival of those individuals, or of the isolate as a whole, will lead to higher probability of isolate persistence (cf. Stanley 1979; Glazier 1987a). Here I consider five such factors. By listing them separately here, I do not mean to imply that they are all truly independent; on the contrary, they interact and some are results or causes of others.

(1) *Population size and stability.* Abundance may be one of the most important variables affecting vulnerability to extinction in established species (e.g. MacArthur and Wilson 1967; Jackson 1974; Diamond 1984a; Stanley 1986); other factors being equal, smaller populations have a higher probability of extinction than larger populations. It is possible, therefore, that isolated populations that are able to expand rapidly to a larger size will have a greater chance of persistence than populations that lack this ability (e.g. Ebenhard 1991).

The difficulty of following the history of large numbers of often ephemeral populations makes this idea difficult to test. Data from small mammals suggest an important role for population size and stability (Anderson 1970; Smith 1974, 1980; States 1976; Glazier 1980, 1986, 1987a). Other studies have stressed the role of stability of population size, independent of what that size is, in isolate persistence (den Boer 1971, 1985; Roff 1974; Diamond 1984b; Glazier 1987a, and references therein). Studies from conservation biology on the 'minimum viable population' necessary to preserve isolated fragments of formerly continuous species ranges also support an important role for population size (e.g. Schoener and Schoener 1983; Diamond 1984b; Gilpin and Soule 1986; Soule 1987).

To the degree that populations are ultimately limited by energy (Odum 1971; Ricklefs 1973; Colinvaux 1978), rather than by predators, interference competition, or competition for non-trophic resources, availability of trophic resources may affect persistence by affecting population size (Belovsky 1987). Glazier (1987a, p. 328) summarizes the argument succinctly: 'Isolates using relatively abundant resources supporting relatively dense populations should have a better chance of surviving long enough to differentiate into new species than those that do not.' The role of resources in speciation is discussed in further detail below.

(2) *Ecological amplitude.* Ecological specialization has often been claimed to correlate positively with speciation rate (e.g. Vrba, 1980).

One interpretation of this correlation is that, as discussed above, specialized species are more likely to form isolates. Another is that these isolates are more likely to persist once they are formed. Eldredge (1979) and Eldredge and Cracraft (1980) argue that sympatric generalists will exclude, while sympatric specialists will accommodate, each other. Stenotopy, as Eldredge (1979, p. 14) puts it, 'increases the probability of survival of new species budded off by accidents of changing geography.' Highly derived species may in general have a greater chance of survival (Lewis 1966; Eldredge 1979, 1989, p. 150). In the initial stages of colonization, however, generalist species may have higher rates of propagule survival (e.g. Ebenhard 1991).

(3) *Environmental rigor.* All other factors being equal, it should be easier to survive in an 'easier' physical environment. Isolates that occur in such environments may therefore show higher rates of persistence than isolates in 'harsher' physical environments (e.g. Stanley 1979). As discussed further below, this may apply in particular to tropical species.

(4) '*Niche space*'. An environment with a larger number of different ways of making a living (i.e. 'more niches') should contain a larger number of discrete and different populations. This has long been a popular explanatory device for different levels of diversity in different communities. More diverse communities have usually been said to have a greater diversity of niches (e.g. MacArthur 1972), the notorious difficulty of independent definition of 'niches' notwithstanding (e.g. Lewontin, 1982, 1983). As is also widely acknowledged, this argument is essentially one of maintenance, rather than of origin, of diversity. Yet it can also be applied to the question of origin of species. Mayr has done just this when he states that '[w]e see again and again that an incipient species can complete the process of speciation only if it can find a previously unoccupied niche' (1963, p. 574). In the same vein, Templeton (1989, p. 18) states that 'for an adaptive transition to persist, a niche must be available for the organisms with the new adaptation.'

(5) (*Adaptation*). All other factors being equal, populations of individuals that possess morphological, behavioral, or biochemical characteristics that enhance their relative fitness in their local environments ('aptations' sensu Gould and Vrba (1982)) should be able to persist longer (and more often) than populations of individuals that lack such features. (The characteristics may arise via natural selection for the function they serve, in which case they are adaptations sensu stricto. Or they may arise for other reasons, in which case they can be called exaptations (Gould and Vrba 1982). Although distinct, for present purposes they will both have the same result, and are not hereafter distinguished.) At least some differences between local populations of a species have functional, and perhaps adaptive, explanations (e.g. Mayr 1963; Endler 1986; Nevo 1986; Wilson 1989). If these populations remain

isolated from each other, such adaptations may allow them to persist for long enough to differentiate into new species, and may in fact be part of the differentiation process (see below).

Stanley (1979) and Glazier (1987a) have stressed that, to the degree that different daughter populations (i.e. 'avatars', Damuth 1985) derived and isolated from the same parent differ in characteristics that render them differentially susceptible to extinction, selection will take place at the level of the isolate, and any characteristic 'which would increase survivorship is expected to be preferentially represented in species newly formed from those isolates' (Glazier 1987a, p. 325). The result of such a process, Stanley (1979, p. 197) suggests, will be bias in the direction of speciation, since descendant species will have been drawn from only a subset of ancestral variation (see also Lloyd 1988 on 'avatar selection').

Many of the factors affecting isolate persistence, discussed above, have been treated by some previous authors as factors affecting extinction rates of established species. The framework presented here, in effect, proposes that attention be shifted backward a step to examine the role of extinction on speciation itself. The distinction is a very important one for evolutionary conclusions. In one case speciation proceeds at the same rate, and only established species are affected. In the other it is the occurrence of speciation itself that changes. It may be that these two are indistinguishable in practice, but this does not change their distinction in theory or in nature.

4. ISOLATE DIFFERENTIATION

Differentiation is in many ways the most complex of the three stages into which the speciation process is here divided, principally because it achieves a variety of results, by a variety of mechanisms, at a variety of times. The nature of population differentiation in speciation can change predictions and conclusions derived only from consideration of isolate formation and persistence and, as already mentioned, can be interwoven with these other two stages in the speciation process. These complexities notwithstanding, consideration of differentiation as a separate aspect of speciation can contribute to greater clarity in judging the role of both it and the other stages.

Differentiation associated with speciation does two things: it produces reproductive isolation, and it (probably usually) produces morphological differences that characterize most closely related species. That these are separable is demonstrated by the existence of all possible combinations between them: morphologically similar populations that are not reproductively isolated (single species); morphologically distinct populations that are reproductively isolated (distinct separate species); morphologically

very similar or identical populations that are reproductively isolated (sibling species); and morphologically distinct populations that are not reproductively isolated (intraspecific variation). Attainment of reproductive isolation itself does not appear to involve much genetic change (Lewontin 1974; Ayala 1975; Templeton 1980, 1981, 1982; Nevo 1989; Lessios and Cunningham 1990).

Two schools of thought have existed on how these two types of differentiation relate in geographic speciation. One view, argued principally by Mayr (1942, 1963; see Mayr 1988 for a recent view), and more recently by Paterson and colleagues (Paterson 1985; see also Butlin 1989 and references therein), suggests that isolating mechanisms are acquired in allopatry as accidental byproducts of genetic differentiation between isolated populations. The other view, argued principally by Dobzhansky (1937, 1940, 1976; see also, for example, Lewontin 1974; Avise 1978), holds that isolating mechanisms are formed only incompletely in allopatry, and become fully developed by reinforcement only after secondary contact.

In ways that are as yet incompletely understood, both types of differentiation result from some combination of at least the following processes: selection (natural and/or sexual), random drift and fixation, enhanced pleiotropy, chromosomal rearrangements, intrinsic molecular processes within the genetic material, and genetic reorganization resulting from processes such as 'genetic revolutions' or 'genetic transilences.' In both the Mayr and Dobzhansky views, selection plays an important role in differentiation, albeit at different times. In Mayr's view selection acts on traits with survival value, rather than on reproductive isolation itself; in Dobzhansky's view, reproductive isolation is selected for directly in secondary sympatry.

Although Mayr's original (1954) preferred genetic explanation for the significance of the founder effect, i.e. a decrease in variation leading to a relatively rapid reorganization of the genome of a small isolated population, has been largely discounted (e.g. Lande 1980; Carson and Templeton 1984; Barton and Charlesworth 1984; since such reduced variation would also reduce potential to react to selection), population bottlenecks may still disrupt 'governing equilibrium genetic relationships' (Bryant and Meffert 1990; see also Goodnight 1987) and so contribute directly to differentiation. Such still poorly understood rearrangement processes aside, rapid differentiation in such small populations remains a valid hypothesis. Variation may in fact remain sufficiently high to allow rapid selection in such populations (Lande 1980). Random fixation by genetic drift is always a potentially powerful force in small populations (Lande 1976). Differentiation by alteration of chromosomal architecture may be more likely in small populations (Lande 1979; Bush 1981; Wright 1982), whether or not they are completely isolated. Processes resembling

genetic 'revolutions' or 'transilences' may occur (Carson and Templeton 1984), but they are not required to grant small populations high potential for differentiation (e.g. Barton and Charlesworth 1984; Nevo 1989; Baker *et al.* 1990).

Differentiation may occur prior to formation of isolated populations (i.e. intraspecific variation), after isolate formation but prior to establishment of sympatry, or, as already mentioned, after establishment of sympatry between partially isolated 'neospecies.' To the degree that processes of intraspecific differentiation such as described in Wright's shifting balance model are conducive to speciation (Wright 1982), population structures consistent with the model will contribute to differentiation prior to physical separation of populations. Among plants at least, as discussed by Antonovics and Levin (1980), intraspecific (among population) differentiation may be greater in high than in low population densities, because of reduced gene flow distances with increased density. Dense, semi-isolated populations may be less affected by gene flow (and so be more likely to differentiate) than less dense populations, which can be more easily 'swamped' by gene flow (Antonovics and Levin 1980, p. 429).

Other mechanisms of differentiation have been suggested as acting prior to population separation, and may be able to act equally well in small and moderately large populations. These include processes operating among chromosomes (e.g. chromosomal rearrangements; White 1978; Capanna 1982) and within the DNA itself (e.g. molecular drive; Dover 1982; Rose and Doolittle 1983).

Following from such specific genetic processes is the general suggestion (e.g. Carson and Templeton 1984) that potential for differentiation may depend often on the nature of the genetic material. Although the exact mechanisms are not yet understood, these authors suggest that 'not all genetic conditions appear to have equal potential for speciation mediated through a founder bottleneck,' and that 'those genetic systems that permit exuberant speciation episodes on oceanic islands have certain properties that distinguish them from colonizing and cosmopolitan species' (Carson and Templeton 1984, p. 103). Perhaps it is here that a comment of Bush (1982, p. 128) applies most appropriately: 'Until we know more about the molecular machinery of adaptation . . . our models of speciation must remain little more than speculation based on the subjective interpretation of equivocal data.'

The occurrence and effects of differentiation thus may depend on nature and behavior of the available genetic material, population size, nature of the environment (i.e. selection pressure), timing and degree of isolation, time and chance. This daunting list of possibilities does not mean that, since differentiation can act at any time in any way, speciation is unanalyzable by stages. On the contrary, it makes analysis of speciation by stages all the more important. The same processes of differentiation

can result in very different speciation processes and patterns if they operate within different contexts of isolate formation and persistence (e.g. selection acting before, during, or after isolation). Similarly, virtually identical processes of isolate formation and persistence can produce different results in the presence of different processes of differentiation (e.g. processes leading to small vs. large populations). Given this complexity, explicit analysis and description are critical.

5. POTENTIAL APPLICATIONS

The three-stage framework discussed here can clarify previous analyses of causal factors underlying both taxonomic and morphological diversification. Below I discuss a number of such analyses, grouped under five headings: adaptive radiation, latitudinal diversity gradients, diversity and trophic resources, diversity and disturbance, and macroevolution.

5.1 Adaptive radiation

Despite wide application of the term, both the concept and process of 'adaptive radiation' are poorly understood (e.g. Hardy 1985; Benton 1988; Mitter *et al.* 1988; Eldredge 1989, p. 177ff; Erwin 1991). Two classes of explanation for this evolutionary divergence and diversification are usually advanced (cf. Simpson 1953): 'empty ecospace' and 'key innovation.' Seldom, however, have explicit suggestions been advanced about exactly how either of these actually causes the taxonomic diversification usually associated with adaptive radiation. The framework presented in this paper may help make these potential explanations more explicit, and provide tests for their validity.

Empty ecospace may originate in two ways. (1) Creation of new habitat space, either geologically (e.g. in the case of new volcanic islands) or evolutionarily (e.g. evolution of new prey or food taxa). (2) Formerly occupied habitat may be opened by extinction of predators and/or competitors. To the extent that empty ecospace is the most important causal factor responsible for radiations, the three-stage framework suggests that it does so largely by increasing the probability of persistence of the small, isolated populations that are incipient species. This is true regardless of the diversity or heterogeneity of such habitats. New species are produced by survival of populations in the new environments.

Empty ecospace must also be accessible evolutionarily. A species cannot invade a new region or enter a new 'adaptive zone' (*sensu* Simpson 1953), without being at least to some degree 'preadapted' to do so. Preadaptations in characters that, in retrospect, can be observed as having apparently

limited the adaptive repertoire of a species, are often said to allow the crossing of adaptive discontinuities into new adaptive zones. The concept of a particular morphological, biochemical, or behavioral trait being central to an event of evolutionary diversification was articulated first by Miller (1950), who referred to 'key inventions' (as well as 'key adjustments,' 'key innovations,' and 'key modifications') as permitting 'adaptive radiation within the group on the new (ecologic) plane because of freedom from former ecologic competitors' and tending 'to develop gaps between the new group and the ancestral group lacking the invention because intermediates which may or may not have existed suffer competition from occupants of both old and new ecologic planes.' Bock (1985, p. 129; see also Simpson 1953, p. 190) defines a 'key innovation' as 'a feature, usually assumed to be adaptive, whose existence is essential to the subsequent adaptive radiation' or 'one whose appearance is pivotal for the exploitation of a new habitat and usually central to an adaptive radiation.' (Lauder and Liem 1989 and Baum and Larson 1991 discuss key innovations from the perspective of phenotypic diversity more or less independent of taxic diversity.)

As Bock also points out, however (1985, p. 135), defining a key innovation by its apparent effects on 'adaptive radiation' means that there are no *a priori* rules by which a key innovation can be identified or distinguished from other evolutionary novelties, adaptive or otherwise. Key innovations are recognized only in retrospect and it is therefore difficult to test a hypothesis that a particular innovation is solely or chiefly responsible for an episode of speciation. This problem has been cited by Raikow (1986) as a fatal flaw in the key innovation concept. Cracraft (1990) has, furthermore, pointed out that many supposed key innovations are actually combinations of features that have appeared in higher taxa over extended periods, and so cannot be used to explain anything individually.

The principal arguments for the crucial evolutionary importance of key innovations have been (1) their apparently high contribution to individual relative fitness, and (2) the coincidence of their appearance with times of rapid, divergent, or prolific speciation. Vermeij (1988, p. 70) suggests that 'the hypothesis that a given factor (or constellation of factors) is important is plausible if . . . underlying theory suggests a causal connection.' Yet no explicit mechanistic theory has ever been put forth linking individual fitness to speciation. It is difficult to test the proposition that a key innovation is responsible for an adaptive radiation when it is not specified exactly how it is supposed to have been responsible. Statements that key innovations 'permit,' 'allow,' 'lead to,' or 'cause' adaptive radiation or other abundant speciation are common in the evolutionary literature and in summaries of the history of life (prominent

Table 1

Examples of 'adaptive radiation' attributed to adaptive value of particular 'key innovations' (possibly by increasing rate of isolate persistence)

Taxon	'Key innovation'	Reference
Angiosperms	Reproductive features, etc. ^a	Doyle (1977)
Scleractinian corals	Symbiotic zooxanthellae	Wells (1956)
Siphonate bivalves	Fused mantle siphons	Stanley (1968, 1977)
Pholadacean bivalves	Wood boring	Hoagland and Turner (1981)
Neogastropods	Siphon and proboscis	Taylor <i>et al.</i> (1980)
Balanoid barnacles	Porous walls	Stanley and Newman (1980)
Comatulid crinoids	Loss of stem	Meyer and Macurda (1977)
Echinoids	Bilateral symmetry, etc. ^b	Smith (1984)
Teleost fishes	Feeding mechanisms	Schaeffer and Rosen (1961)
Cichlid fishes	Pharyngeal jaws	Liem (1973)
Bolitoglossine salamanders	Lunglessness	Lombard and Wake (1976, 1977)
Snakes	Loss of mandibular symphysis	Gans (1961)
Passerine birds	Perching feet	Feduccia (1977, 1979, 1980)
Rodents	Evergrowing incisors	Wilson (1951) Wood (1959)
Proboscideans	Hi-fiber processing skull, teeth	Maglio (1973)
Artiodactyls	Forestomach digestion	Simpson (1953) Van Valen (1971) Stanley (1973)
	Astragalus	Schaeffer (1948) Simpson (1959)

^a 'Rapid, efficient reproduction; insect pollination and capacity for symbiotic interaction with the animal world; rapid and flexible vegetative growth and capacity for producing broad, reticulate-veined leaves' (Doyle 1977, p. 539).

^b 'Irregular echinoids are believed to have evolved from tiny opportunistic regular echinoids . . . which . . . made the adaptive breakthrough to living and feeding on loose mobile sediment early in the Lower Jurassic. Once the threshold had been crossed, diversification progressed rapidly as the group expanded into this major new habitat' (Smith 1984, p. 154). Morphological changes leading up to this 'breakthrough' may have included: (a) development of diamond-shaped teeth, (b) test flattening ('an adaptation for living on loose, unconsolidated sediment. The low profile would have provided greater stability in currents . . . and brought a larger proportion of spines in contact with the sea floor for more efficient locomotion'), (c) decrease in size and increase in number of tubercles and spines

examples from various taxonomic groups are given in Table 1), but aside from vague references to such concepts as adaptive zones, it has never been made clear exactly how this is supposed to come about.

Cracraft (1990, p. 36) implies that putative key innovations can only increase rate of speciation by increasing rate at which populations become isolated, the rate at which novel phenotypes enter populations, or the rate at which such novelties become fixed in populations. Viewing speciation as a three-stage process (Fig. 1), however, leads to an alternative hypothesis. A novel characteristic appears in a species that allows the individuals in this species to exploit some aspect of their environment in a way that was not possible for this species before. Since it is advantageous, the characteristic spreads throughout the species through directional selection. Throughout the subsequent 'normal' history of this species, isolated populations form. Since these populations possess this novel feature, they have an advantage relative to populations lacking such features. They are thus more likely to persist and go on to differentiate to become new daughter species.

This hypothesis may have higher testability than many previous suggestions about key innovations since it is unique among them in putting forth a specific hypothesis linking the innovation to the speciation process itself. Given that the hypothesized innovation has been correctly identified at the appropriate level of the population (Cracraft 1990), populations possessing it will be expected to show higher survivorship than those lacking it. A recent polytypic species of allopatric populations or a sufficiently complete fossil record may offer opportunities for such comparisons (see, for example, McKinney *et al.* 1990 for an example of selection in the fossil record).

5.2 Latitudinal diversity gradients

Considering speciation as a three-stage process allows possible explanations for higher species richness in lower latitudes to be reduced to four categories (see Pianka 1966, 1978; Turner *et al.* 1987 for previous attempts): (1) higher rates of differentiation in the tropics, (2) lower extinction rates for established species in the tropics than in the temperate

(d) loss of the lantern, (e) movement of periproct away from the apex and development of anal sulcus for directing feces 'posteriorly away from the aboral respiratory surface and into the animal's wake,' (f) bilateral symmetry to oral tubercles, (g) 'improvement of oral phyllodes (increasing the number of food-gathering tube feet),' (h) 'appearance of bourrelets (increasing the number of manipulative spines adjacent to the mouth)', and (i) 'development of petaloid ambulacra and tube feet highly adapted for gaseous exchange (presumably allowing [the early irregular echinoid] to become a more active burrower)' (Smith 1984, pp. 118–20).

zone, (3) higher rate of isolate formation in the tropics, and (4) higher rate of isolate persistence in the tropics.

Although available data are sparse, there do not appear to be significant differences in the genetic constitution of, or processes acting on, tropical vs. temperate organisms (e.g. Mayr 1969; Powell 1975).

Lower extinction rates of established species in lower latitudes imply higher extinction rates in higher latitudes. To argue for the latter, one seemingly would have to demonstrate either lower levels of adaptation and/or higher rates of environmental disturbance in higher latitudes. There seems to be little evidence for relatively low rates of species extinction in the tropics (although see below); while high extinction rates outside the tropics might apply to those high latitude areas actually subjected to dramatic climate change (e.g. Pleistocene glaciation), it is not clear that middle latitudes have experienced a greater degree of environmental disruption than low latitudes, either in the Pleistocene or throughout geological time. Furthermore, the fossil record suggests that tropical faunas have suffered higher rates of extinction than temperate faunas at least during periods of mass extinction, since they lacked the option of escaping climatic deterioration by moving into lower latitudes (Stanley 1984; Valentine 1984).

Higher tropical diversity could be due in part to higher rates of isolate formation, if at least some of the current species richness in tropical moist forests is due to their fragmentation into refugia in which isolated populations developed into multiple new species (Haffer 1974; Prance 1982; Whitmore and Prance 1987). Even aside from persistent doubts about details of the refugia hypothesis (e.g. Endler 1982; Connor 1986; Colinvaux 1987; Gentry 1989, and references therein), however, this suggestion must confront the apparently high rate of refuge formation in at least some temperate regions during the Pleistocene (e.g. Fernald 1925; Lindroth 1969; Brown 1971, 1978; Axelrod 1981; Tchernov 1982; Wells 1983). If both tropical and temperate environments were subjected to fragmentation, why are there now more species in the tropics?

That the answer may be a higher rate of isolate persistence in the tropics has been suggested many times in the past, but usually indirectly. Many authors have argued, for example, that more species live in tropical habitats because more species can; the problem of tropical diversity is therefore seen as a problem of the maintenance rather than the origin of diversity (e.g. Connell and Orias 1964; Mayr 1969; Williams 1972; Grubb 1977; Haffer 1977; Janzen 1977; Sale 1977; Whittaker 1977; Huston 1979; Hubbell and Foster 1986; Stenseth 1984; Shmida and Wilson 1985; Gentry 1989; Dorit 1991). Janzen (1977, p. 85) goes so far as to state that 'the number of species in a contemporary habitat has little or nothing to do with speciation rates there or elsewhere.'

Some or all of the factors potentially contributing to maintenance of

higher diversity in tropical communities could do so by lowering rates of extinction of established species. It is primarily for this reason that some authors have suggested that, whether tropical species originate in the tropics or elsewhere, high species diversity is due largely to their simple accumulation in tropical habitats (e.g. Stebbins 1974; Stenseth 1984). As already mentioned, however, these same factors could also lower extinction rates (i.e. raise persistence rates) of incipient species. To the extent that this occurs, the many observations and theories concerning higher probabilities of maintenance of diversity in tropical habitats provide circumstantial support for a higher rate of isolate persistence as the principal explanation for high tropical species richness. Rates of species origin thus may be higher in the tropics because isolated populations have higher probability of survival in the less climatically harsh and variable tropical environments (Dobzhansky 1950; Janzen 1967; Burger 1981; Glazier 1987a).

Recently, Stevens (1989) has put forward a somewhat different hypothesis. He suggests that, as a result of species' narrower geographic ranges in lower latitudes, 'the dispersal powers of individuals near the edge of their preferred microhabitat may extend to unfavorable areas.' This might lead to 'the arrival of individuals in areas where they are able to survive but unable to maintain their population' (1989, p. 251). It is these constantly colonizing species, Stevens argues, that are responsible for high alpha diversities in tropical communities. As with most other hypotheses of tropical diversity, this is essentially a maintenance argument, but with a twist. If Stevens is correct, then the narrower geographic ranges of low-latitude species may lead to a mechanism for abundant species origin by leading to higher rates of isolate formation. High tropical diversity may thus be explicable partly by changes in rate of isolate formation after all, but not as a result of large-scale fragmentation of habitat resulting in refugia.

5.3 Diversity and trophic resources

Connections have often been postulated between trophic resources and diversity of particular taxa, body sizes or trophic levels, in particular communities, or at particular times in earth history. As in other cases discussed above, relationships between resources and diversity have usually been in the context of maintenance of diversity. Yet even when the origin of diversity has been discussed, the precise mechanistic nature of these connections usually has not been articulated. Brown (1975, p. 320; see also Stanley 1979, p. 278) has pointed out that '[t]he production and availability of resources affect species diversity chiefly by limiting the sizes of populations that a habitat can support.' Within the context of the three-stage framework discussed here, the hypothesis therefore can

be put forth that increased levels of trophic resources may lead to increased rates of speciation by increasing rates of isolate persistence. Variable or fluctuating trophic resources may increase speciation rates by leading to range fragmentation and increased rate of isolate formation. If resources are too unstable, speciation may be damped by a decreased rate of isolate persistence.

The higher number of species of small body size is suggestive of a connection between resources and taxonomic diversity (Van Valen 1973; Damuth 1981; Peters and Wassenberg 1983; Peters and Raelson 1984; Brown and Maurer 1987; Kochmer and Wagner 1988). Smaller animals consume less (i.e. have relatively more food available), and therefore their populations can be larger and so more persistent (Gaston and Lawton 1988a,b; Lawton 1989; but see Dixon and Kindlmann 1990). The similarity between pattern of resource availability at different trophic levels (the familiar 'trophic pyramid'; e.g. Odum 1971) and pattern of species diversity at those trophic levels may indicate a similar relationship. For living birds, living reptiles, living mammals, and dinosaurs, for example, more species are present at lower than at higher trophic levels, or in trophic groups that feed on particularly abundant food resources (e.g. insects) (Tables 2-5). In a similar analysis, Glazier (1987b) compared number of species per genus in classes of terrestrial organisms with more than 1000 species. He found that there are more than twice as many species per genus of producers (bryophytes, pteridophytes, angiosperms) than small ectothermic consumers (terrestrial arthropods, including insects), roughly twice as many small as large ectothermic consumers

Table 2

Species richness of living reptiles, classified by approximate trophic group. Group designations reflect presumed relative trophic discreteness and/or homogeneity even if they do not designate trophic level. Aquatic species, for example, probably overlap very little in their diet with terrestrial carnivores or herbivores, even if they do not all eat the same things or at the same trophic level (data from Dowling and Duellman 1978)

Trophic group	No. of species	Total percentage
Fossilorial	456	8.2
Terrestrial carnivorous	1784	32.2
Aquatic	396	7.1
Marine	52	0.9
Arboreal/terrestrial insectivorous	2794	50.4
Herbivorous	39	0.7
Crocodilians	21	0.3

Table 3

Species richness of living birds, classified by approximate trophic group (species numbers from Wallace and Mahan 1975; trophic data largely from Austin and Singer 1971)

Trophic group	No. of species	Total percentage
Piscivores	284	3.3
Herbivores (granivores + frugivores)	1978	23.0
Carnivores	559	6.5
Insectivores	2752	32.0
Omnivores	2494	29.0
Insect-nectivores	473	5.5
Insect-carnivores	26	0.3
Undetermined	35	0.4
	8601	100.0

Table 4

Species richness of the living orders of mammals, with trophic classification. I = insectivorous, C = carnivorous, H = herbivorous, O = omnivorous (data from Vaughan 1978)

Trophic group	No. of species	Total percentage
Insectivora	I	406
Dermoptera	I	2
Chiroptera	I/H	853
Primates	H/O	166
Edentata	H/I	31
Pholidota	I	8
Lagomorpha	H	63
Rodentia	H	1690
Cetacea	H/C	84
Carnivora	C	284
Tubulidentata	I	1
Proboscidea	H	2
Hyracoidea	H	11
Sirenia	H	5
Perissodactyla	H	16
Artiodactyla	H	171
	3793	99.8

Table 5

Number of genera in the major taxa of dinosaurs (data from Lambert 1983) and number of genera of carnivores and herbivores

Infraclass Archosauria	
Order Saurischia	
Suborder Theropoda	
Infraorder Coelurosauria	33
Ornithomimosauria	9
Deinocheirosauria	1
Deinonychosauria	11
Carnosauria	43
Segnosauria	3
?Therizinosauria	1
incertae sedis	15
Suborder Sauropodomorpha	
Infraorder Prosauropoda	20
Sauropoda	62
Order Ornithischia	
Suborder Ornithopoda	80
Stegosaura	11
Ankylosauria	33
Ceratopsia	21
Total	343
Total herbivores	217 (63%)
Total carnivores	126 (37%)

(amphibians and reptiles), and about twice as many of the latter as large endothermic consumers (birds and mammals).

More detailed studies within single clades or communities yield similar results. Mitter *et al.* (1988), for example, show that in pair-wise comparisons of sister groups of insects, phytophagous clades are consistently more diverse. These authors suggest that phytophagy could increase diversification rate in these clades because '[t]he resource base available to primary consumers is larger than that available to higher trophic levels, and this could enhance diversity by itself' (1988, p. 115). Exactly how is left unstated, yet such a pattern is well explained by postulating a higher rate of isolate persistence in phytophagous clades.

Turner *et al.* (1987) have reported that butterfly diversity in Great Britain is highly correlated with availability of solar energy, and argue that this supports the 'species-energy hypothesis' of Wright (1983), which attempts to explain latitudinal diversity gradients by the gradient of available solar energy. Although it is not the only possible causal interpretation of this correlation, the hypothesis of greater energy resulting

in greater survival of individuals and so higher rates of persistence of isolated populations is consistent with it.

Connections between trophic resources and diversification are also suggested by patterns over geological time. The history of perissodactyl and artiodactyl ungulates often has been viewed as an example of 'evolutionary replacement' of a poorly adapted group (perissodactyls) by a better adapted group (artiodactyls) (e.g. Simpson 1953; Van Valen 1971; Stanley 1973). Janis (1989) has shown, however, that the Cenozoic history of the two groups was more complex than this. Perissodactyls are hindgut fermenters, and are better able to make use of large quantities of very fibrous plant material, such as is available in tropical forests (Janis 1989, p. 476). Artiodactyls are mostly ruminants, that is foregut fermenters, which are more efficient in seasonal habitats. As documented by Janis and by Langer (1987), foregut fermenting artiodactyls expanded dramatically in Africa, Asia, North America, and Europe in the Miocene and Pliocene, when climate changes replaced aseasonal forest-dominated biomes with more seasonal, grassland-dominated environments. One of the most dramatic of all mammalian radiations, that of the artiodactyl family Bovidae, occurred coincident with particularly severe climatic changes (Vrba 1984). These patterns suggest a role for the formation of isolated populations during climate changes, but they are also consistent with enhanced isolate survival among artiodactyls owing to their ability to use the available trophic resources more efficiently.

At larger scales, Bambach (1977) has suggested that increases in marine diversity in the Silurian and Cretaceous periods may have been due in part to increases in nutrient influxes from the land, which in turn were due to changes in terrestrial plant communities (the development of land plants and angiosperms, respectively). Lipps and Mitchell (1976) have pointed out the coincidence of the diversification of marine mammals with development of nutrient-supplying upwelling systems in the Miocene (see, for example, Cook and McElhinny 1979; Haq 1981). Caron and Homewood (1983), Corfield and Shackleton (1988) and Leckie (1989) similarly link nutrient availability into evolutionary radiations of planktonic Foraminifera.

Hallock (1987) has proposed a more general theory connecting nutrient availability in the marine realm to processes that could lead to both isolate formation and persistence. She suggests that periods of low nutrient availability will be more stable, allowing greater specialization and diversity of niches. Populations will be small, possibly promoting diversification further. In addition, oceanographic conditions that are responsible for very oligotrophic conditions will also, Hallock suggests, locally create highly eutrophic conditions such as upwelling zones, thus expanding the total 'trophic resource continuum' and allowing for still further niche diversification.

5.4 Diversity and disturbance

Postulated relationships between disturbance and diversity have almost all been in the context of mechanisms of maintenance of diversity. The 'intermediate disturbance hypothesis' (e.g. Janzen 1970; Connell 1978; Hubbell and Foster 1986; Petraitis *et al.* 1989), for example, suggests that diversity of a community will be highest at intermediate levels of disturbance. Yet, as in previous examples, it is possible within the context of the three-stage framework to include disturbance in a theory of speciation as well. Not entirely coincidentally, it is at intermediate levels of disturbance, according to this view, that speciation is most likely (Fig. 2).

The environment provides 'opportunities' for allopatric speciation when it either causes vicariance events, or sets the physical stage for chance dispersal (e.g. Cracraft 1982, 1985; Cronin and Ikeya 1991). These opportunities are often created at times of environmental change or deterioration, usually when formerly continuous habitats become fragmented. At such times, all other factors being equal, the probability of isolate formation increases, and speciation rate may rise. If environmental change continues, however, habitats may be fragmented to such a degree that the resulting isolates are too small to survive. Even more likely, the

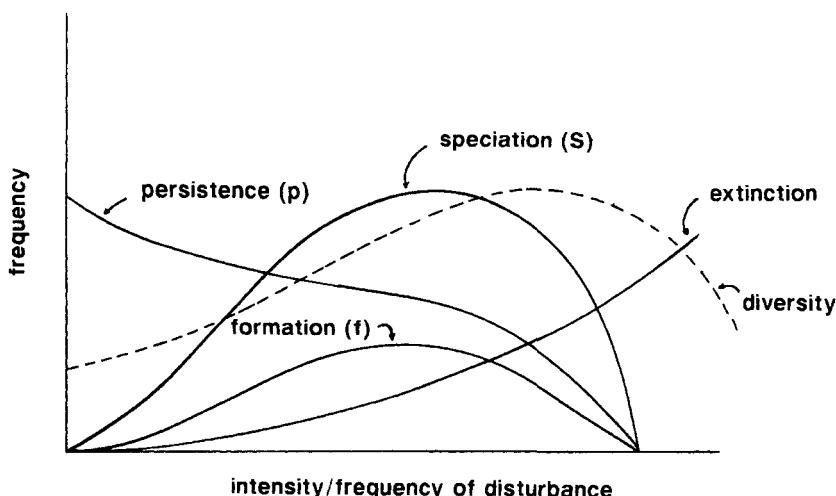


Fig. 2. Model for an 'intermediate disturbance hypothesis' of the origin (as opposed to the maintenance) of diversity. Rate of successful speciation (*S*) is highest at intermediate levels of disturbance since here rates of both isolate formation (*f*) and persistence (*p*) are high. See text for further discussion.

environmental stresses causing fragmentation will begin to affect adversely the persistence probabilities of all isolates, and rate of successful speciation will decline.

These ideas are not entirely new, but have never before been placed in the context of a general theory of speciation. Simpson (1944, p. 89) argued that occurrence of prolific speciation in a group might result from 'a decline in the adaptive status of the ancestral populations, and consequent centrifugal selection and fragmentation of groups imperfectly adapted but tending more or less toward a variety of different adaptive types . . .' Lewis (1962) suggested that speciation in the plant genus *Clarkia* has been more likely during times of adverse environmental conditions (drought) than during more favorable conditions (more rainfall) because harsh conditions can eliminate entire populations by 'catastrophic selection,' leaving only a few survivors possessing resistant traits. Processes of differentiation must also be involved in this case, Lewis argued, since such dramatic population contractions can occur without formation of new species. 'For speciation to occur,' wrote Lewis, 'the result of selection must be essentially irreversible' (1962, p. 268).

Stanley (1979) has suggested that environmental 'adversity would manifest itself most readily in suppressing speciation, rather than in accelerating the extinction of established species [since] . . . small populations involved in the speciation process should be more susceptible to extirpation or to restraint from expansion than a full-fledged species should be vulnerable to total extinction' (1979, pp. 197–8). He clarified the idea in 1986: 'high rates of speciation are actually promoted by less severe environmental deterioration – deterioration severe enough to elevate extinction rates to a moderate degree but not so severe as to cause wholesale extinction' (1986, p. 104). Stanley has designated as the 'fission effect' 'the general phenomenon in which hostile environmental conditions on a broad geographic scale accelerate speciation, while also causing extinction' (Stanley 1986, p. 105). By analogy with the intermediate disturbance hypothesis of diversity maintenance, it can be added to this that at sufficiently low levels of disturbance, isolate formation may well be higher than extinction rates of both isolates and established species, and that net diversity may increase. (Fig. 2).

5.5 Macroevolution

A 'hierarchical view' of macroevolution (e.g. Stanley 1979; Eldredge 1979, 1982; Vrba 1980, 1983a,b, 1985; Gould 1982, 1985; Gilinsky 1986; Vrba and Eldredge 1984) maintains that long-term evolutionary trends are due not to anagenetic change within species lineages but to sorting

among species generated randomly with respect to the morphological direction of the trend. This sorting may, in turn, be due to differential speciation or extinction probabilities in different branches of a clade. Three categories of cause for speciation rate have been considered within this view. (1) Speciation is due to properties of species themselves not reducible to properties of constituent individuals (true species selection; Vrba 1983b; Gilinsky 1986). (2) Speciation is due to extrinsic factors dividing up populations (e.g. Cracraft 1982, 1985). (3) Speciation is due to properties of individual organisms, but acting with incidental effect at the level of the species (the effect hypothesis; Vrba 1983a, 1985, 1987). All three of these causes exclude the traditional role(s) of adaptation (i.e. at the individual level within populations) in macroevolution, and in their most extreme formulations deny any significant role for individual adaptation in macroevolution.

More recently, several authors have argued that, while the traditional Darwinian view may be an oversimplification, adaptation can play an important role in explaining macroevolutionary patterns (e.g. Jackson 1988; Jackson and McKinney 1991; Vermeij 1987a). Implicit in these reassessments, however, is a remaining tension between phenotypic and taxic diversity: both adaptation and speciation are important, but how are they related? Jackson (1988, p. 311) summarizes the problem: 'Because we do not understand how species arise, we cannot say whether ecological processes help to mold their character and origin, or merely sort the species afterwards like so many randomly generated genes . . .'

By providing a conceptual link between adaptation and speciation, the three-stage framework discussed here may help in resolving this impasse. Adaptation may be important in macroevolution, but not only in the traditional reductionist mode within populations. Speciation may, at least occasionally, be due to adaptation, not just as an accidental side-consequence, but as a direct result. Adaptation may increase the rate of isolate persistence, and so successful speciation. The degree to which this applies will be determined by the relative importance of isolate formation and isolate persistence, possible variations of which are illustrated in Fig. 3.

If, as a simplifying assumption, probability of differentiation by all mechanisms is treated as constant, speciation can be seen (and analyzed) as a net result of two varying functions: isolate formation and persistence. When one of these variables is constant, or at least not limiting, the outcome is controlled by the other. If many more isolates are generated than ever go on to become successful species, then rate of speciation is controlled largely by rate of isolate persistence. If, on the other hand, isolate formation is relatively infrequent, and/or most isolates that do form have a fair chance of persisting, then rate of speciation is controlled largely by rate of isolate formation.

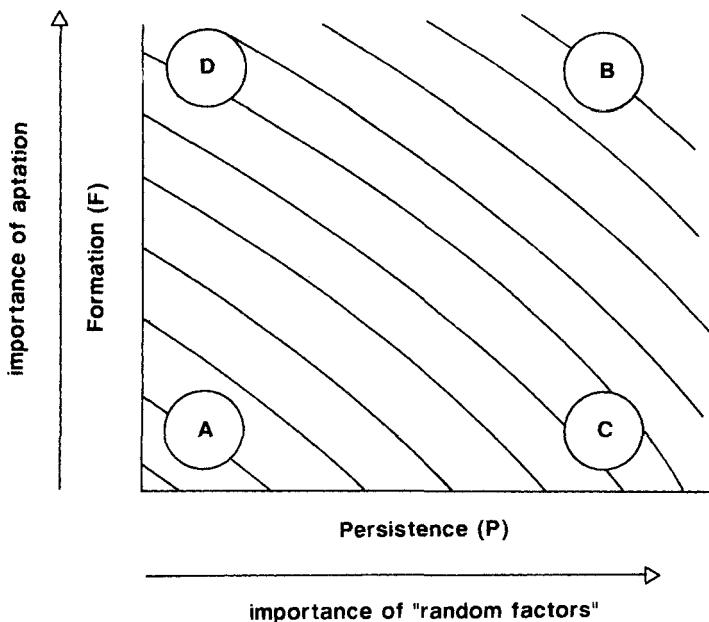


Fig. 3. Graphical representation of possible macroevolutionary significance of relative importance of isolate formation and persistence, assuming probability of isolate differentiation is constant. (A) Isolate formation and persistence are both low, and so is net successful speciation. (B) Isolate formation and persistence are both high, a condition referred to as 'supertaxa' by Stanley (1979). (C) Rate of isolate persistence is high (and so not limiting), and so isolate formation is the most important factor controlling successful speciation. 'Random' factors refer to environmental changes affecting isolate formation but not otherwise affecting individual organisms. (D) Isolate formation is high and not limiting, and so isolate persistence is the most important factor controlling successful speciation. In such a case, individual adaptation may potentially play an important role in the speciation process and so in macroevolution.

6. DISCUSSION

The usefulness of considering speciation as consisting of three stages or components lies in the degree to which it encourages development of more explicit questions or tests concerning what factors are most important in affecting speciation in individual cases. I have outlined above a number of general cases in which I think this stepwise approach can serve this role. More detailed analyses may be permitted in a quantitative statement of the framework.

In discussing rates of recent extinction on islands, Diamond (1984a) has shown that the percentage of species in a group endemic to an

archipelago varies inversely with the average extinction rate of these species in that archipelago. The fraction of an archipelago's species that survive for time t (i.e. the time necessary to evolve into a distinct species) can thus be expressed as e^{-qt} , where q is the probability per unit time that a species present in an archipelago will go extinct.

A preliminary approach to quantifying the three-stage framework discussed here can be derived from a modification of Diamond's expression. The extinction probability of isolated populations can be expressed as

$$S = \frac{Nf}{f + q}$$

where S = number of successful speciation events, N = number of existing (potentially parental) species, f = probability per unit time of isolate formation, and q = probability of isolate extinction per unit time. If $p = 1 - q$ = probability of persistence, speciation can be expressed as a function of both isolate formation and persistence. Putting actual values on these variables is a very difficult task, made even more difficult by the fact that many factors may contribute to both formation and persistence, often in opposite directions. Yet their explicit specification in this manner can contribute to this process by forcing explication of all possibilities.

One of the most significant stumbling blocks to a better understanding of speciation is the multiplicity of effects that a single biological phenomenon can have on different aspects of the speciation. The three-stage framework discussed here may contribute to analysis of such cases. Pomeroy (1990), for example, has recently suggested that mortality rates are lower in flying than in non-flying endotherms, chiefly because of reduced predation. Viewing speciation as a three-stage process raises the possibility that flight has had a positive effect on speciation either by affecting rate of isolate formation in flying birds and mammals, or by increasing the rate of persistence of those isolates. Other examples of multiple potential effects of single factors have been discussed above, including population size, body size, and adaptation. The three-stage framework makes the potential effects of such factors explicit, and forces us to confront and attempt to analyze them.

Finally, it may be noted that the three-stage framework may contribute to studies of speciation in the fossil record, since it makes more explicit the roles of particular categories of factors, many of which are identifiable in paleontological data. Factors potentially increasing the rate of isolate formation, such as environmental changes or predation intensity, can often be identified on geological evidence (e.g. Cracraft 1985; Stanley 1986; Cronin and Ikeya 1991). Dispersal

ability, also potentially affecting isolate formation, may be recognizable in fossils (e.g. Jablonski and Lutz 1983; Taylor 1989). Particular morphological novelties recognizable in fossils may be identified as adaptive, or at least as functional (e.g. Hickman 1988), and so may have affected rates of isolate persistence (e.g. Vermeij 1987a). In individual cases, one or more of these variables may appear more important, leading to an explicit hypothesis about causal factors underlying speciation (e.g. McNamara 1982; Allmon 1988).

7. CONCLUSIONS

(1) Speciation can be viewed as a result of a three-stage process involving (a) formation of an isolated daughter population, (b) persistence or survival of that population, and (c) differentiation of that population to form a distinct new daughter species.

(2) Isolate formation may be caused by a variety of mechanisms. Extrinsic mechanisms include environmental change, predation, and creation of habitat patches. Intrinsic mechanisms include dispersal ability, population structure, and ecological amplitude.

(3) Isolate persistence may be affected by a number of factors including population size and stability, ecological amplitude, environmental rigor, resources, environmental grain, and adaptation.

(4) Isolate differentiation is affected by many factors including population size, selection intensity, nature of the genome, timing and degree of isolation, time, and chance. Different timing, processes, or magnitude of differentiation can yield very different patterns and processes of speciation, even given identical patterns and processes of isolate formation and persistence. Similarly, the same processes of differentiation can result in very different speciation processes and patterns if they operate within different contexts of isolate formation and persistence.

(5) This 'formation–persistence–differentiation framework' for viewing speciation formalizes and makes explicit implicit ideas and suggestions made by many previous authors. As a comprehensive way of analyzing speciation it may be applied to a number of outstanding general evolutionary problems. By making the steps in the speciation process, and the factors potentially affecting them, more explicit and precise, the approach outlined here may encourage development of more explicit and precise predictions and tests on exactly what causes individual episodes of speciation to occur.

(6) This three-stage framework offers an explicit device for identifying possible relationships between adaptation and speciation in macroevolution through consideration of factors affecting formation, persistence, and differentiation of isolated incipient species, and which of these steps is

most likely to be limiting. Considering probability of differentiation to be constant, when isolate persistence is likely and isolate formation is limiting, adaptation may be less important in generating macroevolutionary patterns. When isolate formation is likely and persistence is limiting, adaptation has the potential to affect the occurrence of speciation, and so contribute directly to macroevolutionary patterns, even if these patterns are generated largely by rates of speciation.

ACKNOWLEDGMENTS

I am grateful to R. L. Dorit, D. H. Geary, S. J. Gould, G. C. Mayer, R. J. O'Hara, R. M. Ross, and E. E. Williams for discussions during the development of the ideas presented here. I thank J. Antonovics, M. Foote, D. Futuyma, D. Glazier, P. Morris, and J. Sepkoski for helpful comments on previous drafts of the manuscript. This work was supported in part by a National Science Foundation Graduate Fellowship.

REFERENCES

- Allmon, W. D. (1988). Patterns of first appearance and causes of speciation in Lower Tertiary mollusks. *Geol. Soc. Am. Abst.w/Prog.* **20**, A104.
- Anderson, P. K. (1970). Ecological structure and gene flow in small mammals. *Symp. Zool. Soc. Lond.* **26**, 299–325.
- Antonovics, J. (1976). The input from population genetics: 'The new ecological genetics.' *Ann. Missouri Bot. Garden* **63**, 224–47.
- and Levin, D. A. (1980). The ecological and genetic consequences of density-dependent regulation in plants. *Ann. Rev. Ecol. Syst.* **11**, 411–52.
- Austin, O. L., Jr and Singer, A. (1971). *Families of birds*. Golden Press, New York.
- Avise, J. C. (1978). Genetic differentiation during speciation. in *Molecular evolution*, (ed. F. J. Ayala), pp. 106–22. Sinauer Associates, Sunderland, MA.
- Axelrod, D. I. (1981). Holocene climatic changes in relation to vegetation disjunction and speciation. *Am. Nat.* **117**, 847–70.
- Ayala, F. J. (1975). Genetic differentiation during the speciation process. *Evol. Biol.* **8**, 1–78.
- Baker, A. J., Dennison, M. D., Lynch, A. and LeGrand, G. (1990). Genetic divergence in peripherally isolated populations of chaffinches in the Atlantic Islands. *Evolution* **44**, 981–999.
- Bambach, R. K. (1977). Species richness in marine benthic habitats through the Phanerozoic. *Paleobiology* **3**, 152–67.
- Barton, N. H. and Charlesworth, B. (1984). Genetic revolutions, founder effects, and speciation. *Ann. Rev. Ecol. Syst.* **15**, 133–64.
- Baum, D. A. and Larson, A. (1991). Adaptation reviewed: a phylogenetic methodology for studying character evolution. *Syst. Zool.* **40**, 1–18.

- Belovsky, G. E. (1987). Extinction models and mammalian persistence. In *Viable populations for conservation* (ed. M. E. Soulé), pp. 35–58. Cambridge University Press, Cambridge.
- Benton, M. J. (1988). The nature of an adaptive radiation. *Trends Ecol. Evol.* **3**, 127–8.
- Bleiweiss, R. (1991). Ecological causes of clade diversity in hummingbirds: a neontological perspective on the generation of diversity. In *Causes of evolution. A paleontological perspective* (ed. R. M. Ross and W. D. Allmon), pp. 354–80. University of Chicago Press, Chicago.
- Bock, W. J. (1985). Adaptive inference and museological research. In *Museum collections: their roles and future in biological research*. Occas. Pap. British Columbia Prov. Mus. No. 25, pp. 123–38.
- Brown, J. H. (1971). Mammals on mountaintops: nonequilibrium insular biogeography. *Am. Nat.* **105**, 467–78.
- (1975). Geographical ecology of desert rodents. In *Ecology and evolution of communities* (ed. M. L. Cody and J. M. Diamond), pp. 315–41. Harvard University Press, Cambridge, MA.
- (1978). The theory of insular biogeography and the distribution of boreal birds and mammals. In *Intermountain biogeography: a symposium*. Great Basin Naturalist Memoir 2, pp. 209–27.
- Brown, J. H. and Maurer, B. A. (1987). Evolution of species assemblages: effects of energetic constraints and species dynamics on the diversification of the North American avifauna. *Am. Nat.* **130**, 1–17.
- Bryant, E. H. and Meffert, L. M. (1990). Multivariate phenotypic differentiation among bottleneck lines of the housefly. *Evolution* **44**, 660–8.
- Burger, W. C. (1981). Why are there so many kinds of flowering plants? *BioScience* **31**, 572–81.
- Bush, G. L. (1975). Modes of animal speciation. *Ann. Rev. Ecol. Syst.* **6**, 339–64.
- (1981). Stasipatric speciation and rapid evolution in animals. In *Evolution and speciation* (ed. W. R. Atchley and D. S. Woodruff), pp. 201–18. Cambridge University Press, Cambridge.
- (1982). What do we really know about speciation? In *Perspectives on evolution* (ed. R. Milkman), pp. 119–28. Sinauer Associates, Sunderland, MA.
- Butlin, R. (1989). Reinforcement of premating isolation. In *Speciation and its consequences* (ed. D. Otte and J. A. Endler), pp. 158–79. Sinauer Associates, Sunderland, MA.
- Cain, A. J. and Cook, L. M. (1989). Persistence and extinction in some *Cepaea* populations. *Biol. J. Linn. Soc.* **38**, 183–90.
- Capanna, E. (1982). Robertsonian numerical variation in animal speciation: *Mus musculus*, an emblematic model. In *Mechanisms of speciation* (ed. C. Barigozzi), pp. 155–78. Alan R. Liss, New York.
- Caron, M. and Homewood, P. (1983). Evolution of early planktic foraminifers. *Mar. Micropaleo.* **7**, 453–62.
- Carson, H. L. and Templeton, A. R. (1984). Genetic revolutions in relation to speciation phenomena: the founding of new populations. *Ann. Rev. Ecol. Syst.* **15**, 97–131.
- Colinvaux, P. (1978). *Why big fierce animals are rare: an ecologist's perspective*. Princeton University Press, Princeton.

- (1987). Amazon diversity in light of the paleoecological record. *Quat. Sci. Rev.* **6**, 93–114.
- Connell, J. H. (1978). Diversity in tropical rain forests and coral reefs. *Science* **199**, 1302–10.
- and Orias, E. (1964). The ecological regulation of species diversity. *Am. Nat.* **98**, 399–414.
- Connor, E. F. (1986). The role of Pleistocene forest refugia in the evolution and biogeography of tropical biotas. *Trends Ecol. Evol.* **1**, 165–8.
- Cook, P. J. and McElhinny, M. W. (1979). A reevaluation of the spatial and temporal distribution of sedimentary phosphate deposits in the light of plate tectonics. *Econ. Geol.* **74**, 315–30.
- Corfield, R. M. and Shackleton, N. J. (1988). Productivity change as a control on planktonic foraminiferal evolution after the Cretaceous/Tertiary boundary. *Hist. Biol.* **1**, 323–43.
- Cracraft, J. (1982). A nonequilibrium theory for the rate-control of speciation and extinction and the origin of macroevolutionary patterns. *Syst. Zool.* **31**, 348–65.
- (1985). Biological diversification and its causes. *Ann. Missouri Bot. Gard.* **72**, 794–822.
- (1990). The origin of evolutionary novelties: pattern and process at different hierarchical levels. In *Evolutionary innovations* (ed. M. H. Nitecki), pp. 21–46. University of Chicago Press, Chicago.
- Cronin, T. M. (1985). Speciation and stasis in marine Ostracoda: climate modulation of evolution. *Science* **227**, 60–3.
- (1987). Speciation and cyclic climatic change. In *Climate. History, periodicity and predictability* (ed. M. R. Rampino, J. E. Sanders, W. S. Newman and L. K. Konigsson), pp. 333–42. Van Nostrand Reinhold Company, New York.
- and Ikeya, N. (1991). Tectonic events and climatic change: opportunities for speciation in Cenozoic marine Ostracoda. In *Causes of evolution. A paleontological perspective* (ed. R. M. Ross and W. D. Allmon), pp. 210–48. University of Chicago Press, Chicago.
- Damuth, J. (1981). Population density and body size in mammals. *Nature* **290**, 699–700.
- (1985). Selection among ‘species’: a formulation in terms of natural functional units. *Evolution* **39**, 1132–46.
- den Boer, P. J. (1971). Stabilization of animal numbers and the heterogeneity of the environment: the problem of the persistence of sparse populations. In *Dynamics of populations. Proceedings of the Advanced Study Institute on ‘Dynamics of Numbers in Populations,’ Oosterbeek, Netherlands* (ed. P. J. den Boer and G. R. Gradwell), pp. 77–97. Centre for Agricultural Publ. and Documentation, Wageningen.
- (1985). Fluctuations of density and survival of carabid populations. *Oecologia* **67**, 322–30.
- Diamond, J. M. (1984a). Historic extinctions: a Rosetta stone for understanding prehistoric extinctions. In *Quaternary extinctions: a prehistoric revolution* (ed. P. S. Martin and R. G. Klein), pp. 824–62. University of Arizona Press, Tucson.

- (1984b). 'Normal' extinctions in isolated populations. In *Extinctions* (ed. M. Nitecki), pp. 191–246. University of Chicago Press, Chicago.
- Dixon, A. F. G. and Kindlmann, P. (1990). Role of plant abundance in determining the abundance of herbivorous insects. *Oecologia* **83**, 281–3.
- Dobzhansky, T. (1937). *Genetics and the origin of species*. Columbia University Press, New York.
- (1940). Speciation as a stage in evolutionary divergence, *Am. Nat.* **74**, 312–21.
- (1950). Evolution in the tropics. *Am. Sci.* **38**, 209–21.
- (1976). Organismic and molecular aspects of species formation. In *Molecular evolution*. (ed. F. J. Ayala), pp. 95–105. Sinauer Associates, Sunderland, MA.
- Dorit, R. L. (1991). The correlates of high diversity in Lake Victoria haplochromine cichlids: a neontological perspective. In *Causes of evolution. A paleontological perspective* (ed. R. M. Ross and W. D. Allmon), pp. 322–53. University of Chicago Press, Chicago.
- Dover, G. A. (1982). A role for the genome in the origin of species. In *Mechanisms of speciation* (ed. C. Barigozzi), pp. 435–60. Alan R. Liss, New York.
- Dowling, H. G. and Duellman, W. E. (1978). *Systematic herpetology: a synopsis of families and higher categories*. HISS Publications, New York.
- Doyle, J. A. (1977). Patterns of evolution in early angiosperms. In *Patterns of evolution, as illustrated in the fossil record* (ed. A. Hallam), pp. 501–46. Elsevier, Amsterdam.
- Ebenhard, T. (1991). Colonization in metapopulations: a review of theory and observations. *Biol. Jour. Linn. Soc.* **42**, 105–21.
- Ehrlich, P. R. (1986). Which animals will invade? In *Ecology of biological invasions of North America and Hawaii* (ed. H. A. Mooney and J. A. Drake), pp. 79–95. Springer-Verlag, New York.
- and Raven, P. (1969). Differentiation of populations. *Science* **165**, 1228–32.
- Eldredge, N. (1979). Alternative approaches to evolutionary theory. *Bull. Carnegie Mus. Nat. Hist.* **13**, 7–19.
- (1982). Phenomenological levels and evolutionary rates. *Syst. Zool.* **31**, 338–47.
- (1989). *Macroevolutionary dynamics. Species, niches and adaptive peaks*. McGraw-Hill, New York.
- and Cracraft, J. (1980). *Phylogenetic patterns and the evolutionary process*. Columbia University Press, New York.
- Endler, J. A. (1982). Pleistocene forest refuges: fact or fancy? In *Biological diversification in the tropics*. (ed. G. T. Prance), pp. 641–72. Columbia University Press, New York.
- (1986). *Natural selection in the wild*. Princeton University Press, Princeton, NJ.
- Erwin, D. H. (1992). A preliminary classification of evolutionary radiations. *Hist. Biol.* (in press).
- Feduccia, A. (1977). A model for the evolution of perching birds. *Syst. Zool.* **26**, 19–31.
- (1979). Comments on the phylogeny of perching birds. *Proc. Biol. Soc. Wash.* **92**, 689–96.

- (1980). *The age of birds*. Harvard University Press, Cambridge, MA.
- Fernald, M. L. (1925). Persistence of plants in unglaciated areas of boreal America. *Mem. Am. Acad. Arts Sci.* **15**, 239–342.
- Fryer, G. and Iles, T. D. (1972). *The cichlid fishes of the Great Lakes of Africa*. Oliver and Boyd, Edinburgh.
- Gans, C. (1961). The feeding mechanism of snakes and its possible evolution. *Amer. Zool.* **1**, 217–27.
- Gaston, K. J. and Lawton, J. N. (1988a). Patterns in the distribution and abundance of insect populations. *Nature* **331**, 709–12.
- (1988b). Patterns in body size, population dynamics, and regional distribution of bracken herbivores. *Am. Nat.* **132**, 662–80.
- Gentry, A. H. (1989). Speciation in tropical forests. In *Tropical forests. Botanical dynamics, speciation, and extinction* (ed. L. B. Holm-Nielsen, I. C. Nielsen and H. Balslev), pp. 113–34. Academic Press, London.
- Gilinsky, N. L. (1986). Species selection as a causal process. *Evol. Biol.* **20**, 249–73.
- Gilpin, M. E. and Soule, M. E. (1986). Minimum viable populations: processes of species extinction. In *Conservation biology. The science of scarcity and diversity* (ed. M. E. Soulé), pp. 19–34. Sinauer Associates, Sunderland, MA.
- Glazier, D. S. (1980). Ecological shifts and the evolution of geographically restricted species of North American *Peromyscus* (mice). *J. Biogeogr.* **7**, 63–83.
- (1986). Temporal variability of abundance and the distribution of species. *Oikos* **47**, 309–14.
- (1987a). Toward a predictive theory of speciation: the ecology of isolate selection. *J. Theor. Biol.* **126**, 323–33.
- (1987b). Energetics and taxonomic patterns of species diversity. *Syst. Zool.* **36**, 62–71.
- Goodnight, C. J. (1987). On the effect of founder events on epistatic genetic variance. *Evolution* **41**, 80–91.
- Gould, S. J. (1982). The meaning of punctuated equilibrium and its role in validating a hierarchical approach to macroevolution. In *Perspectives on evolution* (ed. R. Milkman), pp. 83–104. Sinauer Associates, Sunderland, MA.
- (1985). The paradox of the first tier: an agenda for paleobiology. *Paleobiology* **11**, 2–12.
- and Vrba, E. S. (1982). Exaptation – a missing term in the science of form. *Paleobiology* **8**, 4–15.
- Grubb, P. J. (1977). The maintenance of species richness in plant communities: the importance of the regeneration niche. *Biol. Rev.* **52**, 107–45.
- Haffer, J. (1974). *Avian speciation in tropical South America*, Publication 14. Nuttall Ornithological Club, Cambridge, MA.
- (1977). Pleistocene speciation in Amazonian birds. *Amazoniana* **6**, 161–91.
- Hallock, P. (1987). Fluctuations in the trophic resource continuum: a factor in global diversity cycles? *Paleoceanography* **2**, 457–71.
- Hanski, I. and Gilpin, M. (1991). Metapopulation dynamics: brief history and conceptual domain. *Biol. Jour. Linn. Soc.* **42**, 3–16.
- Haq, B. U. (1981). Biogeographic history of Miocene calcareous nannoplankton and paleoceanography of the Atlantic Ocean. *Micropaleo.* **26**, 414–43.
- Hardy, M. C. (1985). Testing for adaptive radiation: the ptychaspid (Trilobita)

- biomere of the Late Cambrian. In *Phanerozoic diversity patterns* (ed. J. W. Valentine), pp. 379–97. Princeton University Press, Princeton, NJ.
- Harrison, S. (1991). Local extinction in a metapopulation context: an empirical evaluation. *Biol. Jour. Linn. Soc.* **42**, 73–88.
- Hickman, C. S. (1988). Analysis of form and function in fossils. *Am. Zool.* **28**, 775–93.
- Hoagland, K. E. and Turner, R. D. (1981). Evolution and adaptive radiation of wood-boring bivalves (Pholadacea). *Malacologia* **21**, 111–48.
- Hubbell, S. P. and Foster, R. B. (1986). Biology, chance, and history and the structure of tropical rain forest tree communities. In *Community ecology* (ed. J. Diamond and T. J. Case), pp. 314–29. Harper and Row, New York.
- Huston, M. (1979). A general hypothesis of species diversity. *Am. Nat.* **113**, 81–101.
- Jablonski, D. and Lutz, R. (1983). Larval ecology of marine benthic invertebrates: paleobiological implications. *Biol. Rev.* **58**, 21–89.
- Jackson, J. B. C. (1974). Biogeographic consequences of eurytopy and stenotopy among marine bivalves and their evolutionary significance. *Am. Nat.* **108**, 541–60.
- (1988). Does ecology matter? (review of *Evolution and escalation: an ecological history of life*, by G. J. Vermeij). *Paleobiology* **14**, 307–12.
- and McKinney, F. K. (1991). Ecological processes and progressive macroevolution of marine clonal benthos. In *Causes of evolution. A paleontological perspective* (ed. R. M. Ross and W. D. Allmon), pp. 173–209. University of Chicago Press, Chicago.
- Janis, C. M. (1989). A climatic explanation for patterns of evolutionary diversity in ungulate mammals. *Palaeontology* **2**, 463–81.
- Janzen, D. H. (1967). Why mountain passes are higher in the tropics. *Am. Nat.* **101**, 233–49.
- (1970). Herbivores and the number of tree species in tropical forests. *Am. Nat.* **104**, 501–28.
- (1977). Why are there so many species of insects? *Proceedings XV Cong. Entomol., Entomol. Soc. Am.*, pp. 84–94.
- Kay, E. A. (1984). Patterns of speciation in the Indo-west Pacific. In *Biogeography of the tropical Pacific* (ed. F. J. Radovsky, P. H. Raven and S. H. Sohmer), pp. 15–31. Bishop Mus. Spec. Publ. No. 72, Assoc. Syst. Coll. and Bernice P. Bishop Museum.
- Kochmer, J. P. and Wagner, R. H. (1988). Why are there so many kinds of passerine birds? Because they are small. A reply to Raikow. *Syst. Zool.* **37**, 68–9.
- Lambert, D. (1983). *A field guide to dinosaurs*. Avon Books, New York.
- Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314–34.
- (1979). Effective deme sizes during long-term evolution estimated from rates of chromosomal rearrangement. *Evolution* **33**, 234–51.
- (1980). Genetic variation and phenotypic evolution during allopatric speciation. *Am. Nat.* **116**, 463–79.
- Langer, P. (1987). Evolutionary patterns of Perissodactyla and Artiodactyla

- (Mammalia) with different types of digestion. *Z. Zool. Syst. Evolut.-forsch.* **25**, 212–36.
- Lauder, G. V. and Liem, K. F. (1989). The role of historical factors in the evolution of complex organismal functions. In *Complex organismal functions: integration and evolution in vertebrates* (ed. D. B. Wake and G. Roth), pp. 63–78. John Wiley and Sons, London.
- Lawton, J. H. (1989). What is the relationship between population density and body size in animals? *Oikos* **55**, 429–34.
- Leckie, R. M. (1989). A paleoceanographic model for the early evolutionary history of planktonic foraminifera. *Palaeogeog., Palaeoclimat., Palaeoecol.* **73**, 107–38.
- Lessios, H. A. and Cunningham, C. W. (1990). Gametic incompatibility between species of the sea urchin *Echinometra* on the two sides of the Isthmus of Panama. *Evolution* **44**, 933–41.
- Levins, R. (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull. Entomol. Soc. Am.* **15**, 237–40.
- (1970). Extinction. In *M. Gerstenhaber* (ed., Some mathematical problems in biology), pp. 77–107. American Mathematical Society, Providence, RI.
- Lewis, H. (1962). Catastrophic selection as a factor in speciation. *Evolution* **16**, 257–71.
- (1966). Speciation in flowering plants. *Science* **152**, 167–72.
- Lewontin, R. C. (1974). *The genetic basis of evolutionary change*. Columbia University Press, New York.
- (1982). The organism as the subject and object of selection. *Scientia* **118**, 65–82.
- (1983). Gene, organism and environment. In *Evolution from molecules to men* (ed. D. S. Bendall), pp. 273–85. Cambridge University Press, Cambridge.
- Liem, K. (1973). Evolutionary strategies and morphological innovations: cichlid pharyngeal jaws. *Sys. Zool.* **22**, 425–41.
- Lindroth, C. H. (1969). The theory of glacial refugia in Scandinavia. Comments on present opinions. *Notulae Entomologicae* **49**, 178–92.
- Lipps, J. H. and Mitchell, E. (1976). Trophic model for the adaptive radiations and extinctions of pelagic marine mammals. *Paleobiology* **2**, 147–55.
- Lloyd, E. A. (1988). *The structure and confirmation of evolutionary theory*. Praeger (Greenwood Press), New York.
- Lombard, R. E. and Wake, D. B. (1976). Tongue evolution in the lungless salamanders, family Plethodontidae. I. Introduction, theory and a general model of dynamics. *J. Morph.* **148**, 265–86.
- (1977). Tongue evolution in the lungless salamanders, family Plethodontidae. II. Function and evolutionary diversity. *J. Morph.* **153**, 39–80.
- Lynch, J. D. (1976). The species groups of the South American frogs of the genus *Eleutherodactylus* (Leptodactylidae). *Occ. Pap. Mus. Nat. Hist. Univ. Kansas* **61**, 1–24.
- (1979a). Leptodactylid frogs of the genus *Eleutherodactylus* from the Andes of southern Ecuador. *Misc. Publ. Mus. Nat. Hist. Univ. Kansas* **66**, 1–62.
- (1979b). The amphibians of the lowland tropical forests. In *The American herpetofauna: Its origin, evolution, and dispersal* (ed. W. E. Duellman), pp. 189–216. Mus. Nat. Hist. Univ. Kansas Monograph No. 7.

- MacArthur, R. H. (1972). *Geographical ecology*. Harper and Row, New York.
- and Wilson, E. O. (1967). *The theory of island biogeography*. Princeton University Press, Princeton, NJ.
- Maglio, V. J. (1973). Origin and evolution of the Elephantidae. *Trans. Am. Phil. Soc.* **63**, 1–149.
- Mayr, E. (1942). *Systematics and the origin of species*. Columbia University Press, New York.
- (1954). Change of genetic environment and evolution. In *Evolution as a process* (ed. J. Huxley, A. C. Hardy and E. B. Ford), pp. 157–80. Allen and Unwin, London.
- (1963). *Animal species and evolution*. Belknap Press of Harvard University Press, Cambridge, MA.
- (1969). Bird speciation in the tropics. *Biol. J. Linn. Soc.* **1**, 1–17.
- (1982). Processes of speciation in animals. In *Mechanisms of speciation* (ed. C. Barigozzi), pp. 1–20. Alan R. Liss, New York.
- (1988). The why and how of species. *Biol. Philos.* **3**, 431–41.
- McKinney, M. L., McNamara, K. J. and Zachos, L. G. (1990). Heterochronic hierarchies: application and theory in evolution. *Hist. Biol.* **3**, 269–87.
- McNamara, K. J. (1982). Heterochrony and phylogenetic trends. *Paleobiology* **8**, 130–42.
- Menge, B. A. and Sutherland, J. P. (1976). Species diversity gradients: synthesis of the roles of predation, competition and temporal heterogeneity. *Am. Nat.* **110**, 351–69.
- Meyer, D. L. and Macurda, Jr, D. B. (1977). Adaptive radiation of the comatulid crinoids. *Paleobiology* **3**, 74–82.
- Miller, A. H. (1950). Some ecologic and morphologic considerations in the evolution of higher taxonomic categories. *Ornithol. Biol. Wiss.* 84–8.
- Mitter, C., Farrell, B. and Wiegmann, B. (1988). The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *Am. Nat.* **132**, 107–28.
- Nevo, E. (1986). Mechanisms of adaptive speciation at the molecular and organismic levels. In *Evolutionary processes and theory* (ed. S. Karlin and E. Nevo), pp. 439–74. Academic Press, Orlando, FL.
- (1989). Modes of speciation: the nature and role of peripheral isolates in the origin of species. In *Genetics, speciation and the founder principle* (ed. L. V. Giddings, K. Y. Kaneshiro and W. W. Anderson), pp. 205–38. Oxford University Press, New York.
- Odum, E. P. (1971). *Fundamentals of ecology*, 3rd edn. W. B. Saunders, Philadelphia.
- Patterson, H. (1985). The recognition concept of species. In *Species and speciation* (ed. E. S. Vrba), pp. 21–30. Transvaal Mus. Monograph No. 4.
- Peters, R. H. and Raeslon, J. V. (1984). Relations between individual size and mammalian population density. *Am. Nat.* **124**, 498–517.
- Peters, R. H. and Wassenberg, K. (1983). The effect of body size on animal abundance. *Oecologia* **60**, 89–96.
- Petraitis, P. S., Lathan, R. E. and Niesenbaum, R. A. (1989). The maintenance of species diversity by disturbance. *Quart. Rev. Biol.* **64**, 393–418.
- Pianka, E. (1966). Latitudinal gradients in species diversity. *Am. Nat.* **100**, 33–46.

- (1978). *Evolutionary ecology*, 2nd edn. Harper and Row, New York.
- Pomeroy, D. (1990). Why fly? The possible benefits for lower mortality. *Biol. Linn. Soc.* **40**, 53–65.
- Powell, J. R. (1975). Protein variation in natural populations of animals. *Evol. Biol.* **8**, 79–119.
- Prance, G. T. (ed.) (1982). *Biological diversification in the tropics*. Columbia University Press, New York.
- Quinn, J. G., Wolin, C. L. and Judge, M. L. (1989). An experimental analysis of patch size, habitat subdivision, and extinction in a marine intertidal snail. *Conserv. Biol.* **3**, 242–51.
- Raikow, R. J. (1986). Why are there so many kinds of passerine birds? *Syst. Zool.* **35**, 255–59.
- Ricklefs, R. E. (1973). *Ecology*, 2nd edn. Chiron Press, Newton, MA.
- Roff, D. A. (1974). Spatial heterogeneity and the persistence of populations. *Oecologia* **15**, 245–58.
- Rose, M. R. and Doolittle, W. F. (1983). Molecular biological mechanisms of speciation. *Science* **220**, 157–67.
- Sale, P. F. (1977). Maintenance of high diversity in coral reef fish communities. *Am. Nat.* **111**, 337–59.
- Schaeffer, B. (1948). The origin of a mammalian ordinal character. *Evolution* **2**, 164–75.
- and Rosen, R. E. (1961). Major adaptive levels in the evolution of the actinopterygian feeding mechanism. *Am. Zool.* **1**, 187–204.
- Schoener, T. W. and Schoener, A. (1983). The time to extinction of a colonizing propagule of lizards increases with island area. *Nature* **302**, 332–4.
- Shmida, A. and Wilson, M. V. (1985). Biological determinants of species diversity. *J. Biogeog.* **12**, 1–20.
- Simpson, G. G. (1944). *Tempo and mode in evolution*. Columbia University Press, New York.
- (1953). *The major features of evolution*. Columbia University Press, New York.
- (1959). The nature and origin of supraspecific taxa. *Cold Spring Harbor Symp. Quant. Biol.* **24**, 255–71.
- Smith, A. (1984). *Echinoid palaeobiology*. Allen and Unwin, London.
- Smith, A. T. (1974). The distribution and dispersal of pikas: consequences of insular population structure. *Ecology* **55**, 1112–9.
- (1980). Temporal changes in insular populations of the pika (*Ochotona princeps*). *Ecology*, **61**, 8–13.
- Soule, M. E. (ed.) (1987). *Viable populations for conservation*. Cambridge University Press, New York.
- Spight, T. M. (1974). Sizes of populations of a marine snail. *Ecology* **55**, 712–29.
- Stanley, S. M. (1968). Post-Paleozoic adaptive radiation of infaunal bivalve molluscs – a consequence of mantle fusion and siphon formation. *J. Paleontol.* **42**, 214–29.
- (1973). Effects of competition on rates of evolution, with special reference to bivalve mollusks and mammals. *Syst. Zool.* **22**, 486–506.
- (1977). Trends, rates and patterns of evolution in the Bivalvia. In *Patterns*

- of evolution, as illustrated in the fossil record (ed. A. Hallam), pp. 209–50. Elsevier, Amsterdam.
- (1978). Chronospecies' longevities, the origin of genera, and the punctuational model of evolution. *Paleobiology* **4**, 26–40.
- (1979). *Macroevolution. Pattern and process*. W. H. Freeman, San Francisco.
- (1984). Temperature and biotic crises in the marine realm. *Geology* **12**, 205–8.
- (1986). Population size, extinction, and speciation: the fission effect in Neogene Bivalvia. *Paleobiology* **12**, 89–110.
- (1991). The general correlation between rate of speciation and rate of extinction: fortuitous causal linkages. In *Causes of evolution. A paleontological perspective* (ed. R. M. Ross and W. D. Allmon), pp. 103–27. University of Chicago Press, Chicago.
- and Newman, W. A. (1980). Competitive exclusion in evolutionary time: the case of the acorn barnacles. *Paleobiology* **6**, 173–83.
- States, J. B. (1976). Local adaptation in chipmunk (*Eutamias amoenus*) populations and evolutionary potential at species' borders. *Ecol. Monogr.* **46**, 221–56.
- Stebbins, G. L. (1969). *The basis of progressive evolution*. University of North Carolina Press, Chapel Hill.
- (1974). *Flowering plants: evolution above the species level*. Belknap Press of Harvard University Press, Cambridge, MA.
- Stenseth, N. C. (1984). The tropics: cradle or museum? *Oikos* **43**, 417–20.
- Stevens, G. C. (1989). The latitudinal gradient in geographical range: how so many species coexist in the tropics. *Am. Nat.* **133**, 240–56.
- Tauber, C. A. and Tauber, M. J. (1989). Sympatric speciation in insects: perception and perspective. In *Speciation and its consequences* (ed. D. Otte and J. A. Endler), pp. 307–44. Sinauer Associates, Sunderland, MA.
- Taylor, A. D. (1990). Metapopulations, dispersal, and predator-prey dynamics: an overview. *Ecology* **71**, 429–33.
- Taylor, J. D., Morris, N. J., and Taylor, C. N. (1980). Food specialization and the evolution of predatory prosobranch gastropods. *Palaeontology*, **23**, 375–409.
- Taylor, P. D. (1989). Major radiation of cheilostome bryozoans triggered by the evolution of a new larval type. *Hist. Biol.* **1**, 45–64.
- Tchernov, E. (1982). Faunal responses to environmental changes in the Eastern Mediterranean during the last 20,000 years. In *Palaeoclimates, palaeoenvironments and human communities in the eastern Mediterranean region in later prehistory* (ed. J. L. Bintiff and W. Van Zeist), pp. 105–87. BAR Int. Series 133, Oxford.
- Templeton, A. (1980). The theory of speciation via the founder principle. *Genetics* **94**, 1011–38.
- (1981). Mechanisms of speciation – a population genetic approach. *Ann. Rev. Ecol. Syst.* **12**, 23–41.
- (1982). Adaptation and the integration of evolutionary forces. In *Perspectives on evolution* (ed. R. Milkman), pp. 15–31. Sinauer Associates, Sunderland, MA.
- (1989). The meaning of species and speciation. In *Speciation and its consequences* (ed. D. Otte and J. Endler), pp. 3–27. Sinauer Associates, Sunderland, MA.

- Turner, J. R. G., Gatehouse, C. M., and Corey, C. A. (1987). Does solar energy control organic diversity? Butterflies, moths and the British climate. *Oikos* **48**, 195–205.
- Valentine, J. W. (1984). Climate and evolution in the shallow sea. In *Fossils and climate* (ed. P. J. Brenchley), pp. 265–78. John Wiley and Sons, New York.
- Van Valen, L. (1971). Adaptive zones and the orders of mammals. *Evolution* **25**, 420–8.
- (1973). Body size and numbers of plants and animals. *Evolution* **27**, 27–35.
- Vaughan, T. A. (1978). *Mammalogy*, 3rd edn. Saunders College Publ., Philadelphia.
- Vermeij, G. J. (1987a). *Evolution and escalation. An ecological history of life*. Princeton University Press, Princeton, NJ.
- (1987b). The dispersal barrier in the tropical Pacific: implications for molluscan speciation and extinction. *Evolution* **41**, 1046–58.
- (1988). The evolutionary success of passerines: a question of semantics? *Syst. Zool.* **37**, 69–71.
- Vrba, E. S. (1980). Evolution, species and fossils: how does life evolve? *S. Afr. J. Sci.* **76**, 61–84.
- (1983a). Macroevolutionary trends: new perspectives on the role of adaptation and incidental effect. *Science* **221**, 387–89.
- (1983b). What is species selection? *Syst. Zool.* **33**, 318–28.
- (1984). Patterns in the fossil record and evolutionary processes. In *Beyond Neo-Darwinism* (ed. M.-W. Ho and P. T. Saunders), pp. 115–42. Academic Press, London.
- (1985). Environment and evolution: alternative causes of the temporal distribution of evolutionary events. *S. Afr. J. Sci.* **81**, 229–36.
- (1987). Ecology in relation to speciation rates: some case histories of Miocene-Recent mammal clades. *Evol. Ecol.* **1**, 283–300.
- and Eldredge, N. (1984). Individuals, hierarchies and aprocesses: towards a more complete evolutionary theory. *Paleobiology*, **10**, 146–71.
- Wallace, G. J. and Mahan, H. D. (1975). *An introduction to ornithology*. 3rd edn. Macmillan, New York.
- Wells, J. W. (1956). Scleractinia. In *Treatise on invertebrate paleontology, Part F, Coelenterata* (ed. R. C. Moore), pp. 328–444. Geol. Soc. Am. and University of Kansas Press, Lawrence, KS.
- Wells, P. V. (1983). Paleobiogeography of montane islands in the Great Basin since the last glaciopluvial. *Ecol. Monogr.* **53**, 341–82.
- White, M. J. D. (1978). *Modes of speciation*. W. H. Freeman, San Francisco.
- Whitmore, T. C. and Prance, G. T. (eds) (1987). *Biogeography and quaternary history in tropical America*. Oxford University Press, Oxford.
- Whittaker, R. H. (1977). Evolution of species diversity in land communities. *Evol. Biol.* **12**, 117–39.
- Wiley, E. O. (1981). *Phylogenetics*. John Wiley, New York.
- Williams, E. E. (1972). The origin of faunas. Evolution of lizard congeners in a complex island fauna: A trial analysis. *Evol. Biol.* **6**, 47–89.
- Wilson, A. C., Bush, G. L., Case, S. M. and King, M. C. (1975). Social structuring of mammalian populations and rate of chromosomal evolution. *Proc. Nat. Acad. Sci. USA* **72**, 5061–5.

- Wilson, D. S. (1989). The diversification of single gene pools by density- and frequency-dependent selection. In *Speciation and its consequences* (ed. D. Otte and J. A. Endler), pp. 366–85. Sinauer Associates, Sunderland, MA.
- Wilson, E. O. (1961). The nature of the taxon cycle in the Melanesian ant fauna. *Am. Nat.* **95**, 169–93.
- Wilson, R. W. (1951). Evolution of the early Tertiary rodents. *Evolution* **5**, 207–15.
- Wood, A. E. (1959). Eocene radiation and phylogeny of the rodents. *Evolution* **13**, 354–61.
- Wright, D. H. (1983). Species-energy theory: an extension of species-area theory. *Oikos* **41**, 496–506.
- Wright, S. (1982). Character change, speciation, and the higher taxa. *Evolution* **36**, 427–43.

This page intentionally left blank

Inferences from allometry and fossils: dwarfing of elephants on islands

V. LOUISE ROTH

1. INTRODUCTION: THE ISLAND RULE FOR MAMMALIAN BODY SIZE

Through their effects on population dynamics and community composition, islands represent settings for evolutionary and ecological ‘natural’ experiments. The consequence of the low immigration rates, high extinction rates, limited geographic range, and maritime climates that islands impose on their colonists is a depauperate and often unique fauna and flora which have been a source of interest to naturalists since Humboldt, Wallace, and Darwin. Research on islands intensified in the decades following publication of MacArthur and Wilson’s work (1967), and islands now serve as models for other isolated (but non-island) systems (Wilcox 1980; Brown 1986). In a world where many natural habitats are becoming increasingly insular, an understanding of the evolutionary and ecological processes occurring on islands assumes even greater significance.

The tendency of small herbivorous mammals to enlarge, and carnivores and ungulates to dwarf on islands ‘seems to have fewer exceptions than any other ecotypic rule in animals’ (Van Valen 1973). Giant rodents and dwarfed proboscideans on islands off California and Southeast Asia and in the Mediterranean are thus intriguing in part simply for the regularity of the trend they exemplify, but in addition because body size is of fundamental importance as a parameter in organismal and population biology. In ecology, body size may affect division of resources (Hutchinson 1959; Brown 1975) or determine an animal’s place in community structure (Brown 1975; Van Valkenburgh 1988); it is correlated with lifespan, reproductive rate, and other demographic variables upon which life history strategies depend (Blueweiss *et al.* 1978; Peters 1983; Stearns 1983; Calder 1984). Because weight, area, and linear dimension change with geometric similarity at different rates, body size influences an animal’s mechanical interaction with its environment (Gould 1966; McMahon 1975), metabolic rate (Kleiber 1961; Schmidt-Nielsen 1984), and the relative generalization or specialization – and hence the evolutionary and ecological potential – of its lineage (Stanley 1973). Furthermore, maintenance of function may limit the size of an animal,

or, inversely, size change may facilitate change in the function associated with a specific morphology (Gould 1971). Change in body size on islands thus is a topic in evolutionary biology whose understanding emerges from the interaction of several major themes in biogeography, ecology, and physiology.

1.1 The pattern

Foster (1964) was the first to note what Van Valen (1973) termed the 'island rule in mammals' in comparing body sizes of 116 species and varieties of mammals living on islands in Europe and western North America with those of their closest relatives on the mainland. The pattern was a striking one, which has since been confirmed for other parts of the world, and for a variety of other island faunas, both Pleistocene and Recent (Hooijer 1967; Orr 1968; Sondaar 1977; Heaney 1978). Although Foster organized his tabulation by taxonomic group – artiodactyls, lagomorphs, and carnivores apparently dwarfed on islands whereas rodents usually evolved larger size – a stronger and more general pattern emerged when species were arrayed according to body size itself. Graphing the ratio of average island and mainland body weights against the log-transformed body weight of the mainland form, Lomolino (1985) noted a negative linear trend that crossed the line representing 'no difference' between island and mainland body size at approximately 250 g (the size of a small rabbit). In a graded trend from the smaller to the larger species, gigantism decreased and dwarfism increased by degree with increasing body size. This tendency (indicated by a positive slope in a regression of log-transformed body mass of the mainland population on that for the island) was significant not only for the entire sample of mammalian species combined, but also within nearly all orders (Insectivora excepted) and nearly all families (the Muridae was also an exception, but was represented in the sample by only two species). Other descriptions of the island rule, as a trend defined by taxonomic groupings (Foster 1964; Van Valen 1973), or varying with island area (Heaney 1978) or distance from the mainland (Foster 1964), yielded less regular patterns than that correlated with body size (Heaney 1978; Lawlor 1982; Roth 1982; Lomolino 1985).

1.2 Proposed causes

Regularities demand a general explanation. Still, multiple, complex causes can sum to produce deceptively simple patterns, while exceptions and special cases command special attention. My focus in this paper will be on an extreme but circumscribed example: the dwarfism of elephants on

Pleistocene islands. I will concentrate on two areas with whose material I am most familiar: the Santa Barbara, California Channel Islands (Roth 1982) and the island of Sicily (Ambrosetti 1968), and I will consider the following question: What can be inferred about the biology (and particularly the evolution) of dwarfed elephants from the gross morphology of their fossils?

In this section, however, I will briefly review some of the causes proposed for the trends for insular mammals overall. I limit the discussion to terrestrial mammals because other terrestrial vertebrates exhibit somewhat different patterns in body size (e.g. Grant 1965; Case 1978), perhaps in part because birds have different constraints on their mobility, and reptiles have different life history patterns and energetic demands.

To discern the selective forces that might have led to the most general features of the island rule, ecological consequences of various body sizes can be considered in the context of the characteristics of island habitats. The relative importance of various ecological factors differs between island and mainland because islands are isolated and limited in area. Island faunas are generally depauperate (species diversity declines with island area: MacArthur and Wilson 1967; Carlquist 1974), so interspecific competition is reduced, or, at least, involves fewer species. Predation is also decreased. Islands support a lower number and diversity of predators, both because they support fewer species overall, and because carnivores require larger home ranges than herbivores and their population densities are lower (McNab 1963). Restrictions on migration affect island immigrants as well as residents. Especially among the smaller species, if large individuals are more successful in surviving overseas dispersal and colonization, the founding population on an island may be initially biased toward large body size ('immigrant selection'; Lomolino 1984). As a result of its confinement to an island, a population may experience limitations in food supply that are more severe than on the mainland. Isolated populations of large mammals that reach high densities unrelieved by emigration may become overcrowded and reduce their resource base yet further by destroying habitat (Laws *et al.* 1975; Johnson 1980).

Heaney (1978) reviewed the relevant relationships between body size and the various ecological factors that are expected to differ on islands, and Lomolino (1985) summarized them: gigantism in small mammals on islands is probably a consequence of immigrant selection and release from competition, whereas dwarfism in large mammals results from resource limitation.

Both very large and very small body sizes can be a defense against predators: the largest herbivores may overwhelm or outrun a predator, whereas the smallest ones can hide from them (Valverde 1964). Although large felids are known to have taken juvenile mammoths as prey on the

North American mainland (Graham 1976), large mammalian predators were absent from islands inhabited by dwarfed elephants in the Pleistocene. Reduced predation on islands should therefore be a permissive factor, allowing small mammals to attain larger sizes and large mammals to attain smaller sizes. It has even been suggested most generally that for all mammals there is an intermediate optimum body size that, in the absence of other selective forces, will be favored for reasons of physiological efficiency (Maiorana 1990).

For small mammals on islands, where the number of competing species is reduced, selection may favor larger body sizes which allow the exploitation of a greater variety and size range of foods (Grant 1965; Brown 1975). This selection for large size should prevail among animals that establish individual feeding territories (Case 1978) and among dietary generalists (Lawlor 1982; Angerbjörn 1985). In addition, for small mammals that continue some growth beyond sexual maturity but that suffer high mortality before reaching maximum size, reduced predation and competition can permit a species to attain larger average size without evolving, simply by extending survival and allowing more individuals to grow larger.

For large mammals, resource limitation due to the small biomass available to inhabitants of small, isolated islands, favors smaller body size. Small size can arise through disparate mechanisms. Stunting, a manifestation of phenotypic plasticity under nutritional deficiency, is by definition not associated with genetic differentiation. Yet the ability to stunt may be enhanced for some genotypes, and can thus be subject to selection. In an environment that fosters stunting, individuals capable of attaining reproductive maturity on a low plane of nutrition will be favored (Bonner 1968; Geist 1978, p. 263; Heaney 1978; Marshall and Corruccini 1978). Small size can also be maintained through several generations by maternal effects: regardless of the cause of their own small size, small individuals tend to give birth to small offspring. In small, polytocous species such as rodents, non-genetic maternal effects on body size are large enough to be measurable (e.g. Roth and Klein 1986); in larger monotocous forms (and hence, probably elephants) the effects are greater and often persist into adulthood (Snow *et al.* 1981).

Group selection has also (implicitly) been invoked. Computer simulations of extinction and body-size evolution within lineages demonstrate that 'evolution toward small size permits the closest tracking of carrying capacity when immigration and emigration are not possible' (Wassersug *et al.* 1979). If differential extinction of populations (in the absence of selection for small size among individuals) were the reason we observe dwarfism of large mammals on islands, however, one could expect at least sporadically to find examples of gigantism or no size change on islands in the fossil record. Instead, almost without exception, dwarfism

predominates (Van Valen 1973; Lomolino 1985).

There are some circumstances for large mammals in which resource limitation is expected *not* to favor size reduction. Lindstedt and Boyce (1985) have argued on the basis of the scaling of fasting endurance that in environments that impose occasional food shortages, selection should in fact favor large body size – a result consistent with Bergmann's rule, and with the success of large mammals in highly seasonal environments. There seems little question that the fat reserves of a larger mammal should allow it to survive fasting for longer than a small mammal, yet it remains a problem to *attain* and *maintain* the large size, which requires absolutely more food per individual (McNab 1971; Boyce 1984). Large herbivores may 'be limited to grazing principally on communities with comparatively high standing crops in order to achieve a rate of food intake sufficient to satisfy requirements,' assuming they cannot compensate by increasing the time they spend feeding (Clutton-Brock and Harvey 1983, p. 653), a distinct possibility for elephants, which typically devote 15–19 h daily to foraging (Owen-Smith 1988). Lindstedt and Boyce suggest that when food shortages cause high mortality, survivors subsequently face low competition and abundant resources. Such conditions could permit survivors and their descendants to become large, but what is selectively favored in times of shortage is not large size *per se*, but rather ability to survive and reproduce despite the shortages. The problem becomes a matter of allocation, where 'structural' body mass – the frame of bone and muscle on which energy reserves in the form of fat are supported – the frame of bone and muscle on which energy reserves in the form of fat are supported – itself is costly to build. In sub- and young adult mammals, growth priority for structural tissues is higher than that of adipose tissue, which is to say bone and muscle are laid down relatively early, establishing the frame on which fat reserves are built up later. If, during the ontogeny of an individual, resources are in short supply and do not alternate predictably (e.g. seasonally) between limitation and superabundance, then given alternatives of (1) investing in a large body frame or (2) halting growth at a smaller size and shifting priority to the storage of fat, which is more readily mobilized to serve metabolic needs in times of shortage than are bone or muscle, the second may be favored. Several field studies on large mammals, including red deer (Clutton-Brock *et al.* 1982, 1985), African elephants (Laws and Parker 1968), reindeer (Skoglund 1990), and mountain sheep (Geist 1971), provide empirical data suggesting that smaller individuals, which invest less in skeletal growth and presumably more in body fat or reproduction, tend to have higher fitness when populations are food-stressed (see also Boyce 1984, p. 436; Merritt and Merritt 1978).

2. THE OCCURRENCE OF DWARFED ELEPHANTS

2.1 Modern analogues

For our purposes, an island inhabitant is dwarfed if it typically attains a smaller size when fully grown than do forms from the mainland. The occurrence of dwarf elephants in the fossil record is relatively unambiguous: osteological specimens provide hard evidence of body size and state of physical maturity. Elongation of limb bones ceases and growth is complete once epiphyses fuse. Regardless of how small a tooth is, for elephants large numbers of lamellae (enamel loops) and a relatively elongate shape signal a mature individual.

By contrast, controversy persists over whether a pygmy elephant distinct from but closely related to the modern bush elephant, *Loxodonta africana africana*, exists in modern tropical African forests. What has been called the 'forest elephant,' *Loxodonta africana cyclotis*, is well accepted (Laursen & Beckoff 1978) and well documented by small but relatively mature osteological specimens in museum collections. In linear dimensions, *L. a. cyclotis* reaches about 80 per cent the size of the larger subspecies. Existence of an even smaller (less than 2 m in shoulder height; cf Table 1) 'pygmy' form, *Loxodonta pumilio*, however, is problematic. Reports of pygmy elephants less than 2 m in shoulder height (cf Table 1) are anecdotal, often field observations in which heights were estimated and pregnancy or possession of long tusks taken as signs of physical maturity. Yet female elephants typically reach sexual maturity at just 80 per cent

Table 1
Maximum asymptotic shoulder heights^a (in meters) of modern elephants (populations within each taxon vary in their size ranges)

	<i>Elephas maximus</i>	<i>Loxodonta africana africana</i>	<i>L. a. cyclotis</i>
Asian elephant		African bush elephant	African forest elephant
Females	2.5	2.9	2.3
Males	3.2	3.4	2.7
Recorded maximum	3.4	4.0	3.0*
Reference source	Shoshani and Eisenberg (1982)	Owen-Smith (1988)	Dorst and Dandelot (1969)
			*Morrison-Scott (1947) and Eisentraut & Böhme (1989)

^a An asymptotic shoulder height is a limit approached by a continuous growth curve.

of their asymptotic shoulder height (Perry 1954, Laws 1966), and sizes of tusks vary tremendously among individuals, the largest more likely to be found in populations inaccessible to selection by ivory poachers. Shapes and presence or absence of tusks appear to have some degree of heritability, and Eltringham (1982) has speculated that the tusklessness of Sri Lankan elephants may be a consequence of selection by poaching.

The type specimen of *L. pumilio* was a captive individual (Noack 1906), a small animal from Gabon whose age was estimated and who died young (Morrison-Scott 1947). Possibly stunted or in poor condition, it did not live to attain full height. In museum collections in Britain and North America I have not seen any specimens of individuals under 2 m in height that were sufficiently advanced in ossification to be considered close to fully grown (Roth 1984). No study has yet documented the existence of a modern pygmy elephant with voucher specimens of precisely determined dental stage or stage of ossification, placed in a context of intraspecific variability. In elephants, individual variation and sexual dimorphism are marked. Elephants at birth are less than 20–30 per cent of their heights as adults; attainment of full height takes several decades, years after sexual maturity, and senescence is accompanied by further changes in morphology. The thinning of the mandibular ramus and the changes in mandibular shape cited by Eisentraut and Böhme (1989) as evidence of the morphological distinctiveness of one putative pygmy specimen are features I have commonly observed in old individuals with extensively worn teeth. The one study (a thesis by Oberdörfer in 1984 at the University of Bonn) cited by Eisentraut and Böhme (1989) and Greenwell (1990) that evaluated quantitatively the morphological distinctiveness of *L. a. pumilio* from *L. a. cyclotis* found no significant difference. Diminutive elephants among the extant fauna would be a welcome discovery, but at present, living populations of adult elephants less than 2 m high are still in the purview of cryptozoology.

2.2 Pleistocene insular dwarfs

Table 2 and Fig. 1 show the locations of named forms of fossil dwarf proboscideans. The material evidence varies from discoveries of a few isolated teeth, to extensive cave deposits yielding statistical samples. Size reduction on the islands is believed to have been rapid (Gould 1975; Hooijer 1975): individuals on the islands vary considerably in size (reflected in Table 2's listing of more than one named form per island; see also Roth in press), but in few or no cases are fossils intermediate in size also clearly intermediate in stratigraphic position. Moreover, the size variation that exists is, as far as it is known, not correlated in any simple way with time, suggesting fluctuation rather than monotonic decrease, and possibly repeated invasions (and introgression) of larger

Table 2

Dwarfed proboscideans from Pleistocene islands. M. = *Mammuthus*, E. = *Elephas*. Stegodontids are extinct elephant-like proboscideans more closely related to mastodons than to true elephants (which constitute the family Elephantidae, comprising *Elephas*, *Mammuthus*, and *Loxodonta*)

Map reference number (Fig. 1)	Islands	Taxa	Sample literature sources
1.	San Miguel and Santa Rosa, CA	<i>M. exilis</i>	Stock & Furlong (1928)
A.	2. Malta and Sicily	<i>E. falconeri</i> <i>E. melitensis</i> <i>E. mnaidriensis</i>	Busk (1867) Vaufrey (1929) Ambrosetti (1968)
	3. Sardinia	<i>M. lamarmorae</i>	
	4. Cyclades: Kithnos Milos Delos Naxos Serifos	<i>Elephas spp.</i>	For overview, see Caloi <i>et al.</i> (1989, and in press), Theodorou (1986), and Sondaar (1977)
	5. Tilos		
	6. Rhodos		
	7. Crete		
	8. Cyprus		
B.	9. Mindanao and Luzon		For overview, see Sondaar <i>et al.</i> (1989), Sondaar (1977), Vos (in press), and Bergh <i>et al.</i> (in press)
	10. Sulawesi		
	11. Java		
	12. Sumba		
		Various stegodontids	
	13. Flores		
	14. Timor		
15.	Miyako, Okinawa, Japan		

individuals from the mainland (Orr 1968; Burgio and Cani 1988; Caloi *et al.* 1989).

For most sites, many questions remain to be addressed that will require further excavation and careful stratigraphy and dating. Yet already much can be learned even from single fragmentary specimens. Provided they can be identified anatomically and taxonomically, and their size, shape, and state of maturity can be assessed, skeletal and dental material permit us to deduce (1) body masses, which are in turn correlated with many aspects of

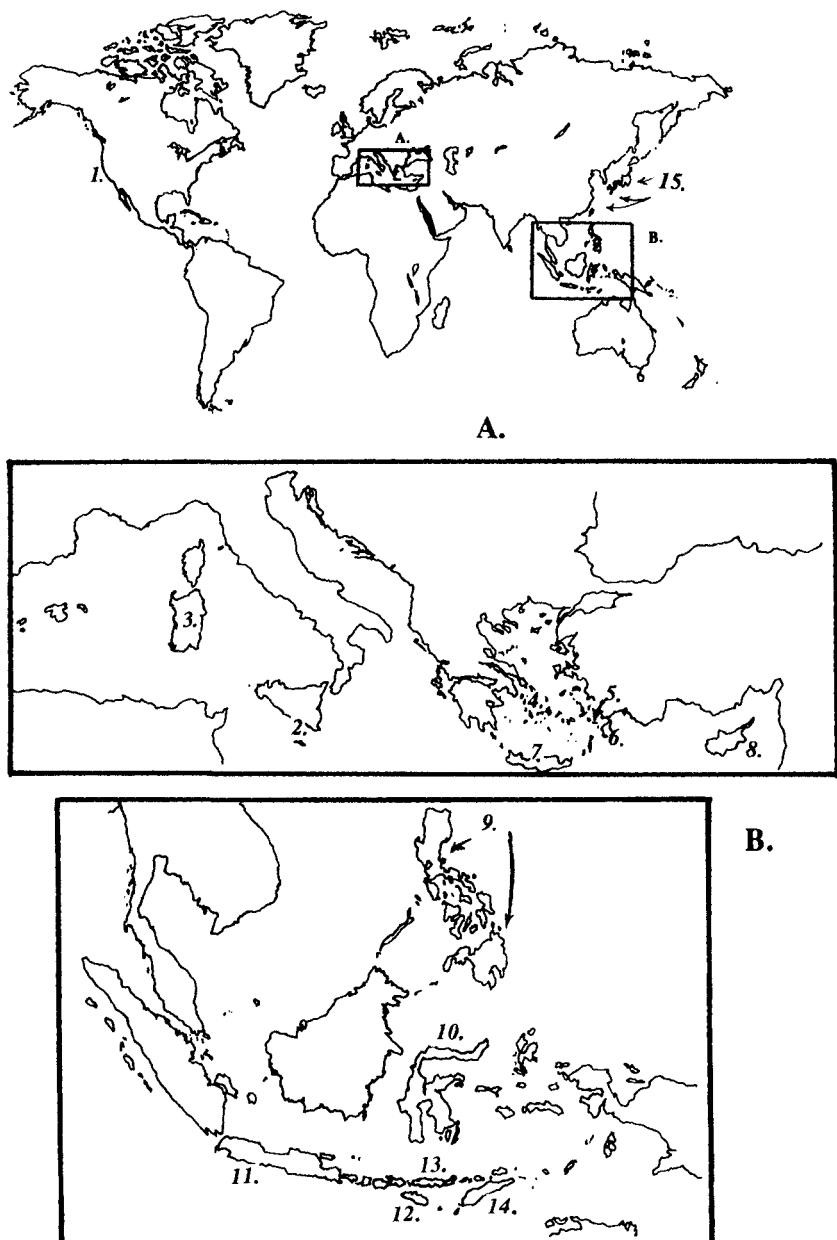


Fig. 1. Islands bearing Pleistocene fossil deposits of dwarfed proboscideans. See Table 2 for key. Some of these islands, and others, also bear fossils of giant rodents, swans, tortoises, and dwarfed versions of hippos, deer, or other large mammals.

life history; (2) body proportions, and their implications for heterochronic evolution and biomechanical function; and (3) important size-related features of population dynamics and evolutionary pattern and process.

3. INFERENCES

3.1 Body mass and life history

Among mammalian quadrupeds, lengths, diameters, and circumferences of limb bones scale in regular ways with body size (Alexander *et al.* 1979; Anderson *et al.* 1985); as a consequence, these dimensions can be used to estimate body mass for fossils. I have described methods of (and sources of error in) estimating body masses for fossil elephants of various sizes elsewhere (Roth 1990). Small elephants have relatively robust limb bones, and for (juvenile or dwarfed) elephants less than 2 metric tons, lengths yield the best estimates, whereas above 2.5 tons, regressions based on circumferences or on shoulder heights of modern populations are more reliable.

For reference, it will be helpful to keep in mind that fully grown modern female African elephants typically attain masses of 2.5–3.2 tons, whereas large individual males have been recorded at 6.6 tons and estimated at up to 10 tons (Owen-Smith 1988). All of the sources of variation cited in the previous section (sexual dimorphism, interindividual and interpopulational variation, etc.) for shoulder heights of modern populations of elephants apply as well to body mass, with the additional complication that individuals of identical height can differ by as much as a factor of 2 or 3 in mass, depending on body build (e.g. *Elephas* vs. *Loxodonta*) and physical condition. Male Asian elephants weigh up to 5.4 tons, and females are known to reach 4.2 tons, but more typically attain 2.7 tons. I will use as my standard in some of the comparisons that follow a typical female African elephant, fully grown at 3 tons.

Shoulder heights and limb-bone circumferences of *Mammuthus columbi*, the Columbian mammoth that was ancestral to the dwarf mammoths inhabiting the Pleistocene California Channel Islands, suggest adult body masses of 5–11 tons. *Elephas antiquus*, which is considered ancestral to the Mediterranean island dwarfs (and, according to Maglio (1973), synonymous with *E. namadicus*, described from Asia), may have weighed between 8.7 and 17 tons, if Kurtén's (1968) and Osborn's (1942) estimates of 3.9–4 m in shoulder height are accurate. (A height of 4 m is not unusual for Pleistocene elephants, notwithstanding the more modest size of the better known wooly mammoth; Mol and Agenbroad MS). By contrast, the insular form *M. exilis* (which varied between 1.2 and 2.5 m in height) ranged from 200 kg to 2 tons when physically mature. For the smallest of insular dwarf elephants, *E. falconeri*, 1 m high and represented

by excellent samples from Sicily and Malta, males weighed on average approximately 130 kg, and females 80 kg, when fully grown (Roth 1990).

It is clear that size reduction for insular elephants was extreme not only in absolute terms, but also in terms of the empirical relationship Lomolino (1985) derived for mammals generally. According to Lomolino's (1985) log-linear regression of island on mainland body size for his full sample of mammals, from animals of 10 tons on the mainland one would predict a reduction to 6 tons on an island. By the equation for artiodactyls (the largest animals – and those most analogous ecologically to elephants – included in his samples), insular dwarfs of 3.3 tons are expected. The forms known from islands are generally considerably smaller than this. Hence, a relationship between mainland and island body sizes that extrapolates smoothly out to the body size of elephants may be log-curvilinear, rather than log-linear.

From the reciprocal relationship between the scaling of body mass with population density and metabolic rate, Damuth (1981) observed that the energy used by local populations of mammals is independent of body size. This empirical observation suggests that body size reduction on islands may have evolved in association with high population density as a compensatory reduction of energy use (Roth 1990). However, it turns out that the difference in energy use between populations at the maximum sustainable densities known for modern elephants, and the more comfortable densities sustained by unstressed populations, is insufficient to account for the magnitude of the size reduction from full-sized *E. antiquus* to its descendant *E. falconeri* (the tiniest of the insular dwarfs). Assuming this model is appropriate, one must imagine, instead of a single compensatory adjustment, either that a declining resource base (from habitat destruction, for example) accompanied the size reduction, or that a more iterative process was involved: overcrowding, followed by dwarfing, followed by still further increases in density and further dwarfing (Roth 1990).

Mainland forms of Pleistocene elephants were commonly larger than any living quadruped, and many existing empirical allometric equations do not extend even into the size range of living elephants. As a consequence, some of the calculations that follow will by necessity involve extrapolation beyond the size range for which the regressions were empirically derived. There are flaws in this approach, but we have few alternatives when there exist no living analogues. All of the inferences I make below are implicitly accompanied by the proviso 'if one can assume that scaling relationships known for other mammals apply.'

Fasting endurance (Table 3) for ancestral elephants from the mainland is estimated to exceed 1 year, a staggering figure that is derived from the allometry of body fat, calculations of its stored energy, and an assumption of metabolic expenditures near basal level (Lindstedt and Boyce 1985).

Table 3
Predicted values of life-history parameters for elephants of various sizes

		Predictions (original reference source of equation)			
Representative species	Body mass	Fasting endurance		Gestation period	Postnatal time to 98% full size (3)
		(1)	(2)		
<i>E. falconeri</i> (Sicily)	100 kg	71 days	1.7	189 days	4 years
<i>M. exilis</i> (California islands)	200 kg	96 days	1.45	223 days	4 years 9 months
	1 ton	194 days	1	326 days	7 years 3 months
<i>L. africana</i> (modern Africa)	2 tons	264 days	0.76	384 days	8 years 8.5 months
	3 tons predicted:	315 days	0.64	424 days	9 years 8.5 months
	known:	—	1+ (twinning known on rare occasions)	~660 days	15–20 years
<i>M. columbi</i> (mainland North America)	10 tons	535 days	0.3	564 days	13 years 3 months
<i>E. antiquus</i> (mainland Europe)	15 tons	640 days	0.16	622 days	14 years 8 months

(1) Lindstedt & Boyce (1985); (2) Millar (1981); (3) Brody (1945), Peters (1983).

Lindstedt and Boyce (1985) argue that this last assumption is realistic for mammals in an 'energy survival crisis,' but an elephant keeping relatively inactive for many months is likely to encounter other problems before energy supplies become critical. The figures provide at least a qualitative index of ability to survive a fast, and endurance of such magnitude is a favorable trait in an animal able to migrate long distances to exploit new food sources. However, when animals must remain in an area beyond the time at which they destroy habitat and the resource base deteriorates from overuse (Laws *et al.* 1975), simply waiting (fasting) for conditions to improve may be a losing proposition. McCullagh (1969) calculated that the diet of African elephants in a normal dry season is only just sufficient to provide the protein required for maintenance. If food shortage is a chronic, rather than a seasonal problem, individuals with reduced

requirements may be favored relative to those simply able to endure a fast. Dwarfed elephants on Sicily and the California Islands were restricted in their fasting endurance to a few months; yet, given the 0.75 exponent scaling metabolic expenditure to body size (Kleiber 1961; c.f. Heusner 1982; Nagy 1987), the metabolic requirements of a 200 kg individual are only 5 per cent of those of a 10 ton relative, and the small individual could survive on an absolutely lower plane of nutrition.

Dietary selectivity increases with decreasing body size in mammalian herbivores (Jarman 1974). Forage of low nutritive value is sufficient for animals above about 300 kg, which tend to range over large geographic areas. At body sizes less than this, mammalian herbivores tend to feed more selectively, on particular growth stages or parts of plants of higher nutritive value. Gut capacity scales isometrically, so retention time for gut contents declines with decreasing body size, whereas mass-specific metabolic rate rises, necessitating the shift to more readily digestible foods (Demment and Van Soest 1985). Hindgut fermenters such as elephants can tolerate (and in fact probably require; Laws *et al.* 1975; Eltringham 1982) higher-fiber diets at lower body sizes than can ruminants; they consume a larger volume of food to make up for low quality of forage (Janis 1976). The proportion of grass in the diets of modern elephants is typically about 50 per cent, although amounts vary greatly with availability, habitat, and season (Eltringham 1982; Sukumar 1989), preference apparently being greatest in the wet seasons when grass is actively growing, rich in nutrients, and low in silica and fiber. The remainder of the diet comprises foliage of a wide variety of browse plants, forbs, and some bark and woody portions of trees. In his studies of the (small) African forest elephants in the Ivory Coast, Merz (1977, cited by Eltringham 1982) observed that seeds (the fruits of which tend to be calorie-rich and low in fiber and toxins by comparison with foliage and wood) constituted up to 35 per cent of the dry weight of elephant droppings during fruiting seasons. Of course, without detailed knowledge of types of soils and available vegetation it is difficult to make inferences about sources and limitations of nutrients other than calories and protein for any animal, fossil or extant. However, dwarfed elephants are likely to have experienced something of a tradeoff: a reduction in total metabolic requirements, on the one hand, coupled with a need for higher-quality browse, on the other. Moreover, small elephants would not have access to the taller vegetation that full-sized animals commonly reach or knock down (Laws *et al.* 1975). Nevertheless, even the smallest adult insular dwarfs would have found 60 per cent of the browse taken by large modern elephants within easy reach (Guy 1976). Thus, despite the complicating factors – and although the total food resources available to dwarfed elephants (as herbage within reach of the trunk, and as nutrients made accessible by digestion) are probably less than for larger forms – still, as

a net effect, energy requirements would have been more readily met for small individuals.

Regressions of litter size and gestation time on body weights of mammals do not predict accurately the values known for living elephants, but the deviations may (perhaps not coincidentally) be mutually compensatory. Successful litters consist of whole individuals, so an allometric relationship predicting a litter size less than one cannot be accurate. If, however, the scaling relationship reflects size-related trends in (or constraints on) energy expenditure on reproduction, a prediction of 0.3 (for example) may simply indicate that producing a single offspring requires corresponding compensatory adjustment in other physiological parameters for that species, e.g. a longer gestation period than would be predicted on the basis of size alone. Accordingly, a 3 ton elephant has a gestation period a factor of 1.6 times the value predicted for this body mass (Table 3). With a litter of one, as it happens, her litter size is an equal factor (1.6) greater than expected.

With reduction in body size, an animal with an adult mass of 1 ton that bears a single offspring in 11 months of gestation will conform to the mammalian curves (Millar 1981). Alternatively, one might predict the retention of a relatively long gestation period, with an increase in the frequency of twinning. It is difficult to predict whether gestation time would or would not track body size in dwarfed elephants; examples of closely related mammals are known both to vary (e.g. species of canids, Wayne 1986, and cricetine rodents, Creighton and Strauss 1986) and to be extremely conservative (breeds of dogs, Wayne 1986) in gestation length. Even if gestation time declined with size in the way Millar's (1981) equations predict, at 100 kg (the size of *E. falconeri* of Sicily), twins become expected more frequently than singletons.

In their time to reach full size or sexual maturity, modern elephants stray far from the mammalian curve, and for both processes the time is twice that predicted on the basis of body mass. We can only speculate that if elephants conform to mammalian trends in their pattern of scaling, if not in predicted value, then the duration of growth for the Channel Island dwarfs would have been cut to between 33 and 65 per cent (depending on ultimate size) of that of their ancestors. For *E. falconeri* of Sicily, the figure is approximately 27 per cent.

The minimum time required for doubling of modern populations of elephants is approximately 15 years, assuming the maximum annual population growth rate of 4.7 per cent calculated by Eltringham (1982, p. 134) from interbirth intervals and ages at first reproduction known for modern African elephants. If population doubling time scales with body mass to an exponent between 0.26 and 0.36 (Calder 1984, pp. 308–9), we infer a corresponding 23- to 27-year minimum doubling time for 15 ton animals, 21–23 years for 10 tons, 5.7–6.2 years for 200 kg, and 4.4–6.2

years for 100 kg animals. The time required for dwarfed populations to recover from episodes of high mortality is accordingly greatly reduced.

3.2 Shape as evidence for mechanical function and heterochronic evolution

What are the morphological consequences of dwarfism in a graviportal animal? The skeletons of full-sized elephants carry enormous body weights, and their morphological adaptations reflect this. I will summarize conclusions below; except where other citations are provided, the data and analyses on which they are based are presented in Roth (1982) in preliminary form, and will be given elsewhere in more detail (Roth in preparation).

Sondaar (1977) has noted relatively shortened distal segments in the limbs of fossil dwarfed goats and hippos, as well as elephants, from Mediterranean islands. Together with fusion between the radius and ulna and between tibia and fibula (Ambrosetti 1968), which limits pronation and supination, this feature suggests a 'low-gear locomotion' well suited to the rugged, uneven terrain separating patches of grazing area on mountainous islands (Sondaar 1977).

The shaft of the femur in large elephants twists along its long axis in such a way that the feet are placed more centrally beneath the weight of the body they support; as Ambrosetti (1968) observed, however, this twist of 50° in large elephants is reduced to 30° in the tiny Sicilian dwarf *E. falconeri*.

Limb posture in large elephants is columnar. As a consequence, the mass of the animal is supported less by muscular force than by the compressive strength of bones whose centers of mass are aligned. Several pieces of anatomical evidence suggest, however, that limb posture in dwarfed elephants was more flexed. Skeletally mature (fully fused) femoral specimens of *M. exilis*, the dwarf from the California Islands, differ from those of *M. columbi*, its huge next-of-kin from the mainland, in the orientation of the distal articulation. In full-sized animals, the main axis of the femur in antero-medial view is shifted caudally on its condyles, which articulate with the tibia in a relatively extended position. In the dwarfs (as well as young specimens of *E. maximus*) the distal end of the shaft rests more anteriorly on the condyles. Femora of both *M. exilis* and *E. falconeri* exhibit a more bowed posterior surface, and a less well-developed patellar groove than larger animals, all of which suggest that they were maintained in a more horizontal posture, transmitting weight through an archlike structure, rather than through the pillarlike femur of a graviportal giant. Similarly, it appears from the orientation of the humeral condyles on the shaft of *E. falconeri* that the elbow must have been held relatively flexed in position. Small elephants, like small-

medium-sized mammals generally (Gregory 1929), show greater angulation of the limbs at the knee and elbow joints, a flexed posture associated with generally friskier locomotion and the rapid accelerations and decelerations involved in jumping or sudden changes in direction (Biewener 1989). For large adult elephants, by contrast, leaping is high-risk behavior. Modern elephants move deliberately in a walk or amble, with virtually no vertical displacement of their center of mass, and at least one foot in contact with the ground at every phase of the locomotor cycle (Gambaryan 1974).

The dimensions of single limb bones could in theory vary according to any of several patterns: geometric similarity (maintaining similar shapes at all sizes), ontogenetic scaling (adult dwarfs having proportions like those of young individuals of the larger forms), or modified scaling associated with some clearly identified genetic syndromes. One such syndrome that may show Mendelian inheritance as an autosomal dominant, and which is associated with dwarfing in a wide variety of mammalian species, is achondroplasia. In chondrodystrophic anomalies of the skeleton like achondroplasia, elongation of the diaphyses of bones stops early, but periosteal deposition continues. The result is a dwarf with short but relatively thick limb bones and exaggerated bony prominences on the skull (Rimoin *et al.* 1970), features that should be readily identifiable in fossils (Prothero & Sereno 1982), but are not seen in dwarf elephants (Roth *in press*). Though the limb bones of insular dwarfed elephants are thick in relation to those of other mammals of the same stature, the same is true of the bones of juvenile elephants, which are geometrically similar to those of larger elephants (Roth 1990, and *in preparation*). There is no evidence from skull morphology that insular dwarfs were achondroplastic, though chondrodystrophies probably do occur in elephants (personal observation from photograph of New Delhi zoo animal courtesy of I. Peterson and S. Rachootin).

The shapes of postcranial elements of dwarfed elephants are consistent with a truncation in growth. In most respects the shapes of scapula, ulna, humerus, tibia, and femur are intermediate between those of the juvenile and full-grown elephants from the mainland. That is, on bivariate plots of various measurements against lengths of diaphyses, adult *M. exilis* fall in a band where one also finds *M. columbi* of similar size but a younger stage of growth. Moreover, *E. maximus* and *E. falconeri* often, though with lower frequency, fall within the same swath of variation. Large and small, young and old elephants appear to occupy positions on a single continuum of sizes and shapes (although note that even though the dimensions of the bones within a species follow a particular pattern with the length of the diaphysis, it remains possible that relationships between other particular pairs of measurements show a different pattern of scaling with size or with ontogenetic stage). In addition, comparisons of overall

shape of individual bones of equivalent size or ontogenetic stage in living species reveal extraordinary variability. If, at each size or stage, the ancestral population exhibited intraspecific variation that was broad enough to encompass a wide range of body proportions, almost any morphology expressed in a dwarf then becomes consistent with paedomorphic size reduction. For plots in which length of diaphysis is used as a baseline, ontogenetic and geometric scaling cannot be distinguished. This observation may arise in part from simple heterochrony (probably progenesis; Alberch *et al.* 1979) and growth that is isometric. However, the same observation could also arise, in part, from a more complicated modification of non-isometric growth, superimposed on high variability. Variability characterizes even the earliest stages of ontogeny (bony parts of mandibles from single neonatal individuals, for example, are markedly asymmetrical) and the dentition (Roth 1992), which in other mammals tends to be highly canalized and species-specific in its variation (Yablokov 1974). Sexual dimorphism is one source of variation; augmenting any genetic influence on phenotypic variation among individual elephants is a period of growth spanning an enormous time and extending over a considerable range of size. In any event, the fact that individual bones of dwarfed elephants retained shapes that were often found among juveniles of its ancestors suggests some functional (locomotor) similarities between the dwarfs and juveniles.

For Asian and African elephants, the relationships among lengths of different limb bones, unlike the shapes of bones individually, exhibit highly consistent trends with size (Roth 1984). Bones that are last to complete fusion at all epiphyses tend also to show the greatest total growth increments, and proportionately the greatest increases in length relative to other bones. Proximal segments (scapula, femur) grow faster than more distal ones. Some adult dwarfed mammoths show the intermembral proportions of a juvenile elephant of equivalent size, but other individuals appear to show proportions more typical of extremely large adults (Roth 1984). Such individual variability in the limb proportions of the dwarfs, when contrasted with the uniformity of intermembral proportions across several genera of full-sized elephants, suggests developmental instability resulting from frequent influx of genes from the mainland, relaxed or rapidly changing selection pressures, or simply imperfect canalization (Roth in press).

The development of pneumatized bone in the skulls of dwarfed elephants is considerably less than in large fully grown elephants (Sondaar 1977; personal observation), but the skull morphology is not fully paedomorphic – muscle scars and frontal bossing signal fully grown animals. Well-developed tusks are another non-juvenile characteristic of adult dwarfed elephants, as are the large and elongated cheek teeth. Relatively large teeth are commonly observed in dwarf mammals (Gould 1975), and may

result from any of several mechanisms of size reduction, including stunting. Relatively large teeth are also functionally advantageous to the extent that they wear less rapidly than teeth that are scaled more proportionately. Another way the teeth of dwarfed elephants are unlike scaled-down replicas of giants' is their fewer (and consequently, proportionately thicker) lamellae, which Maglio (1973) interpreted as a functional necessity: if the thickness of enamel is reduced too greatly, the durability of its shearing edges diminishes. The unusually prolonged ontogeny and physical plasticity that characterizes elephant dentitions permits the teeth physically to adjust, both ontogenetically and phylogenetically, to a wide range of jaw morphologies, allowing elephants to adapt more readily to a tremendously wide range of body sizes than dentally more typical mammals (Roth 1989).

3.3 Evolutionary patterns and processes

From the material evidence of dwarfed elephant fossils, what can one infer about evolutionary processes?

Size reduction in insular populations of elephants probably began with stunting; indeed, the high frequency of partial lamellae in teeth of *M. exilis* (Roth 1989) may be evidence that the island populations were stressed, as dental anomalies of this type have been observed with similar, relatively high frequencies in the most stressed populations of African elephants (Laws *et al.* 1975). As discussed above, the skeletal proportions of the dwarfs are also consistent with a hypothesis of stunting. However, it strains belief to imagine that a size reduction of the magnitude exhibited by insular descendants of *E. antiquus* in its transition to *E. falconeri* could be produced simply by a reduction in food supply without a shift in the genetic composition of the population. Selection pressures for small body size were discussed in the first section of this paper, and as the size reduction appears morphologically in many respects to be a manifestation of truncated growth, it may represent a genetic assimilation of stunting (Bonner 1968; Waddington 1975; Heaney 1978; Marshall and Corruccini 1978).

A quantitative-genetic model of natural selection for dwarfism that makes use of allometric relationships between metabolic rate and fitness in insular dwarfing will be presented elsewhere (Roth and Mercer MS). Despite the relative rapidity with which dwarfing is believed to have occurred on islands, rates of dwarfing predicted by this model are consistent with information currently available from the fossil records of insular deer and elephants.

Some dwarfing syndromes such as achondroplasia are produced by single-allele substitutions, and some workers have suggested that once such alleles arose they could become fixed rapidly in the small founding

population of an island. However, dwarfing has been too frequent, and the resultant morphologies too varied, to suggest a simple genetic basis for insular dwarfism. As indicated above, morphological evidence argues against the specific hypothesis of achondroplasia.

Genetic variation is rapidly lost through drift in small populations, and for selection to have a noticeable effect it must be relatively strong, or the population must be sufficiently large. Franklin (1980) and subsequently Frankel and Soulé (1981) argued that a minimum effective population size of approximately 500 is necessary to preserve adequate genetic variation for drift to be negligible in relation to selective forces of measurable intensities. Using estimates of population density and composition (sex ratio and proportion of juveniles), and the areas of the islands, one can determine whether populations of Pleistocene elephants are likely to have met this requirement.

Sukumar (1989) tabulated proportions of adult males and females in six elephant populations which I used for calculations of effective population size (N_e). The total populations corresponding to $N_e = 500$ range from 829 to 2137 individuals, and are conservative estimates given that male elephants reach social maturity a decade or so later than sexual maturity, generations overlap extensively, and populations are unlikely to be truly panmictic (Moss 1983). (Moreover, at least some dwarfed elephants were sexually dimorphic (Ambrosetti 1968), which suggests that they may have retained some features of the mating system of their larger kin.) The average density taken from 15 studies of African elephants surveyed by Eltringham (1982, p. 87) is 1.7 individuals/km², which indicates that the area required would be 488–1257 km². Taking 3 tons as the mass of an adult African elephant, and assuming population density scales with body mass to an exponent of -0.75 (Damuth 1981), we can calculate corresponding areas for elephants of other body sizes (Table 4). On Fig. 2, two white parallel lines of positive slope delimit the minimum island areas calculated to support $N_e = 500$ (the upper line corresponding to $N = 2137$, the lower to $N = 829$).

Figure 2 also shows predictions from a model of populational persistence. In demographic terms, the viability of a population depends on its growth, which depends on the body size of its members, and is affected by environmental variation (Goodman 1987). Belovsky (1987) used Goodman's model to predict the number of individuals required in a population for its probability of persistence for a specified time interval to exceed 95 per cent. Using mass-scaled population densities (e.g. Table 4), I calculated from Belovsky's equations the area needed to support a population for 1000 years. Belovsky also gave the solution for 100 years, but this period is very short in relation to both the generation time of elephants (30 years) and the time spans (multiples of 10⁴–10⁵ years) over which fossil elephants on islands have been dated. Gingerich (1983) has

Table 4
Minimum areas estimated to support viable populations of elephants of various body sizes

Body mass	Mass-scaled density (no./km ²)	Estimated area (km ²) required for $N_c = 500$	Estimated areas (km ²) required for 1000 year persistence, with environmental variance	
			Low	High
100 kg	22	38– 98	140	3600
200 kg	13	64– 160	180	4700
1 ton	3.8	220– 560	310	9000
2 tons	2.3	360– 930	390	12 000
10 tons	0.69	1200–3100	690	22 000
15 tons	0.51	1600–4200	800	25 000

assembled empirical data on rates of morphological evolution, and from these I calculate an expected net change in mean body mass for 1000 years to be less than or equal to 10 per cent. Secular changes in the environment and evolutionary changes in the population would make extrapolation of the model much beyond 1000 years questionable. As Fig. 2 shows, the population size or area required by the demographic model is extremely sensitive to the level of environmental variation, and to a lesser extent to body size. (Note that for both the demographic and the genetic-variance models, the minimum areas required for the minimum-sized population are truly minimum: not all parts of any geographic region are necessarily habitable for elephants, especially where terrain is rugged, and the estimates allow for no seasonal movement of the animals.)

Considering the sizes of various islands in relation to Fig. 2, one notes that modern Sri Lanka has an area over 66 000 km², which would appear to be more than adequate for long-term persistence and for maintenance of genetic variation in elephants of any body size. Indeed, the *E. maximus* populations currently inhabiting this island appear as vigorous as any in modern times (making allowance for human impact, which provides the greatest threat to their survival), and the elephants show no signs of being dwarfed. It is probably no coincidence that these full-sized island inhabitants are part of a more fully balanced fauna, which includes a selective agent for large body size in the form of leopards (Eisenberg and McKay 1970).

Sicily has at present an area (approximately 26 000 km²) six times that minimally needed to maintain genetic variation in populations of 15 ton

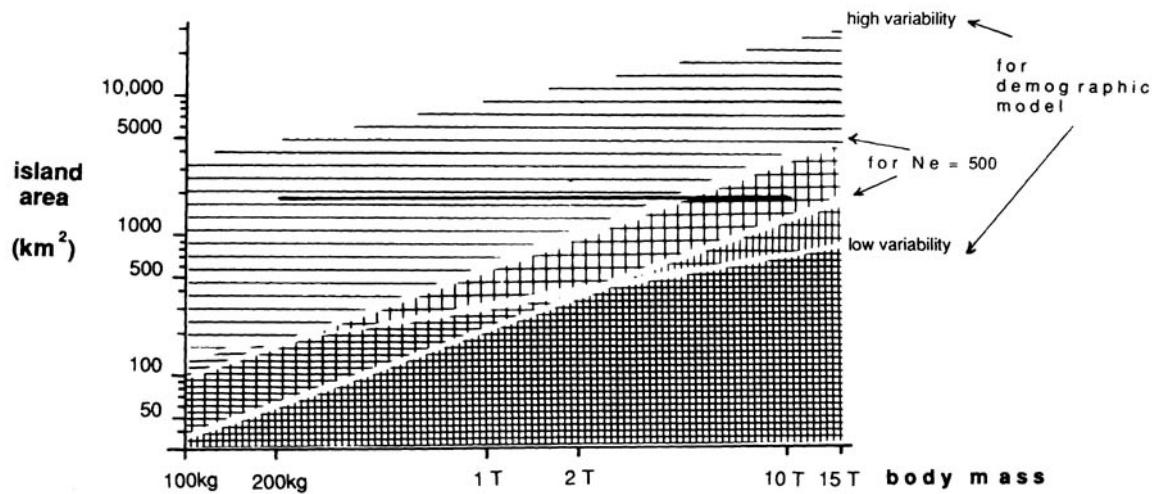


Fig. 2. Persistence of elephant populations as a function of island area and body mass (both logarithmically scaled). On this graph, vulnerability to extinction increases with increasing density of hatching. Two models of population viability, solved at two values each, are represented here. (Vulnerability is likely to be a continuous function of body mass and island area, but for simplicity, differing vulnerabilities are shown at two thresholds.) **Horizontal hatching** reflects vulnerability to extinction according to Belovsky's (1987) and Goodman's (1987) demographic model, the two thresholds representing 95 per cent probability of persistence for 1000 years in environments of, respectively, high and low variability. **Vertical hatching** reflects vulnerability to extinction as a consequence of fixation of deleterious alleles through genetic drift in populations of $N_e < 500$ (Franklin 1980). The two levels shown for this model reflect different population structures that have been observed for elephants. The horizontal line segment in the middle of this graph spans the body sizes estimated for the *Mammuthus columbi*-*M. exilis* lineages, and is set at the area (1800 km^2) of the Northern California Channel Islands at lowest sea stands during the late Pleistocene. Note that dwarfed mammoths (estimated from fossil evidence at between 200 kg and 2 tons) might be expected to persist on a land mass of this area in all but the most variable environmental conditions; full-sized mammoths would be more vulnerable to extinction both for demographic reasons and as a consequence of depletion of genetic variation.

giants (Table 4), but no large mammalian predators that could provide selection pressure for maintaining large body size are evident in its fossil record. According to Belovsky's model, its area is also sufficient for large elephants to persist even in the face of high environmental variability.

By contrast, the largest of the northern California Channel Islands now is less than 250 km². In total area they are just over 500 km² (Power 1980), but at the lowest Pleistocene sea stands (~120 m; Vedder and Howell 1980) they were united into a single land mass whose area I calculate from Vedder and Howell's (1980) and Orr's (1968) maps to be approximately 1800 km². This island area is marked on Fig. 2 as a horizontal bar toward the middle of the graph. It thus appears unlikely that the California islands could have supported genetically and demographically viable populations of full-sized (10 ton) *M. columbi* for long during even the lowest Pleistocene sea stands without periodic influx of immigrants from the mainland. Introgression would have retarded evolutionary size reduction, but by contributing genetic variation at loci other than those affecting size, could have reduced the likelihood of fixation of deleterious alleles, delaying extinction, and (paradoxically) enhancing opportunities for evolutionary response to selection for small size. Repeated invasion of the California Islands from the mainland is consistent with the fossil record: large individuals are known to have occurred both before and after the smallest forms (Orr 1968), but no dwarfs have been reported from the mainland. (Emigration from the islands may have been difficult, either because currents in the channel prevented reciprocal dispersal from island to mainland, because separation between islands and the mainland was greatest at the times the smallest forms arose, because small elephants are less seaworthy than large ones, or all three.) At times of higher sea level, when the islands were reduced in area and separated by greater distances from the mainland, populations of elephants larger than 2 tons may have been too small to have remained viable for long enough to leave an appreciable fossil record. Although greater isolation from the mainland reduces the influx of genetic variation, it would at the same time permit the island population to diverge more rapidly in response to selection on body size; smaller body size, in turn, permits higher population densities, greater total population size, and reduced likelihood of extinction. One envisions a scenario in which size reduction resulting from selection for dwarfism was initiated on a large-bodied population inhabiting large islands at low sea-stand, and accelerated, with rising sea levels, as distance from the mainland increased and mean body size declined.

These calculations, however approximate and sensitive to assumptions, suggest some qualitative predictions. Islands that are both small and inaccessible are unlikely to support elephants of large body size for very long, and the populations that do colonize them are unlikely to persist

long enough for selection for small body size to take effect. San Clemente and Santa Catalina Islands, off Southern California further south, are cases in point. Both lack the extensive fossil deposits of mammoths characteristic of Santa Rosa and San Miguel, although each is today similar in area to Santa Rosa and much larger than San Miguel. Bathymetric profiles of the surrounding channel indicate that neither the areas of these two islands nor their distance from the mainland would have been greatly affected by a 120 m lowering of sea level, and with their present areas (each less than 200 km²) they could not support reasonable-sized populations of elephants any larger than the smallest *M. exilis* (Fig. 2). If elephants were ever able to reach them in the first place, these islands may still have been too small to support adequate populations of large elephants, and too inaccessible to permit the minimal immigration from the mainland required to maintain genetic variability and responsiveness to selection. However, dwarfed elephants *could* be expected to evolve on equally remote islands the size of Sicily. Though now relatively small, San Miguel and Santa Rosa Islands may also have been well poised for evolution of dwarfed populations because of their changing areas and relative accessibility from the mainland. In the Mediterranean and elsewhere, islands with fossil elephants are sufficiently numerous that as their fossil records and geographic history (changing areas and distances from the mainland with fluctuations in sea level) become better known, these predictions can be tested.

4. CONCLUSIONS

Elephant populations confined to islands are restricted in their migrations, and face limitations in food resources which (in the absence of countervailing selection by large mammalian predators) select for individuals that reach reproductive maturity on a low plane of nutrition, i.e. at small body size. Correlated shifts in a vast suite of other characteristics, including body proportions, generation time, etc., could be expected to accompany size reduction. However, selection for maturity at small size (as opposed to selection for some other character) is likely to have been primary, for several reasons. These include (1) the observation that body sizes among the insular forms are consistently small, but morphologies are diverse; (2) the experience of livestock breeders (Bichard 1968; Taylor 1968), which suggests that mature body size is more responsive to selection than are many features of relative growth or body composition; and (3) calculations (described here) that show the reduction in food required by the dwarfs to be impressive, but the decrease in time to maturation still to encompass several seasonal cycles. Even so, the demographic resilience of populations of small body size probably has had an influence on what

we observe in the fossil record. The time period and number of islands available for colonization by elephants was sufficient for differential extinction of populations – group selection – plausibly to play some role. Populations unable to respond sufficiently quickly to selection for small body size would be too ephemeral to leave an appreciable fossil record. In morphology, to the extent that regular patterns can be deduced, the body proportions of dwarfs suggest a role for heterochrony and, in particular, progenesis – a shift to attainment of sexual maturity at a juvenile morphology. From the standpoint of individual fitness, however, the most important consequence would *not* have been the shift to juvenile morphology, but rather a shift to reproductive maturity at a juvenile size.

Careful reconstruction of morphology from anatomical details of the skeleton has been the basis for inferring the functional morphology and natural history of extinct animals since the time of Cuvier. As our understanding of the regularities of scaling relationships has expanded, so have our capabilities of inferring details of biology that do not leave their imprints directly on the bones. Elephants from Pleistocene islands are especially conducive to allometric analysis because they span an enormous range of sizes. They are also of special interest because their repeated occurrences, on the same and on separate islands, provide replicates which allow us to test predictions with samples larger than one.

ACKNOWLEDGMENTS

I thank J. Mercer for discussion and J. Antonovics and A. Lister for constructive comments on the paper. A manuscript first drafted some years ago by A. K. Behrensmeyer and L. G. Marshall provided some important leads and reference citations. Work described here was supported in part from grants from the (US) National Science Foundation and Duke University Research Council.

REFERENCES

- Alberch, P., Gould, S. J., Oster, G. F. and Wake, D. B. (1979). Size and shape in ontogeny and phylogeny. *Paleobiology* **5**, 296–317.
- Alexander, R. McN., Jayes, A. S., Maloiy, G. M. O. and Wathuta, E. M. (1979). Allometry of the limb bones of mammals from shrews (*Sorex*) to elephant (*Loxodonta*). *J. Zool. (Lond.)* **189**, 305–14.
- Ambrosetti, P. (1968). The Pleistocene dwarf elephants of Spinagallo (Siracusa, south-eastern Sicily). *Geologica Romana* **7**, 277–398.
- Anderson, J. F., Hall-Martin, A. and Russell, D. A. (1985). Long-bone circumference and weight in mammals, birds and dinosaurs. *J. Zool. (Lond.)* **207**, 53–61.

- Angerbjörn, A. (1985). The evolution of body size in mammals on islands: Some comments. *Am. Nat.* **125**, 304–9.
- Belovsky, G. E. (1987). Extinction models and mammalian persistence. In *Viable populations for conservation* (ed. M. E. Soulé), pp. 35–57. Cambridge University Press, Cambridge.
- van den Bergh, G. D., Sondaar, P. Y., de Vos, J. and Aziz, F. (in press). The proboscideans of the Southeast Asian Islands. In *The Proboscidea: Trends in evolution and paleoecology* (ed. J. Shoshani and P. Tassy). Oxford University Press.
- Bichard, M. (1968). Genetic aspects of growth and development in the pig. In *Growth and development of mammals* (ed. G. A. Lodge and G. E. Lamming), pp. 309–25. Butterworths, London.
- Biewener, A. A. (1989). Scaling body support in mammals: limb posture and muscle mechanics. *Science* **245**, 45–8.
- Bluweiss, L., Fox, H., Kudzma, V., Nakashima, D., Peters, R. and Sams, S. (1978). Relationships between body size and some life history parameters. *Oecologia (Berlin)* **37**, 257–72.
- Bonner, J. T. B. (1968). Size change in development and evolution. *J. Paleontol.* **42**, 1–15.
- Boyce, M. S. (1984). Restitution of r- and k-selection as a model of density-dependent natural selection. *Ann. Rev. Ecol. Syst.* **15**, 427–47.
- Brody, S. (1945). *Bioenergetics and growth, with special reference to the efficiency complex in domestic animals*. Reinhold Publ. Corp., New York.
- Brown, J. H. (1975). Geographical ecology of desert rodents. In *The ecology and evolution of communities* (ed. M. L. Cody and J. M. Diamond), pp. 522–36. Belknap Press, Cambridge, MA.
- (1986). Two decades of interaction between the MacArthur-Wilson model and the complexities of mammalian distributions. *Biol. J. Linn. Soc.* **28**(1 & 2), 231–51.
- Burgio, E. and Cani, M. (1988). Considerazioni sulla successione filogenetica degli elefanti fossili Siciliani. In *Abstracts of the international conference on early man in island environments (Oliena, Sardinia, 25 Sept.–2 Oct. 1988)*. Comitato Corbeddu, Nuoro.
- Busk, G. (1867). Description of the remains of three extinct species of elephant, collected by Capt. Spratt, C.B., R.N., in the ossiferous cavern of Zebbug, in the island of Malta. *Trans. Roy. Soc. London* **6**, 227–306.
- Calder, W. A. (1984). *Size, function, and life history*. Harvard University Press, Cambridge, MA.
- Caloi, L., Kotsakis, T., Palombo, M. R. and Petronio, C. (1989). The dwarf elephants of Mediterranean islands. In *Abstracts of papers and posters, Fifth International Theriological Congress, Rome*, pp. 141–2.
- Kotsakis, T., Palombo, M. R. and Petronio, C. (in press). The Pleistocene dwarf elephants of Mediterranean islands. In *The Proboscidea: Trends in evolution and paleoecology* (ed. J. Shoshani and P. Tassy). Oxford University Press.
- Carlquist, S. (1974). *Island biology*. Columbia University Press, New York.
- Case, T. J. (1978). A general explanation for insular body size trends in terrestrial vertebrates. *Ecology* **59**, 1–18.

- Clutton-Brock, T. H. and Harvey, P. H. (1983). The functional significance of variation in body size among mammals. In *Advances in the study of mammalian behavior* (ed. J. F. Eisenberg and D. G. Kleiman), pp. 632–63. Spec. Publ. Amer. Soc. Mammal. No. 7.
- Clutton-Brock, T. H., Guinness, F. E. and Albon, S. D. (1982). *Red Deer*. University of Chicago Press, Chicago.
- Clutton-Brock, T. H., Albon, S. D. and Guinness, F. E. (1985). Parental investment and sex difference in juvenile mortality in birds and mammals. *Nature (Lond.)* **313**, 131–3.
- Creighton, G. K. and Strauss, R. E. (1986). Comparative patterns of growth and development in cricetine rodents and the evolution of ontogeny. *Evolution* **40**, 94–106.
- Damuth, J. (1981). Population density and body size in mammals. *Nature* **290**, 699–700.
- Demment, M. W. and Van Soest, P. J. (1985). A nutritional explanation for body-size patterns of ruminant and nonruminant herbivores. *Am. Nat.* **125**, 641–72.
- Dorst, J. and Dandelot, P. (1969). *A field guide to the larger mammals of Africa*. Houghton Mifflin Co., Boston.
- Eisenberg, J. F. and McKay, G. M. (1970). An annotated checklist of the recent mammals of Ceylon with keys to the species. *Ceylon J. Sci. (Bio. Sci.)* **8**(2), 69–99.
- Eisentraut, M. and Böhme, W. (1989). Gibt es zwei Elefantenarten in Afrika? *Zeitschr. Köln. Zoo.* **32**, 61–8.
- Eltringham, S. K. (1982). *Elephants*. Blandford Press, Poole, UK.
- Foster, J. B. (1964). The evolution of mammals on islands. *Nature* **202**, 234–5.
- Frankel, O. H. and Soulé, M. E. (1981). *Conservation and evolution*. Cambridge University Press, Cambridge.
- Franklin, I. R. (1980). Evolutionary change in small populations. In *Conservation biology: An evolutionary-ecological perspective* (ed. M. E. Soulé and B. A. Wilcox), pp. 135–49. Sinauer Assocs., Inc., Sunderland, MA.
- Gambaryan, P. P. (1974). *How mammals run: Anatomical adaptations*. John Wiley and Sons, New York.
- Geist, V. (1971). *Mountain sheep*. University of Chicago Press, Chicago.
- (1978). *Life strategies, human evolution, and environmental design*. Springer-Verlag, New York.
- Gingerich, P. D. (1983). Rates of evolution: effects of time and temporal scaling. *Science* **222**, 159–61.
- Goodman, D. (1987). The demography of chance extinction. In *Viable populations for conservation* (ed. M. E. Soulé), pp. 11–34. Cambridge University Press, Cambridge.
- Gould, S. J. (1966). Allometry and size in ontogeny and phylogeny. *Biol. Rev.* **41**, 587–640.
- (1971). Geometric similarity in allometric growth: a contribution to the problem of scaling in the evolution of size. *Am. Nat.* **105**, 113–36.
- (1975). On the scaling of tooth size in mammals. *Amer. Zoo.* **15**, 351–62.
- Graham, R. W. (1976). Pleistocene and Holocene mammals, taphonomy, and

- paleoecology of the Friesenhahn Cave local fauna, Bexar County, Texas. PhD dissertation, University of Texas, Austin.
- Grant, P. R. (1965). The adaptive significance of some size trends in island birds. *Evolution* **19**, 355–67.
- Greenwell, J. R. (1990). New evidence supports existence of pygmy elephant. *The ISC Newsletter* **9**(1), 1–6.
- Gregory, W. K. (1929). Mechanics of locomotion in the evolution of limb structure as bearing on the form and habits of the titanotheres and the related odd-toed ungulates. In *The Titanotheres of Ancient Wyoming, Dakota, and Nebraska*, pp. 727–56. US Government Printing Office Monograph 55, Washington, DC.
- Guy, P. R. (1976). The feeding behaviour of elephant (*Loxodonta africana*) in the Sengwa area, Rhodesia. *S. Afr. J. Wildl. Res.* **6**, 55–63.
- Heaney, L. R. (1978). Island area and body size of insular mammals: evidence from the tri-colored squirrel (*Callosciurus prevosti*) of Southeast Asia. *Evolution* **32**, 29–44.
- Heusner, A. A. (1982). Energy metabolism and body size. I. Is the 0.75 mass exponent of Kleiber's equation a statistical artifact? *Respir. Physiol.* **48**, 1–12.
- Hooijer, D. A. (1967). Indo-Australian insular elephants. *Genetica* **38**, 143–62.
- (1975). Quaternary mammals west and east of Wallace's line. *Netherl. J. Zool.* **25**, 46–56.
- Hutchinson, G. E. (1959). Homage to Santa Rosalia or why are there so many kinds of animals? *Am. Nat.* **93**, 117–25.
- Janis, C. (1976). The evolutionary strategy of the Equidae and the origins of rumen and cecal digestion. *Evolution* **30**, 757–74.
- Jarman, P. J. (1974). The social organization of antelope in relation to their ecology. *Behaviour* **48**, 215–67.
- Johnson, D. L. (1980). Episodic vegetation stripping, soil erosion, and landscape modification in prehistoric and Recent historic time, San Miguel Island, California. In *The California Islands: Proceedings of a multidisciplinary symposium* (ed. D. M. Power), pp. 103–21. Santa Barbara Museum of Natural History, Santa Barbara.
- Kleiber, M. (1961). *The fire of life*. John Wiley, New York.
- Kurtén, B. (1968). *Pleistocene mammals of Europe*. Weidenfeld and Nicolson, London.
- Laursen, L. and Bekoff, M. (1978). *Loxodonta africana*. *Mammalian Species* **92**, 1–8.
- Lawlor, T. E. (1982). The evolution of body size in mammals: Evidence from insular populations in Mexico. *Amer. Nat.* **119**, 54–72.
- Laws, R. M. (1966). Age criteria for the African elephant *Loxodonta a. africana*. *E. Afr. Wildl. J.* **4**, 1–37.
- and Parker, I. S. C. (1968). Recent studies on elephant populations in East Africa. In *Comparative nutrition of wild animals* (ed. M. A. Crawford), pp. 319–59. Symposium of the Zoological Society of London, Number 21.
- , Parker, I. S. C. and Johnstone, R. C. B. (1975). *Elephants and their habitats. The ecology of elephants in North Bunyoro, Uganda*. Clarendon Press, Oxford.

- Lindstedt, S. L. and Boyce, M. S. (1985). Seasonality, fasting endurance, and body size in mammals. *Am. Nat.* **125**, 873–8.
- Lomolino, M. V. (1984). Immigrant selection, predation, and the distributions of *Microtus pennsylvanicus* and *Blarina brevicauda* on islands. *Amer. Nat.* **123**, 468–83.
- (1985). Body size of mammals on islands: The island rule reexamined. *Amer. Nat.* **125**, 310–16.
- MacArthur, R. H. and Wilson, E. O. (1967). *The theory of island biogeography*. Princeton University Press, Princeton, NJ.
- Maglio, V. J. (1973). Origin and evolution of the Elephantidae. *Trans. Amer. Philos. Soc.* **63**(3), 3–149.
- Maiorana, V. C. (1990). Evolutionary strategies and body size in a guild of mammals. In *Body size in mammalian paleobiology: Estimation and biological implications* (ed. J. Damuth and B. J. MacFadden), pp. 69–102. Cambridge University Press, Cambridge.
- Marshall, L. G. and Corruccini, R. S. (1978). Variability, evolutionary rates, and allometry in dwarfing lineages. *Paleobiology* **4**, 101–19.
- McCullagh, K. G. (1969). The growth and nutrition of the African elephant. II. The chemical nature of the diet. *E. Afr. Wildl. J.* **7**, 91–8.
- McMahon, T. A. (1975). Allometry and biomechanics: limb bones in adult ungulates. *Amer. Nat.* **109**, 547–63.
- McNab, B. K. (1963). Bioenergetics and the determination of home range size. *Amer. Nat.* **97**, 113–40.
- (1971). On the ecological significance of Bergmann's rule. *Ecology* **52**, 845–54.
- Merritt, J. F. and Merritt, J. M. (1978). Population ecology and energy relationships of *Clethrionomys gapperi* in a Colorado subalpine forest. *J. Mammal.* **59**, 576–98.
- Millar, J. S. (1981). 1981. Pre-partum reproduction characteristics of eutherian mammals. *Evolution* **35**, 1149–63.
- Mol, D. and Agenbroad, L. (MS). Metapodials and shoulder height of *Mammuthus columbi* compared with Eurasian *Mammuthus* species.
- Morrison-Scott, T. C. S. (1947). A revision of our knowledge of African elephants' teeth, with notes on forest and 'pygmy' elephants. *Proc. Zool. Soc. London* **117**, 505–27.
- Moss, C. J. (1983). Oestrous behaviour and female choice in the African elephant. *Behaviour* **86**, 167–96.
- Nagy, K. A. (1987). Field metabolic rate and food requirement scaling in mammals and birds. *Ecol. Mono.* **57**, 111–28.
- Noack, T. (1906). A dwarf form of the African elephant. *Annals and Mag. Nat. Hist. Ser. 7* **17**, 501–3.
- Orr, P. C. (1968). *Prehistory of Santa Rosa Island*. Santa Barbara Museum of Natural History, Santa Barbara.
- Osborn, H. F. (1942). *Proboscidea*, vol. 2. American Museum Press, New York.
- Owen-Smith, R. N. (1988). *Megaherbivores: The influence of very large body size on ecology*. Cambridge University Press, Cambridge.
- Perry, J. S. (1954). Some observations on growth and tusk weight in male and female African elephants. *Proc. Zool. Soc. London* **124**, 97–104.

- Peters, R. H. (1983). *The ecological implications of body size*. Cambridge University Press, Cambridge.
- Power, D. M. (1980). Introduction. In *The California Islands: Proceedings of a multidisciplinary symposium* (ed. D. M. Power), pp. 1–4. Santa Barbara Museum of Natural History, Santa Barbara.
- Prothero, D. and Sereno, P. C. (1982). Allometry and paleoecology of medial Miocene dwarf rhinoceoses from the Texas Gulf Coastal Plain. *Paleobiology* **8**, 16–30.
- Rimoin, D. L., Hughes, G. N., Kaufman, R. L., Rosenthal, R. E., McAlister, W. H. and Silberberg, R. (1970). Endochondral ossification in achondroplastic dwarfism. *New Engl. J. Med.* **238**, 728–35.
- Roth, V. L. (1982). Dwarf mammoths from the Santa Barbara, California Channel Islands: Size, shape, development, and evolution. PhD diss. Yale University, New Haven. University Microfilms, Ann Arbor.
- (1984). How elephants grow: heterochrony and the calibration of developmental stages in some living and fossil species. *J. Vert. Paleontol.* **4**, 126–45.
- (1989). Fabricational noise in elephant dentitions. *Paleobiology* **15**(2), 165–79.
- (1990). Insular dwarf elephants: A case study in body mass estimation and ecological inference. In *Body size in mammalian paleobiology: Estimation and biological implications* (ed. J. Damuth and B. J. MacFadden), pp. 151–79. Cambridge University Press, Cambridge.
- (1992). Quantitative variation in elephant dentitions: Implications for the delimitation of fossil species. *Paleobiology* **18**(2), 184–202.
- (in press). Dwarfism and variability in the Santa Rosa Island mammoth: An interspecific comparison of limb bone sizes and shapes in elephants. In *Proceedings of the Third California Islands Symposium* (ed. F. G. Hochberg), in press. Santa Barbara Museum of Natural History, Santa Barbara.
- and Klein, M. S. (1986). Maternal effects on body size of large insular *Peromyscus maniculatus*: Evidence from embryo transfer experiments. *J. Mammal.* **67**, 37–45.
- and Mercer, J. M. (MS) A model of natural selection for dwarfism in mammals on islands.
- Schmidt-Nielsen, K. (1984). *Scaling: Why is animal size so important?* Cambridge University Press, Cambridge.
- Shoshani, J. and Eisenberg, J. F. (1982). *Elephas maximus*. *Mamm. Sp.* **182**, 1–8.
- Skoglund, T. (1990). Density dependence in a fluctuating wild reindeer herd; maternal vs. offspring effects. *Oecologia* **84**, 442–50.
- Snow, M. H. L., Tam, P. P. L. and McLaren, A. (1981). On the control and regulation of size and morphogenesis in mammalian embryos. In *Levels of genetic control in development* (ed. S. Subtelny and U. K. Abbott), pp. 201–17. Alan R. Liss, New York.
- Sondaar, P. Y. (1977). Insularity and its effect on mammal evolution. In *Major patterns in vertebrate evolution* (ed. M. K. Hecht, P. C. Goody and B. M. Hecht), pp. 671–707. Plenum, New York.
- , Aziz, F., Bergh, G. v. d. and Husain, S. T. (1989). The insular proboscideans of S. E. Asia. In *Abstracts of papers and posters, Fifth International Theriological Congress, Rome*, pp. 161–62.

- Stanley, S. M. (1973). An explanation for Cope's Rule. *Evolution* **27**, 1–26.
- Stearns, S. C. (1983). The influence of size and phylogeny on patterns of covariation among life-history traits in the mammals. *Oikos* **41**, 173–87.
- Stock, C. and Furlong, E. L. 1928. The Pleistocene elephants of Santa Rosa Island, California. *Science* **48**, 140–1.
- Sukumar, R. (1989). *The Asian elephant: Ecology and management*. Cambridge University Press, Cambridge.
- Taylor, St. C. S. (1968). Genetic variation in growth and development of cattle. In *Growth and development of mammals* (ed. G. A. Lodge and G. E. Lamming), pp. 267–90. Butterworths, London.
- Theodorou, G. E. (1986). Pleistocene elephants from Crete (Greece). *Modern Geology* **10**, 235–42.
- Valverde, J. A. (1964). Remarques sur la structure et l'évolution des communautés de vertébrés terrestres. *La Terre et la Vie, Paris* **111**, 121–54.
- Van Valen, L. (1973). Pattern and the balance of nature. *Evol. Theory* **1**, 31–49.
- Van Valkenburgh, B. (1988). Trophic diversity in past and present guilds of large predatory mammals. *Paleobiology* **14**, 155–73.
- Vaufrey, R. (1929). Les éléphants nains des îles méditerranéennes et la question des isthmes pléistocènes. *Arch. Inst. Paleont. Humaine (Paris), Mem.* **6**, 1–220.
- Vedder, J. G. and Howell, D. G. (1980). Topographic evolution of the Southern California borderland during late Cenozoic time. In *The California Islands: Proceedings of a multidisciplinary symposium* (ed. D. M. Power), pp. 7–31. Santa Barbara Museum of Natural History, Santa Barbara.
- Vos, J. D. (in press). Pleistocene larger mammal faunas of the Indonesian archipelago and their paleo-ecological and paleogeographical implications. *Proceedings of the international conference on early man in island environments (Oliena, Sardinia, 1988)*.
- Waddington, C. H. (1975). *The evolution of an evolutionist*. Cornell University Press, Ithaca, NY.
- Wasserzug, R. J., Yang, H., Sepkoski, J. J., Jr. and Raup, D. M. (1979). The evolution of body size on islands: a computer simulation. *Amer. Nat.* **114**, 287–95.
- Wayne, R. K. (1986). Developmental constraints on limb growth in domestic and some wild canids. *J. Zoology (Lond.)* **210**, 381–99.
- Wilcox, B. A. (1980). Insular ecology and conservation. In *Conservation biology: An evolutionary-ecological perspective*. (ed. M. E. Soulé and B. A. Wilcox), pp. 95–117. Sinauer Assocs., Inc., Sunderland, MA.
- Yablokov, A. V. (1974). *Variability of mammals*. Amerind Publ. Co., New Delhi.

Morphology, the study of form and function, in modern evolutionary biology

MARVALEE H. WAKE

1. INTRODUCTION

A ‘renaissance’ in morphology was proclaimed during the last decade (Wake, D. 1982a; Liem and Wake 1985; Gans 1985) to herald the vigor with which the discipline of morphology is incorporating new techniques and theory, expanding into analysis of development, ecology, systematics, and evolution, and generating new ideas and approaches for a synthetic approach in modern biology. The ‘renaissance’ is really the product of 20 years of increasingly intensive study of morphology, the product of both new techniques and renewed interest in the biology of the organism as a whole. Natural historians, from Aristotle through the 19th century, studied the form and function of organisms through description. However, the reorientation of biological investigation at the end of the 19th century led to a search for causation. The reductionistic examination of the parts that contribute to the whole became the *modus operandi* of most biologists. Study of component parts remains important, but emphasis is returning to the whole organism, its history, and its place in its environment. Thus, Benson (1989) declared that the organism is biology’s phoenix.

Morphology now includes sub-disciplines that call themselves ‘functional morphology,’ ‘biomechanics,’ ‘ecomorphology,’ and ‘evolutionary morphology.’ Morphology provides information for developmental biology, neurobiology, physiology, genetics, paleontology, behavior, and systematics, and these fields reciprocally illuminate morphological analysis. Workers in the sub-disciplines of morphology, and of biology in general, with some notable exceptions, are really just beginning to talk to each other, so that research on protists, plants, invertebrates, and vertebrates – fossil and extant – increasingly is being considered in problem-oriented contexts. It is becoming more common to cross the boundaries of disciplines, sometimes under the label of ‘evolutionary morphology.’ ‘Evolutionary morphology’ is not a restrictive descriptive term, as are ‘particle physics’ or ‘quantum mechanics,’ but it is a general descriptor of a part of biology that uses morphological data of all sorts in an

evolutionary (historical, often phylogenetic) context to analyze a diversity of problems (Wake, M. 1991). There is no unity of concept of what evolutionary morphology includes, but its very breadth and dynamism preclude other than the most general precepts (see Wake, M. 1991). This heightened level of activity and integration includes the search for, and proposal of, new methodologies. This makes the current study of morphology exciting, dynamic, and enormously stimulating. It returns to the pantheon of science as its contributions are being recognized in terms of explanatory power and generation of new hypotheses by both biologists in the discipline and especially those in once peripheral or not even allied areas.

It remains true that morphology may not be 'evolutionary,' or even comparative, *per se*; for example, a careful physical and biological analysis of a single species may help understanding of the biomechanics of butterfly flight. The broader question of the origin of butterfly flight, however, would necessitate comparative biology, preferably in a phylogenetic context – here considered evolutionary morphology. Indeed, Funk and Brooks (1990) consider phylogenetic systematics 'the appropriate context for all studies in comparative biology.' No matter what the approach, the description or context of the problem, the applicability of the methodology, and the rigor of the analysis are what is important. Explicitness is essential. There has been a recent tendency to apply the term 'evolutionary' to morphological studies that led to results that could be used in an evolutionary context (by others). The label 'evolutionary' seems to have some cachet today, but it need not be used seemingly to validate work that does not need, and should not have, that label (Wake, M. 1990). We must recognize that many, if not most, workers develop questions and problems in steps, usually hierarchically. Functional analysis may be a final goal, or it may be a first step in a progression to comparative and then evolutionary analysis.

The contribution of morphology (and thereby morphologists) to the development of evolutionary theory has been questioned, and generally viewed as minimal at best. Ghiselin (1980) declared that morphologists failed to assimilate Darwinism, and contributed nothing to the evolutionary synthesis, because morphology is a descriptive science. Thus 'morphology tends to be the sort of discipline that will follow, rather than lead, in the development of evolutionary theory,' a conclusion also accepted by Coleman (1980). This idea was countered by Waisbren (1988), who demonstrated the contribution to the evolutionary synthesis of the morphologists of the Oxford group. E. S. Goodrich, for example, used morphology to search for evolutionary pathways. He worked from the assumptions that natural selection was the major force giving direction to evolution, and that structure was the best indicator of evolutionary 'advance.' Goodrich had a thorough understanding of variation, and used

morphological data to assess variation in fitness of organisms, though Coleman (1980) considered morphologists to be interested in similarity for identification of 'groups,' rather than differences or variation. Goodrich may have been the first to define 'genotype' and 'phenotype' (1924), and urged that the terms be adopted to avoid confusion about the concept of the 'unit character.' Goodrich had a strongly selectionist perspective, but it often was not represented in his significant contributions to morphology, such as *The Structure and Development of Vertebrates* (1930). Waisbren correctly noted that the assessment of contribution to the synthesis was dependent on which works were analyzed. Many of Goodrich's students also contributed to the synthesis. Julian Huxley (1954) recognized that biologists were trying to clarify the process of evolution, rather than its mechanism. He urged that methods of studying evolution be broadened and unified. Huxley, and a student of both Goodrich and Huxley, Gavin de Beer, explored growth gradients and heterochrony. Gould (1977) stated that de Beer's *Embryos and Evolution* (1930; later *Embryos and Ancestors* 1940) 'was the first and shortest' of a 'series of remarkable books that established the synthetic theory of evolution.'

Why, then, is morphology still considered non-contributory? Medawar and Medawar (1977) believed that comparative anatomy declined because 'the greater part of the work has already been done,' and that there is 'modern impatience with research as slow moving as comparative anatomy.' However, they state 'the fact that the study of comparative anatomy is an exacting and formally very beautiful discipline . . . almost a biological art form . . . a biologist who cannot appreciate and marvel at Edwin Goodrich's *Studies in the Structure and Development of Vertebrates* deserves sympathy.' I believe that there are two additional reasons for the lack of appreciation for morphology: the way it was (and too often still is) taught, and the continuing schism between 'experimentalists' and 'natural historians.' Maynard Smith (1991) declares 'a strong prejudice against comparative anatomy, acquired when an undergraduate . . .' because he was taught 'a course in nineteenth century anatomy' that was not 'intellectually respectable.' I think that he is, regrettably, representative of a host of biologists who are skeptical of the possibility of synthetic work in morphology even at the present, because of their undergraduate experiences. Even today, many courses in morphology are taught in ways that do not reflect current areas of research, even those of their instructors. This, too, is changing, but too slowly, and positive change may be inhibited further as positions in morphology are filled by non-morphologists, thus denying students access to modern concepts. My second point, that of a schism of some sort between experimentalists and natural historians, with morphology the province of description, natural history, and systematics, all devoid of experimentation, is a perception that is held by many biologists (and

others) today. The perception further contributes to the idea that there is nothing 'modern' in morphology. This notion is highly inaccurate, as I hope to convince you in this essay. Experimentation and statistical analysis are significant parts of the now fast-moving field of evolutionary morphology, and are contributing to a new understanding of the natural history of organisms.

Of constructive value to the practice of morphology was Gould and Lewontin's (1979) criticism of the 'adaptationist programme.' The practice of examining the traits (often morphological features) of organisms and then attempting to explain them in terms of adaptation, or some optimization of function, without consideration of any other biological processes, was indeed faulty. Such explanations were rarely testable, and alternative explanations were rarely entertained. The Gould and Lewontin paper met with many direct responses, ranging from appeals for increased breadth of technique and of analysis to thoughtful challenges to their indictment of adaptationism (see section 7.2). Maynard Smith (1991) states that 'practitioners of the adaptationist program have cleaned up their act during the last ten years.' Alternative explanations are considered, and the comparative method is used much more rigorously to test hypotheses. In fact, under current practice, knowledge of the evolutionary relationships of the species examined is required in order to test hypotheses of adaptation, and that knowledge is based on the methodology called cladistics (see section 6 for discussion). Adaptation is the subject of study for many evolutionary morphologists, but with awareness that there may be alternatives to adaptive explanations of changes in morphology. Explicit tests and rigorous statistical analysis of data provide not only new insights into structure-function relationships, but also to pattern and process of evolution.

I propose to show that much of current biological thought, with its renewed attention to morphology, is producing new ideas for evolutionary biology. I illustrate recent research in evolutionary morphology with but a few examples from biomechanics, ecomorphology, functional morphology, developmental morphology, and systematics, deliberately avoiding a number of excellent, but well known and often cited, examples. I focus on new or novel research, and studies that emphasize particular points, concentrating on the literature of the last decade. I show the application of tools 'borrowed' from different fields of biology, physics, engineering, and mathematics, and how these approaches generate new methodologies and new ideas. Several examples of research that are synthetic studies of form, function, and evolution are used to elucidate and discuss current philosophies of morphology. Finally, I demonstrate the ways that morphologists, among others, are contributing to the potential for an expanded evolutionary synthesis that recognizes some of the limitations of Darwinism, and contributes new approaches that deal

with structural, functional, epigenetic, and historical properties that allow understanding of the evolution of organisms. Most of all, I hope to convey a sense of the excitement, breadth, and dynamism of morphology as it is currently practiced.

2. FUNCTIONAL MORPHOLOGY

The domain of functional morphology is the study of the way that form, or structure, causes, permits, and even constrains organisms to function or perform. Functional morphology often is considered to include biomechanics, ecomorphology, and aspects of comparative physiology, systematics, and development (Gans 1974), but in order to indicate their distinguishing properties and the current attention being paid these areas, I shall treat them separately.

Gans (1965, 1966) and Bock and von Wahlert (1965) proclaimed a new orientation in functional morphology based on the development of new tools, better understanding of the properties of cells and tissues, and especially a new conceptual basis of the field that is a direct consequence of the implications of the theory of evolution. He commented that past limitations were due to delayed comprehension of these implications by functional morphologists, physiologists, and other experimental biologists (though naturalists, ecologists, and *morphologists* [my italics] perceived them earlier). Rigorous study of function is still rather limited. Homberger (1988) listed several handicaps for functional anatomical work, among them lack of established guidelines for selecting relevant structures for investigation, lack of methods for assessing the accuracy of description of morphology, and the need to integrate results from various approaches and often different disciplines. She advocated sequential selection of data, constructing a structural model based on the morphological description; then constructing a functional model based on the structural model that integrates information from several approaches and levels of the analysis of structure and function, and finally testing the functional model observationally or experimentally. Lauder (1990) noted that while workers often recognize the importance of data on function, many papers that present 'functional analyses' largely infer function from the study of structure. In fact, Gans (1966) asserted that function *cannot* be read from structure; many do not agree, and inference is a common practice. Relatively few studies present quantification of function in organisms and provide direct examples in which measured functional attributes are useful for understanding problems in evolutionary biology. However, these are precisely the areas in which new advances are being made – new models are being constructed, and much more precise tests of mechanisms are being performed. Herring (1988) indicated that the lack of a unified

paradigm in evolutionary morphology is due to the inherent complexity of the structure and function of organisms, and to the great increase of new experimental techniques. While new techniques allow new analyses, techniques can constrain the practice of functional morphology (and other areas of science). The limits of techniques serve to determine the questions that are asked, in many cases.

There are other reasons for some concern about current focus in evolutionary morphology. For example, much research in vertebrate functional morphology has emphasized muscle–bone functional units, with occasional attention given to the neurophysiology of these units. There are several reasons for such focus: bones are easy to interpret as simple lever systems, and the forces exerted by muscles on such systems are now easily measured. Further, bones are usually the only parts of extinct vertebrates available; consequently they permit only functional inferences. Study of neurophysiology is providing input to functional analysis in terms of control and integration properties, but it is technically limited. Finally the complexity of organisms has thus far precluded functional morphological studies that examine simultaneously all the aspects of form (plumbing, wiring, support, covering, etc.) and function (movement, integration, fluid dynamics, digestion, etc., etc.) of whole organisms. A diversity of techniques must be brought to bear on each system, and the results of studies of parts do not necessarily give an accurate or even adequate summary of the structure–function relationships of the whole organism.

Still, the increasing sophistication of studies of structure–function relationships within and among parts of organisms gives promise of more synthesis. For example, feeding and locomotion are particularly well suited to analysis of functional morphology because (1) the behaviors are usually repeatable in the lab and often measurable in the field; (2) the components involved can be separated for analysis; (3) experimental techniques can often be employed; and (4) function of fossil forms can often be inferred from that of living relatives or apparent functional analogues. Many examples exist of studies of the functional morphology of feeding and of locomotion, and I will cite only a few, because they illustrate particular points.

2.1 Case 1: Mollusc and bird flapping locomotion

The examination of the pteropod mollusc *Clione limacina* and the pigeon, *Columba livia* (Fig. 1) to determine how flapping (lift-based propulsion) as a means of locomotion is effected (Welsford *et al.* 1991) provides an example of the use of rigorous technique to ascertain pattern and process. The pigeon's wings are of the bird bone–muscle–skin–feather morphology; *Clione* is a planktonic mollusc whose reduced foot is modified as lateral

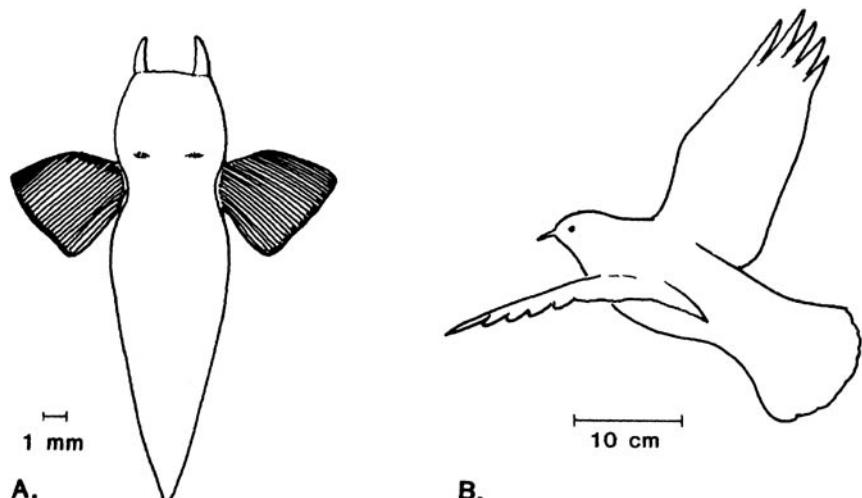


Fig. 1. 'Flapping' animals. (A) *Clione*, a pteropod mollusk (redrawn from Satterlie *et al.* 1990). (B) *Columba*, a pigeon.

wing-like structures that beat synchronously to move the animal through the water. Flapping is dependent on the dynamics of the wing and of the medium (Denny 1990), and also requires a level of neuromuscular organization sufficient to adjust wing speed in a changing environment.

Welsford *et al.* (1991) followed-up previous work on the central and peripheral organization of the mollusc wing (summarized in Satterlie *et al.* 1990), noting that it has small motoneurons organized for slow swimming and large motoneurons for fast swimming. The pectoralis, the wing adductor of the pigeon, is composed of two populations of muscle fibers that are different in size and histochemistry, suggesting that the muscle might also be organized for slow and fast flying. Indeed, electromyography shows that the large fast-glycolytic fibers are active during takeoff and landing, and the smaller, fast-oxidative fibers during constant velocity flight. Experimental techniques not previously used to analyze locomotion in birds include *in vivo* glycogen depletion and particular methods of motor unit isolation and testing, intracellular recording, and nerve filament stimulation studies (methodologies are summarized in Welsford *et al.* 1991), all successfully employed on the pigeons.

The comparison of the neuromuscular organization of two very diverse animals using the same experimental techniques is instructive of similarity of pattern; however, one must not leap to conclusions about the evolution of form and function based on data for two unrelated taxa. The two animals show strong parallels in design. Both have motor units with

contractile properties that are matched to the power output demand. Both have two fiber types, but their histochemical features differ. *Columba* maintains a nearly constant wingbeat frequency during flight maneuvers, in contrast to *Clione*, but the wingbeat frequency increases nearly 25 per cent during takeoff and landing (Dial *et al.* 1987). The authors predict that the two types of fibers in the pigeon will prove to differ significantly in their twitch contraction times. The muscle organization in *Columba* does not show distinct populations of large and small motor units, so the large motor units might be recruited when large power forces are needed. *Columba* and *Clione* have populations of both fatigable and non-fatigable motor units, although *Columba* has a larger number with 'intermediate' levels of fatigue. A series of intracellular recording experiments has not yet been able to produce recordings from a motoneuron that activates large fatigable fibers; this may be either a technical problem, or large fibers may be driven by combined activity of several small, synaptically coupled motoneurons, so that stimulation of one neuron may not excite the necessary number of neurons to activate the fiber. Work on *Columba* with reduced preparations, as is common in invertebrate neurophysiology, must be done. The goal is to develop a preparation that allows the characterization of a neural correlate to flight, similar to that obtained in *Clione* that has permitted advances in understanding of the neuronal circuitry that facilitates swimming. Such a neural correlate likewise may be necessary for understanding the circuitry that permits bird flight.

This example of functional morphology is broadly comparative, and deals with the examination of a mode of locomotion in highly divergent lineages. The study is one in evolutionary morphology, broadly defined, in that it provides a means of analyzing convergent patterns, though it does not deal with the lineage-dependent history of development of the pattern (i.e. its evolution). It also illustrates the reciprocal illumination, as well as the potential constraints, of the use of current experimental techniques in studies of functional morphology.

2.2. Case 2: Heaters in fish heads

Cell and molecular techniques allow an analysis of structure and function at the molecular level, which provides new insight into whole animal function. It is important that functional morphologists and comparative physiologists *not* avoid reductionism, but use molecular techniques to unravel questions not approachable by other experimental methods. Molecular approaches are complementary, and useful to expand the scope of organismal biologists in investigating major questions in morphology and physiology. Exploration of the biochemical and genetic changes that gave rise to a pathway provides a mechanistic explanation of the origin of the novel character. Cell and molecular techniques now allow an

understanding of the way changes in gene expression result in pathways that lead to new or altered morphologies, so that it soon may be possible to understand the mechanism of transformation from ancestral to derived states (Block 1991).

The 'heater' of certain fishes is a novel phenotype, and its origins are sought by examining how muscle has been modified to produce the organ. Block studies the way that an elevated brain temperature is maintained while the body temperature is allowed to fluctuate in species of the monophyletic assemblage of billfishes (Xiphiidae and Istiophoridae, swordfishes, etc.). The ability of fishes to elevate tissue temperatures has evolved three times in the Scombroidei (Fig. 2), and each venture has used different morphological and physiological means (literature is summarized in Block 1991). Tunas and bonitos (Thunnini of the family Scombridae) maintain elevated temperatures by conserving heat generated in highly oxidative red aerobic swimming muculature, and it is dependent

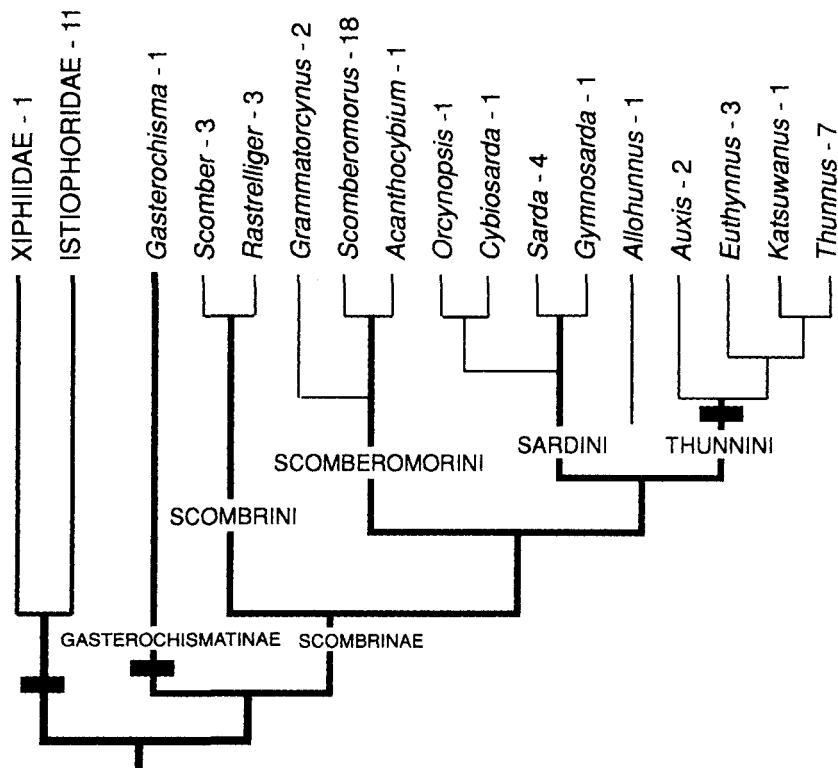


Fig. 2. Cladogram of relationships of scombroid fishes (after Block 1991). Endothermy, indicated by black bars, evolved in three different lineages, each using different morphological and physiological characters (see text for discussion).

on high densities of mitochondria, high myoglobin content, and well developed membrane systems for controlling calcium movement. Less is known of the mechanism of the second group, the subfamily Gasterochismatinae of the Scombridae, but the heating modifications appear similar to those of the Thunnini. However, billfish only generate heat in the brain and eye regions; they use a bilateral modification of the superior rectus (one of the extrinsic muscles of the eyeball) into a heat-generating tissue. Most of the muscle contains fibers that generate heat without producing tension. The carotid artery courses through the part of the muscle rich in these fibers and convectively carries heat from the blood to the retina and the brain. The carotid circulation at the base of the muscle contains a counter-current heat exchanger and retards heat loss. Among billfish, there is a correlation of extent of heater with the thermal ecology of the species; fishes in the coldest environments have the superior rectus largely modified as a heater, whereas surface species have less of the muscle mass converted to heat-generating fibers.

The study elucidates the molecular pathways for the generation of heat. Motoneuron excitation leads to calcium release, as in all muscles, and is apparently a conserved innervation pattern. Heater cells have hypertrophied cellular membrane systems and mitochondrial volumes, and have reduced numbers of contractile filaments. Release of calcium does not induce a force-generating cycle with the absence of actin and myosin; however, calcium ions keep pumping back into the membrane system and hydrolysis of ATP continues. Heat is generated as a by-product of ATP hydrolysis as long as calcium continues to leak into the cytoplasm, perhaps because of the high mitochondrial volume. Other work has demonstrated that the brain temperature remains elevated even during very deep dives lasting as long as 12 h (see Block 1991 for references). Block uses molecular data to propose further tests that could elucidate the trajectory for transformation from muscle to heater in a mechanistic way, and outlines a series of experiments to this end.

This study illustrates the kinds of molecular techniques that can be absorbed and integrated into the repertoires of comparative physiologists and functional morphologists, in order to resolve study of the origin of complex traits. Block emphasizes that organismal biologists are in a unique position because of their interest in the evolution of complex organisms. Organismal biologists are better able to grasp the tools of molecular biology because, unlike molecular biologists, they search for solutions to problems that have an integrative message. Therefore it is the responsibility and the opportunity of organismal biologists to accept useful tools from molecular biology, simultaneously introducing that field to the broad array of questions in evolutionary biology. Molecular biology thereby can contribute, via technique and interpretation, to several levels of understanding of the structure-function relationship.

3. BIOMECHANICS

The domain of biomechanics is the analysis of form–function relationships using principles of engineering and physics – principles of the physical world – to analyze work done by and in organisms, as well as the complexity of form and function. Gans (1974), Alexander (1975), Wainwright (1988), Wainwright *et al.* (1976), Vincent and Currey (1981), and Vogel (1981, 1988) have produced major texts in the field. Examples abound of the power of techniques adapted from engineering, physics, and chemistry. The goal of many workers is the characterization of the working of structural units, which is sufficient for answering some important questions. For instance, the development of prosthetic limbs is dependent on a detailed understanding of the way the human arm or leg works. If principles of biomechanics are employed in a comparative mode, using the history of development (evolution) of structure and function of organisms within a lineage, the study can elucidate the evolutionary morphology of the organism (see Liem 1989; Lauder 1990b; Rayner 1990). As noted above, problems often are delineated in hierarchical steps. Thus some work in biomechanics, begun as idiographic experimental analysis of a question, can develop into exploration of the evolution of a system and the work that it does.

3.1 Case 1: The evolution of insect wings

Kingsolver and Koehl (1985) used a modeling approach to biomechanics with the goal of sorting among competing hypotheses about the functional effects of the wings of insects. Not only was one hypothesis supported in terms of the initial increase in wing size, but the data and their analysis have significant implications for evolutionary biology, because they indicate that function changed dramatically *without* change of geometry. A simple increase in size may open different physical regimes facilitating major functional modification without drastic redesign.

Kingsolver and Koehl examined a number of aerodynamic and thermoregulatory hypotheses of adaptive factors in the evolution of wings from small segmental protrusions. Physical models of Paleozoic insects (Fig. 3) were tested in a wind tunnel to examine the potential effects of wings for increasing gliding and dispersal distance during parachuting, as well as for improving stability and elevating body temperatures. Body size and shape, wing length, number, and venation; and meteorological conditions were analyzed. Hypotheses consistent with both movable and fixed wings were tested. The models represented a range of possible forms comparable to general types of known Paleozoic insects. Aerodynamic tests used resistance-type strain gauges; lift and drag were measured with gauge and beam devices. Several different angles of attack for four to

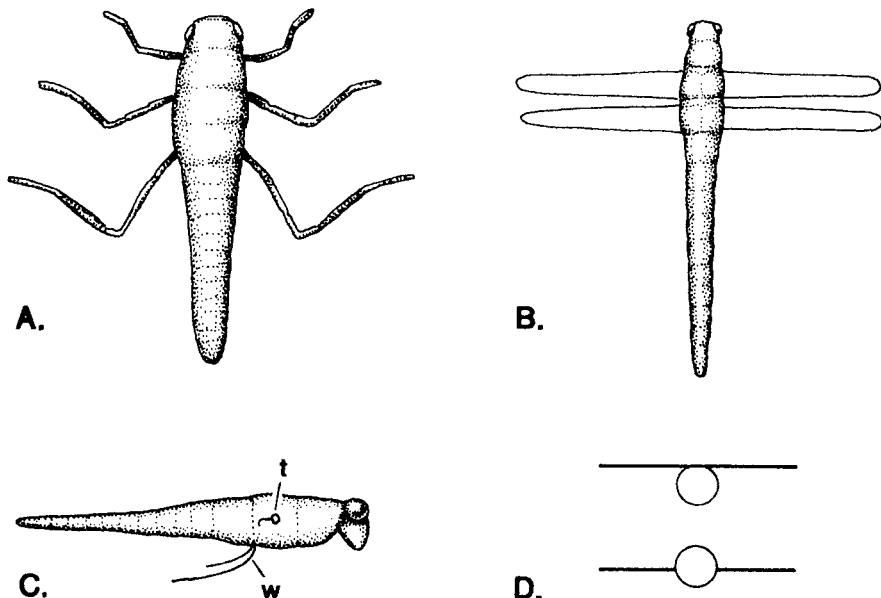


Fig. 3. Models of Paleozoic insects used in aerodynamic-thermoregulatory experiments. (A) Dorsal view of wide flattened body shape with legs. (B) Dorsal view of slim cylindrical shape, with dorsally mounted wings. (C) Side view of wide flattened body with position of thermocouple (t) and thermocouple wires (w) indicated. (D) Cross-sections of mesothorax of a slim cylindrical body showing dorsally (top diagram) and laterally (bottom) mounted wings. (Redrawn from Kingsolver and Koehl 1985.)

six different wing lengths at several wind speeds were measured. For tests of thermoregulation, the models were painted black, mounted, placed in the wind tunnel, and heated with a photoflood lamp. Air and body temperatures were recorded; steady-state body temperature was measured at specific radiation, wind speed, and air temperature for each model as a function of wing length and wing thermal conductivity (Fig. 4).

Short wings have no significant effects, but long wings do, relative to wingless models, on *aerodynamic* characteristics. However, short wings have significant *thermoregulatory* effects, but long wings do not affect this performance, relative to wingless models. There is a wing length below which there are significant *thermoregulatory* effects of increased wing length at any body size, and above which there are significant *aerodynamic* effects of increased wing length. The relative wing length for the transition decreases with increased body size. Kingsolver and Koehl conclude that there can be no effective selection for increased wing length in wingless or short-winged insects relative to increased

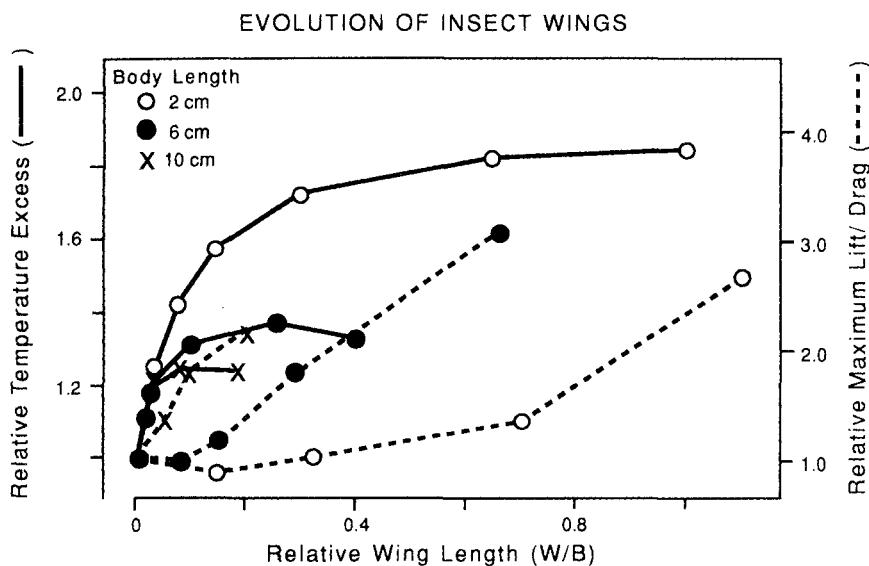


Fig. 4. Diagram of relative temperature excess (left ordinate, solid lines) and maximum lift/drag ratio (right ordinate, dashed lines) as a function of relative wing length (wing length/body length) for several body lengths of insect models. (Redrawn from Kingsolver and Koehl 1985.)

aerodynamic capacity. However, the results support the hypothesis that wings initially were thermoregulatory structures and became aerodynamic only at larger wing lengths and/or body sizes. Therefore, thermoregulation was the property favored in the initial evolution of wings, and it was a preadaptation for the evolution of flight. This is a mechanism by which an isometric change in body size may produce a qualitative change in the function of a given structure. Kingsolver and Koehl propose a new hypothesis in which the transition from thermoregulatory to aerodynamic function for wings involved only isometric changes in body size; they argue that major change in function was not conditioned by change in body form. A consequence of such differential scaling is that a particular structure may have different functional properties at different stages of its evolution. Two scenarios are proposed for the transition from thermoregulatory to aerodynamic function. One is that the transition occurred as a result of incremental small changes in wing length, after wings reached lengths of approximately 1 cm. Alternatively, an isometric increase in body size may yield a change in function from thermoregulatory to aerodynamic capacity. Their example is that of an insect with wing lengths 50 per cent of the body length. These wings could function for thermoregulation in a small insect, but as aerodynamic structures in a

large insect. As qualitative shifts in function of a structure occur, the conditions favoring that structure are likely to change as well.

Their elegant but straightforward approach has opened new avenues for the investigation of the relationship of body form to function, as well as the evolution of insect flight. It illustrates the usefulness of a modeling approach, conceived to determine parameters of mechanical function, and the influence of their conclusions on approaches to evolutionary morphology. The models were not lineage-constructed, but generalizations that allow testing of new hypotheses about the evolution of structure and function within lineages.

3.2 Case 2: The evolution of insect flight

Insect flight lends itself particularly well to biomechanical analysis because (1) insect systems are amenable to modeling, (2) insects are usually abundant, and (3) many lineages of insects include great diversity of form and function. Dudley (1991) has used a comparative biomechanical approach to explore the evolutionary diversification of flying insect morphology. He examined the aerodynamic consequences of the evolution of asynchronous flight muscle, which is found in members of three of the four largest insect orders (Hymenoptera, Diptera, and Coleoptera) and 75 per cent of all insect species. Asynchronous muscle facilitates higher wingbeat frequencies, and has permitted increased wing loading (reduction in wing area relative to body mass) and increased flight speed. The flight trajectories are more nearly independent of ambient air motion. Dudley infers that the forewings of beetles have become elaborated as protective devices, the elytra, because they are freed of their aerodynamic role, as in dipterans the hindwings have become specialized as gyroscopic halteres. However, flight behavior, wing kinematics, and flight speeds in natural conditions are unknown for nearly all insects. Therefore Dudley executed a broad comparative survey of flight biomechanics in order to understand the functional significance of patterns of morphological evolution in winged insects. Biomechanical analysis is used to evaluate the functional correlates of morphological diversity, to assess the effectiveness of existing design, and to suggest past scenarios of insect morphology.

Dudley focused his study on asynchronous flight muscle, which has appeared repeatedly in insect lineages. It is a derived state, characterized by repeated contractions of the muscle in response to a single nervous impulse. The primitive synchronous muscle has a one-to-one correspondence between nerve impulse and muscle contraction (literature summarized in Dudley 1991). However, the characterization of both types of muscle is based largely on morphology, rather than an explicit investigation of their physiological properties. Asynchronous muscle is stretch-induced, permitting rates of contraction exceeding the operating frequency of

synchronous muscle by as much as an order of magnitude. Wingbeat frequencies are elevated, those of bees and flies being 110–300 Hz. (Wings of most synchronous fliers beat below 50 Hz.)

Dudley surveyed the taxonomic distribution of asynchronous muscle, and found a strong correlation between insect diversity and the presence of asynchronous flight muscle, which appears to facilitate diversification of insects. He then investigated the aerodynamic and biomechanical consequences of high wingbeat frequencies exhibited by asynchronous fliers, and considered the ecological and evolutionary implications of high wingbeat frequencies and the concomitant reduction of wing area and increased flight speeds.

Dudley used several principles of aerodynamics in his analysis: wing loading (see above), advance ratio of the wings (ratio of the forward airspeed to the flapping velocity of the wing tip), and quasi-steady analysis, an approach that assumes that the motions of wing flapping can be reduced to a series of static conditions of steady-state flow, so that conventional aerodynamic theory can be employed to estimate resultant forces (literature summarized in Dudley 1991). The velocity of moving wings is proportional to their flapping frequency and to wingbeat amplitude, so that an increase in wingbeat frequency results in a disproportionate increase in aerodynamic forces (fig. 5A). The magnitude of the increase depends on forward airspeed, and therefore on the advance ratio. Aerodynamic forces on the wings are less dependent on wing flapping velocity at higher forward airspeeds, so the relative effect of an increase in wingbeat frequency is less pronounced (Fig. 5B). Hence, development of asynchronous muscle will have maximal aerodynamic consequences for slower-flying insects.

As the force from the flapping wings is required to offset body weight, increased wingbeat frequency permits reduction of wingbeat amplitude, decrease in muscle strain, and a decrease in wing area. Reduction in wing area increases wing loading (demonstrated empirically for many flying animals), thus increased wingbeat frequencies increase flight speed. Dudley points out that there may be a cost to increased wingbeat frequencies. Wings must be strengthened to resist deformation from inertial forces, and, in taxa other than beetles and flies, fore- and hindwings must be complexly coupled to assure beat frequency. Morphology suggests that there are elements in the flight apparatus that allow storage of elastic energy, so that it is used to reaccelerate wingbeats.

The acquisition of asynchronous muscle and subsequent modification of the wings may allow the evolution of small insects, because they typically require wingbeat frequencies greater than 100 Hz to maintain flight. Small insects (less than 10 mm in length) fly at airspeeds below 1 m/s, which are less than the likely ambient air flow. Therefore small increases in flight speed will confer some advantage, especially if flight

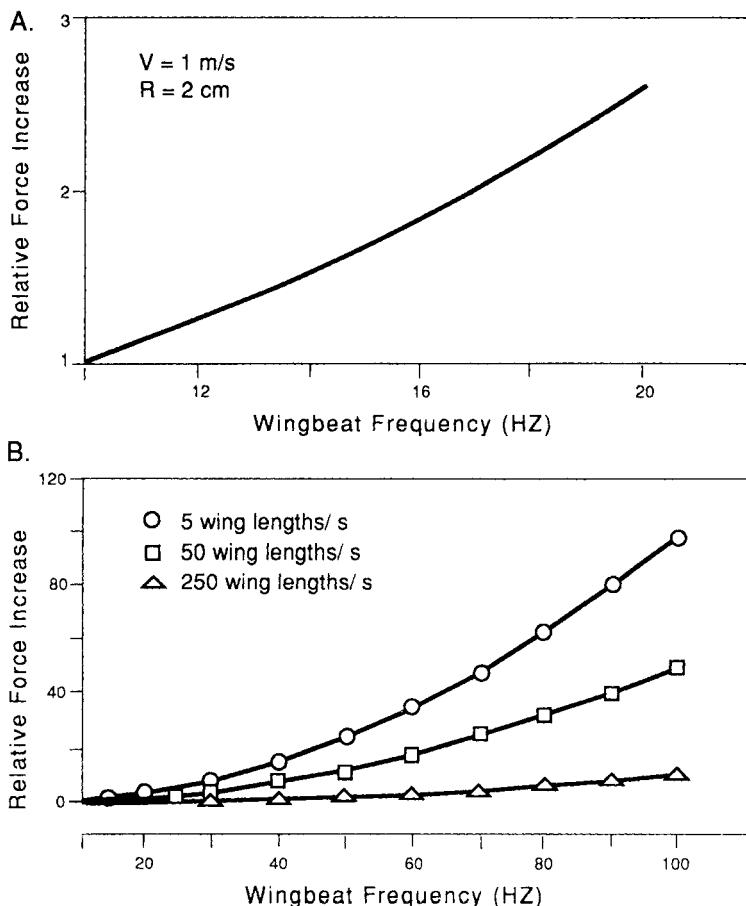


Fig. 5. Aerodynamic force increases for insects relative to wingbeat frequency. (A) Force relative to force produced at wingbeat frequency of 10 Hz, forward airspeed 1 m/s, assumed wing length 2 cm. (B) Force increase relative to the same wingbeat frequency, for relative forward speeds of 5, 50, and 250 wing lengths/s. The advance ratio at five wing lengths is 0.12, or that of hovering flight; that at 250 wing lengths is 6.0, or that of fast forward flight. From Proceedings of the Fourth International Congress of Systematic and Evolutionary Biology, Copyright 1991 by Dioscorides Press (an imprint of Timber Press, Inc.). Reprinted by permission.

is to a specific destination such as a mate or a food source, reasons Dudley. The key point is that an increase in flight speed increases the effectiveness of flight behaviors, and may have facilitated the evolution of small insects and modified wing structures in dipterans and coleopterans. More data on all parameters of insect flight patterns are needed to test these hypotheses, as is investigation in a rigorous phylogenetic framework.

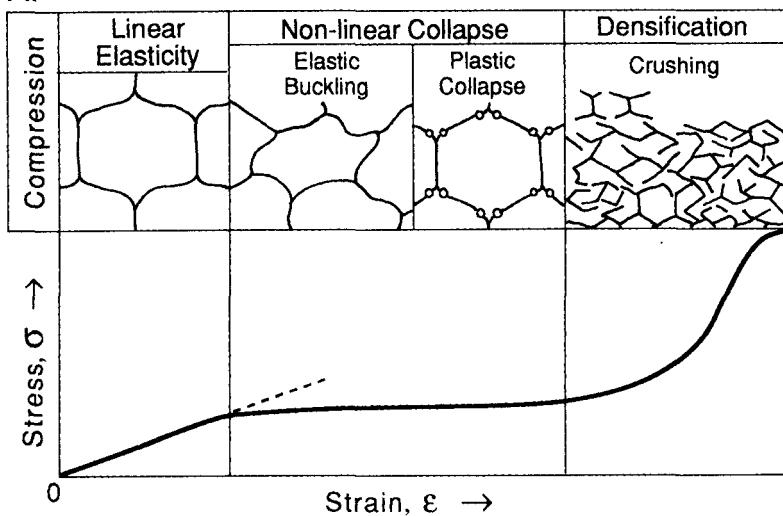
3.3 Case 3: The evolution of multicellularity in plants

Workers on animal organisms traditionally pay little attention to the contributions to studies of biomechanics of researchers who examine plant structure and function (the converse seems less true). Niklas (1989, 1990) has used biomechanical principles to examine the behavior of tissues in recent and fossil plants, and to suggest patterns in vascular plant evolution. Niklas and Kaplan (1991) have explored the mechanical effects of multicellularity in plants and the advantages conferred by partitioning the protoplast into cells.

Niklas and Kaplan (1991) have reviewed the mechanical attributes of plant cells and tissues, and found that the partitioning of the protoplast by cell walls confers structural, physiological, and reproductive advantages. Compartmentalization of the protoplast by cell walls or their analogues resulted in stiffer plant bodies, and was elaborated during plant evolution. The multicellular plant body is a very specialized mode of development in which cytokinesis and nuclear division are correlated precisely. The protoplast of the plant body is a single indivisible feature, *contra* the tenets of the cell theory. Niklas and Kaplan therefore reject the ideas that the plant body is fundamentally divided into cells of initially equal developmental and morphological rank, that differentiation and morphogenesis involve competitive interactions among cell lineages, and that the individuality of the organism results from competitive interactions. Their basis for argument is that most plant cells have a cell wall with a geometry and chemical composition that dictate the mechanical attributes of cells and the texture of tissues. The cell wall is flexible during its development, rigid and stiff in mature cells. Therefore mechanical differences in cell walls may be a consequence of developmental differences, or age of cells. Further, the patterns of microfibrillar deposition and orientation in cell wall layers affect directions of cell expansion, and influence the geometry of the cell wall and the morphology of individual cells. It is important to recognize that different degrees of protoplasmic partitioning among species result in morphological (form) convergences among those species; convergently, anatomical (structure) convergence among multicellular species can obscure morphological differences. Therefore form (morphology) and structure (anatomy) may provide two independent biological attributes susceptible to selection. (Botanists pay much more attention to distinction between morphology and anatomy than do zoologists, who usually treat structure as a component of form, and label both structure and form 'morphology'.)

The biomechanical properties conferred by multicellularity have to do with the protoplasm behaving as a fluid, and the cell wall as a solid. They have very different individual properties, such as capacities for recovery from deformation, and exhibit particular behaviors when they

A.



B.

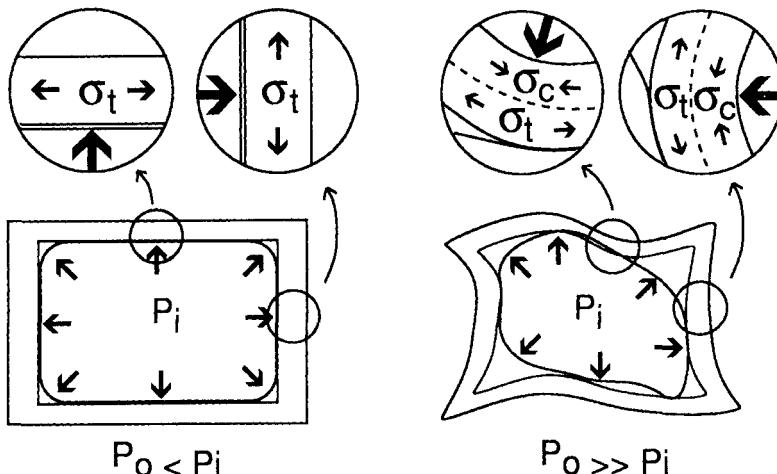


Fig. 6. Biophysics of plant tissue. (A) Stress-strain diagram for a ‘honeycomb’ cellular solid under compressive stress. The stress–strain curve has three regions: initial linear elastic response, in which strain is proportional to stress; non-linear ‘collapse,’ in which elastic buckling or plastic collapse can occur; and a region of densification in which the cellular units within the cellular solid are crushed together and the density of the cellular solid greatly increases. (B) Hydrostatic pressure and mechanical stresses of the cell wall. The protoplast exerts a hydrostatic pressure (P_i) on the cell wall, placing the wall in tensile stress (t), which limits the extent to which the cell wall can deform when under compression.

operate in synchrony. Both the cell wall and the protoplasm are viscoelastic materials that structurally undergo stress relaxation, but their properties are not the same. For example, during growth the cell wall behaves as a solid when subjected to high stress, but when mature has a greater linear elastic range (Fig. 6A). The protoplast behaves as a fluid with a wide range in viscosity that varies with developmental state and other conditions.

Niklas (1989) showed quantitatively that the cell wall and the protoplast act as a hydrostat. This is intuitively obvious; plants deprived of water wilt. In hydrostatically stable plants, the protoplast exerts pressure on the cell walls so they are less likely to buckle under the weight they bear. Under water deprivation, the protoplast volume is reduced, and exerts less pressure on the cell wall, so that the cell more likely deforms. Thickness of the cell wall influences the degree of deformation under a given load. The longitudinal axis of the cell wall is stressed by the internal pressure so that uniform extension or contraction occur. Further, if turgor pressure exceeds ambient pressure, circumferential stress is tensile and radial stress is compressive. When the cell wall thickness is approximately 20 per cent of the cell radius, the circumferential stress and the ambient pressure on the cell wall converge (Fig. 6B). Therefore if the relative volume of the protoplast is great, the hydrostatic mechanism predominates; if the cell wall fraction is high, the cell or tissue is mechanically protected against variation in water availability to the plant.

This also influences how long a branch or twig or shoot can be before it deflects under its own weight. Niklas (1990) evaluated this problem in terms of critical aspect ratios of cylindrical beams of tissues as a function of the elastic modulus of the tissues for which ratios of cell wall to protoplasm have been calculated. In its density range, cellulose is the strongest material known; all plant tissues have a lower elastic modulus than cellulose. Wet cellulose has a lower elastic modulus than dry; cell walls in living plants are somewhat hydrated. However, lignified plant fibers are weaker than non-lignified. Lignin, though, is hydrophobic, so when present reduces the hydration of cell walls. Therefore lignified cell walls will vary less in stiffness, conferring a design factor with a margin of safety against reduction of stiffness because of cell wall hydration. Niklas and Kaplan suggest that as terrestrial plants evolved, they acquired

Fig. 6. Continued

When the protoplast is deflated, the cell wall deforms when a compressive stress is applied (see arrows on upper and side inserts), and it undergoes tensile (*t*) and compressive (*c*) stresses as it buckles. From Proceedings of the Fourth International Congress of Systematic and Evolutionary Biology, copyright 1991 by Dioscorides Press (an imprint of Timber Press, Inc.). Reprinted by permission.

new secondary tissues that made the plant body stiffer and larger. This is a change from a hydrostatic system with thin cell walls to one mechanically supported by thicker cell walls.

The main point of Niklas and Kaplan's argument is that hydrostatic plant bodies are size-limited by the availability of water. Above a given size, the infrastructure of the cell wall becomes mechanically desirable. A hollow tubular cell or stem can resist reduced hydrostatic pressures if it is internally reinforced, and multicellularity is one way of effecting reinforcement. Bending failure is reduced by concentrating cell wall material at the perimeter of cylindrical structures, or by inserting struts such as nodal septa or trabeculae that span the lateral external cell walls. Niklas and Kaplan point out that understanding the mechanical benefits of infrastructure gives little information on the developmental mechanism that gave rise to the infrastructure, nor a sound basis for interpretation of evolutionary modifications of plants.

These three cases illustrate well the use of techniques largely derived from engineering to evaluate the structural capacity of performance of organisms. They offer new insights into structure–function relationships, and the performance of organisms in the environments that provide the physical (and biological) forces with which they must cope. While these studies do not assess the evolutionary history of groups or lineages, they do provide sound generalizations that may be analogies with which to examine the evolutionary history of monophyletic groups. Such studies open new approaches to evolutionary morphology, and should (must) not be constrained by the presumption that evolution can only be evaluated within groups for which there is a phylogenetic hypothesis of relationships (see section 6 for further discussion of this concern). Most importantly, the use of real measurements of physical forces in the environment and the forces generated by and within organisms allows better evaluation of the capacity for change, and the limits to change, given particular morphologies.

4. ECOMORPHOLOGY

The domain of ecomorphology is the assessment of form and function as it correlates with the organism's environment. Ecomorphology may have begun with Darwin, as he associated the morphology of related species of finches in the Galapagos Islands with the attributes of their habitats. This example continues to fascinate biologists, and new ideas have emerged from recent studies of those animals (Grant and Grant 1989). Yet as Wainwright (1991) notes, the ecomorphological paradigm that the functional design of organisms relates to their ecology has suffered historically from lack of a rigorous framework for its analysis. However,

this area has been receiving new attention. More rigorous new techniques that permit quantification of morphology and habitat structure have improved correlations of morphological and ecological patterns, toward the goal of understanding the nature and origin of current associations and how they came to be. Liem (1989) argues that co-existing and interdependent genealogical and ecological hierarchies must be recognized as part of the study of functional morphology. He calls this principle 'symecomorphosis,' and states that it can be used to examine the roles, and even the causal relationships, of various designs in producing morphological diversity in different lineages. Much of the effort of ecomorphologists is devoted to developing methodologies for evaluating ecological/morphological correlations.

I am concerned that, as the methodologies are derived from the empiricisms drawn from the investigations of particular lineages, they are highly specific and do not have the generality that their authors advocate. However, this will only be demonstrated, and models and methods improved, as they are applied to other lineages.

4.1 Case 1: Morphological patterns in communities of Darwin's finches

Schluter and Grant (1984) re-examined arguments about the processes that result in the morphological properties of species communities. They studied the patterns of the morphology of the generalized granivores among Darwin's finches, the populations of *Geospiza* that feed on seeds in the dry season. They focused on interspecific variation in beak size, specifically depth, because of its dietary significance. Other beak variables, as well as body mass, are correlated with beak depth.

A methodology for evaluating alternative explanations for morphological patterns in communities was developed and applied by Schluter and Grant. Their first step was the computation of expected population density as a function of beak depth for a hypothetical solitary finch species on an island as a function of beak depth. They had data on food characteristics for 15 Galapagos islands, so were able to perform such calculations. The second step was the construction of hypothetical finch communities for those islands using five different models: random assembly/evolution, partly directed assembly, directed evolution, directed assembly with competitive exclusion, and directed coevolution under interspecific competition. The third step was the comparison of the predictions of the models to actual communities, so that the roles of food supply and competition could be examined. They found that the beak sizes of species present on each island correspond to local maxima in expected density, but two species never occupy the same or even closely adjacent maxima. The first three models listed above assumed varying

degrees of influence of food supply, and did not adequately explain morphological differences among species. The two models that included effects of interspecific competition in addition to food supply closely accounted for the observed morphological structure in the bird communities.

Schluter and Grant suggest that it is unusual that a single morphological dimension can summarize resource use of a community of species, but their methodology can be expanded to consider other species and other sorts of communities, including mainland. The use of a greater number of morphological dimensions and a more complex set of functions that relate morphological distance to ecological differences could be employed to assess communities that include diverse species with complicated microhabitat specializations.

This case demonstrates the careful and judicious use of a modeling approach with rigorous empirical and statistical tests of correlation of morphological features with habitat attributes. Schluter and Grant clearly delineate a program of research equally applicable to the evolution of Darwin's finches and to other problems in ecomorphology. Quantitative analysis of the relevant data opens new venues for evolutionary morphology.

4.2 Case 2: Fish feeding mechanisms

Wainwright (1991; Wainwright *et al.* 1991) proposes a methodology for exploring experimentally the ecological consequences of variation in morphology. The central concept is that morphology limits the ability of the individual to perform major tasks. First, the effect of morphological variation on performance (behavior) can be tested experimentally. The behavioral capacity of an organism defines the range of ecological resources that the organism can potentially use (the potential niche). Second, the potential niche can be compared with measurable patterns of resource use (the realized niche). A quantitative assessment of the significance of the maximal capabilities of the organism in determining patterns of resource use can then be made.

Wainwright (1991) illustrates his points with a study of the ecomorphology of mollusc eating by the Caribbean hogfish, *Lachnolaimus maximus* (Wainwright 1987). The upper size limit of the mollusc prey is determined by the jaw gape of the fish, and the crushing force that the fish can exert. Both gape and crushing strength increase during ontogeny, and the range of prey sizes and shell strengths that can be eaten by fishes of various sizes also increases (Fig. 7B). Therefore the range of prey that potentially can be consumed by hogfish increases as the fish increases in size. Strength seems to be more important than gape in shaping the diet of the fishes in natural populations. The stomach contents of 59 hogfish taken from a

population on the Belizean barrier reef show that stronger fish eat larger as well as smaller snails, and that crushing strength appears to be the factor that constrains predation by fish of all sizes. Hogfish tended to eat snails up to the maximal size they can crush, and the hardness of the snail shells is such that the fish seem never to test their gape limitation.

Wainwright recently extended his method of analysis to include a population dimension through examination of intraspecific trophic divergence in the snail-eating sunfish, *Lepomis gibbosus* (Wainwright *et al.* 1991). He compared populations in a lake in which snails are abundant prey items, and in a lake from which snails are absent. The pharyngeal jaw muscles and bones were significantly larger in the snail eaters (Fig. 7A). However, only one of the five muscles examined showed a major difference in its pattern of use, and it varied least morphologically between lakes (Fig. 7C). Patterns of change in muscle morphology therefore are not congruent with changes in the snail-crushing motor pattern. The morphological and motor pattern changes underlie the differences in ability to feed on hard-shelled prey between the populations. In spite of the highly integrated, complex feeding mechanism of fishes, morphology and muscle activity patterns frequently change independently, and may be driven directly by ecological factors. A causal connection between crushing strength and mollusc-eating is demonstrated through biomechanical analyses, behavioral performance experiments, and ecological investigations. The method provides a direct means of exploring experimentally the ecological consequences of morphological innovation.

4.3 Case 3: Ecomorphology of Bornean tree frogs

Ecomorphological investigation can be extended to communities of organisms. Emerson (1991) tested the assumption that there is a causal relationship between morphology and ecology that is mediated by performance. She examined intra- and interspecific correlations among morphological-functional and habitat variables for eight species of rhacophorid frogs from stream and forest localities in Borneo. The group is fairly similar in morphology and ecology, and has a common phylogenetic history. Emerson innovatively used biomechanical models of jumping and sticking ability to make predictions about the relationship between morphology and performance, and morphology and ecology. If morphological differences are functionally significant and related to fitness, the predicted relationships should be valid at any level of investigation. This principle of uniformity has not been rigorously tested, especially with the use of biomechanical models. Jumping and sticking ability are likely correlated with fitness, for locomotion is used to change location and escape predators, and the ability to stick, or hold position on leaves or tree trunks rather than on the ground, may influence nesting and

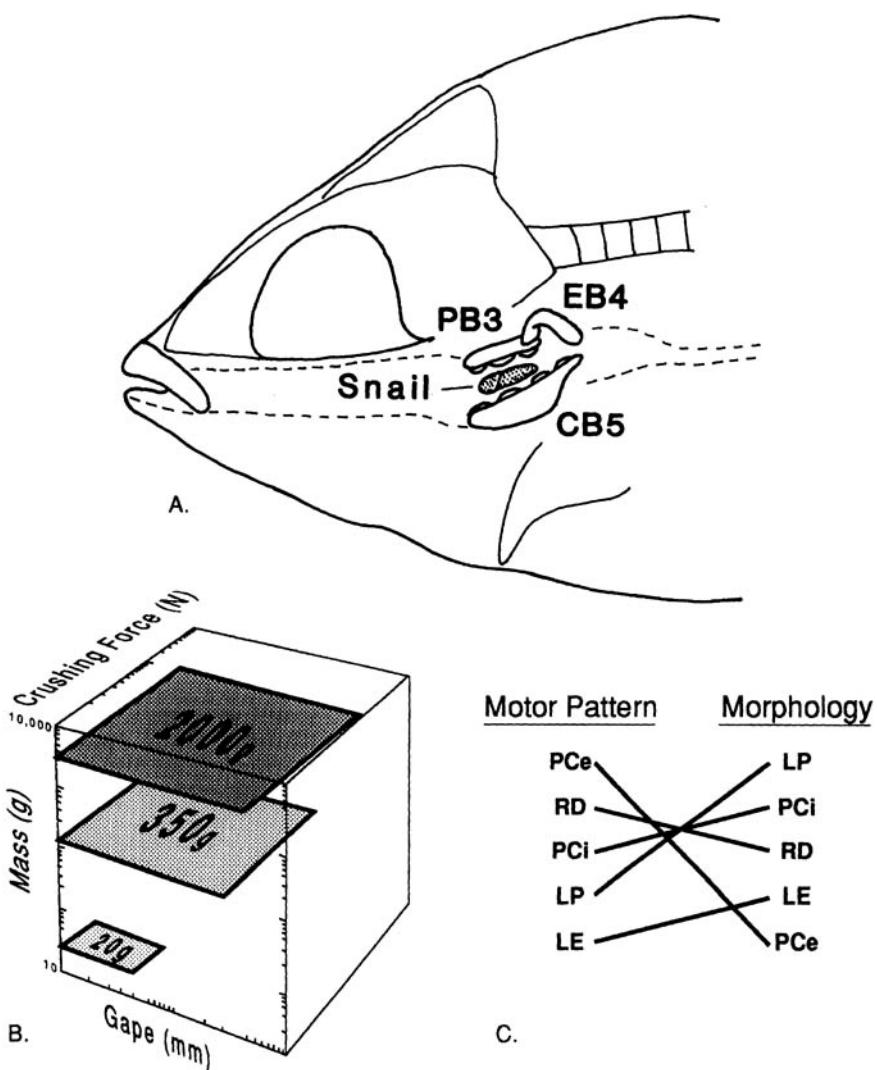


Fig. 7. Ecomorphology of fish feeding. (A) Lateral view of the head of a sunfish showing the position of the pharyngeal jaw apparatus and its muscles. During snail eating, the lower jaw (CB5) is relatively stationary, and the upper jaw exerts the crushing force as it is pressed against the prey item by rotation of epibranchial 4 (EB4) onto the dorsal surface of the upper jaw (PB3) (after Wainwright *et al.* 1991). (B) Three-dimensional diagram of the scaling of pharyngeal jaw gape and crushing strength in the Caribbean hogfish. Ranges of prey sizes and hardnesses that fish can handle are shown for individuals of 20, 350, and 2000 g. Axes are \log_{10} scales. Crushing strength and gape increase during ontogeny, so that the range of prey hogfish can feed on increases. (After Wainwright 1991a) (C) Rankings of population differences for five pharyngeal jaw muscles examined in

reproductive success. Not all probable determinants of sticking and jumping were measured, but selected features were tested to see if they influenced performance in the direction predicted by the biomechanical models. Distance and acceleration of jumping and body mass, and capillarity (the major mechanism of sticking-force generation by toe tips and abdomen of frogs), measured by toe pad area, also correlated with body mass, were used to construct the models. The morphological variables correlated with habitat for both jumping and sticking intraspecifically, with larger individuals (longer hind limbs, greater mass, greater toe pad area) jumping further and sticking better, and were associated with forest locations. Interspecifically, the variables correlated with substrate type; larger animals were correlated with rough bark and trunk substrates, and sticking ability was correlated with being found on leaves. Size and shape variables showed the predicted performance correlation with ecology both intra- and interspecifically. However, the relationship between toe pad area and sticking ability differed intra- and interspecifically. Thus the correlation of morphology, performance, and fitness was not found to be uniform at different levels of investigation. This indicates that the relationship of morphology-function relationships to ecology may be much more complex than assumed, and more careful empirical work is required in such studies.

These examples illustrate the value of quantification of structure-function relationships relative to the environment of the organism. The major contribution, to my mind, is the dissection of behavior, morphology, and control mechanisms as correlated with ecology. Morphology can change, though the motor action (control) pattern does not; there may be changes in the motor action pattern that are not reflected in morphological differences among taxa. Therefore behavior may *not* drive morphological change in all cases, but morphological change may permit new behavioral repertoires. Further, correlations may be obtained at

Fig. 7. Continued

morphological and motor pattern analyses (PCe = pharyngocleithralis externus muscle; RD = retractor dorsalis muscle; PCi = pharyngocleithralis internus muscle; LP = levator posterior muscle; LE = levator externus muscles). Ranking in each column was determined by averaging *F*-ratios from ANOVA significance tests for lake effect. The top muscle in each column is the most different between lakes, and the bottom muscle is least different. Lines connect the muscles in each column and show the incongruence of morphological and motor pattern divergence in sunfish pharyngeal jaws (after Wainwright *et al.* 1991). (A) and (C) from Proceedings of the Fourth International Congress of Systematic and Evolutionary Biology, copyright 1991 by Dioscorides Press (an imprint of Timber Press, Inc.). Reprinted by permission.

one level of investigation, but not at others (e.g. intra- vs. interspecific comparisons).

Hypothetical causal connections between morphology and ecology can be tested by demonstrating the effect of morphology on performance, and performance on ecological resource use. Biomechanical models can be used to construct testable predictions about the correlation of structure–function relationships and ecology. The formal description of the influence of functional morphology on patterns of resource use provides insight to several concepts in ecomorphology, such as ‘key innovations,’ or transformations of organismal design that cause changes in the performance gradient of a lineage, and therefore in the range of resources they can use. The historical correlation between morphological innovation and ecological consequence can be assessed by examining congruence with a phylogenetic hypothesis in the way proposed by Lauder and Liem (1989) (see section 7). The transportability of these approaches to other organisms will test and refine their applicability.

5. DEVELOPMENTAL MORPHOLOGY

The domain of developmental biology is the integration of data on early development and ontogeny of organisms in the analysis of the evolution of structure. Advances in developmental biology, recently through application of cellular and molecular techniques, also permit analysis in an evolutionary framework – when members of the two disciplines bring their approaches to bear on such questions (Alberch 1980). Several recent volumes and reports, usually the products of symposia or meetings (Bonner 1982; Goodwin *et al.*, 1982; Maynard Smith *et al.* 1985; Müller *et al.* 1989; D. Wake *et al.*; 1991), reflect discussions that attempt to find a common ground and to develop a research strategy for analysis of all aspects of developmental morphology (such as cellular and molecular processes, and genetic and epigenetic phenomena) as they relate to, reveal, and direct patterns in evolution.

Müller (1991) has appraised the current status of the attempt to generate a new field of research – evolutionary embryology, to use his term – that combines understanding of developmental and evolutionary approaches. The arena is dominated by theoretical concepts, rather than empirical investigations. Müller considers that five domains of development – recapitulation, timing, thresholds, induction, and plasticity – are amenable to experimentation in order to assess their role in morphological evolution. In addition to the mechanisms that underlie the variation of phenotypes, the generative and integrative capacities of developmental systems must be better understood.

'The search for a theory of organismic form' has followed two avenues that have not been interrelated, according to Müller. The first is the historical approach in which patterns and processes of morphological change over time are studied; the second is the generative approach, in which the processes that create form within individual ontogenies are examined. The latter, embryology, has not been effectively included into the (a) theory of evolution (Mayr 1980). Developmental biologists have needed to consider the evolutionary modification of developmental mechanisms in their search for a theory of development, and evolutionary biologists have recognized that the synthetic theory, with its genomic base, inadequately explains morphological and organismal evolution. A comprehensive theory of organismic form is needed that relates embryonic development to longer-term processes of evolution, with the rationale that phenotypic structural change must originate through changes in ontogeny, and the two processes must be linked.

A major new concern of the new domain is investigation of the mechanistic relationship between evolutionary and developmental processes. Attention is placed on analysis of causal interactions among genes, development, and evolution, rather than embryology as a source of data for systematic analysis. The questions now focus on the influence of developmental mechanisms on morphological evolution, and how ontogenetic processes are modified in phylogeny. Techniques of modern experimental embryology, especially those of molecular biology, can be applied to several hypotheses or concepts. Mechanisms of evolutionary change in developmental systems, including heterochrony, threshold effects, induction, and plasticity, and the effects of developmental parameters on morphological evolution, such as the basis of homology and of developmental constraint, are experimentally testable.

5.1 Case 1: Heterochrony

Heterochrony, the evolutionary modification of relative rates and timing of developmental processes within and among lineages (Gould 1977; Alberch *et al.* 1979; Alberch 1980), currently is receiving much attention. Various workers are attempting to determine whether heterochrony is a pattern or in fact a mechanism of evolutionary change. Many recent studies show that heterochrony is almost a universal occurrence among classes of organisms, levels of development, and diverse organ systems. A number of developmental parameters have been identified as targets for heterochronic modification (onset and offset times, mitotic and growth rates, spatio-temporal expression of molecules, etc.). Crucial to the analysis of heterochrony is the assumption of dissociability of developmental processes in time and space. Study of heterochrony is inherently

comparative, for its analysis must reveal differences in timing of ontogenies of related species by experimentally perturbing timing in development.

Raff and his co-workers (Raff 1987; Parks *et al.* 1989; Raff and Wray 1989) have compared the mesenchymal cell lineages of direct-developing species of sea urchins and species that have pluteus larvae. They found major differences in cleavage pattern, mitotic rate, and mesenchyme-specific antigen expression in pattern and timing. They are now using single-cell marking techniques to compare cell lineages and to ascertain modifications of cellular arrangements in the embryos in time and space. Such results demonstrate that early development is not conservative and constrained relative to later development, as had long been assumed. Further, Raff *et al.*'s results demonstrate the dissociability of larval and adult mesenchymal skeletogenesis, so that they propose a heterochrony model for the origin of direct development in sea urchins. They speculate that through compartmentalization, larval developmental programs can be suppressed and adult programs accelerated, so that major phylogenetic changes are established early in ontogeny but embryogenesis is not disrupted at all.

A second approach to analysis of heterochrony is the manipulation of developmental timing differentially of component parts of a structure. For example, Hanken and Summers (1989) implanted pellets loaded with thyroid hormone into the heads of frog tadpoles. This induced early metamorphic transformation of skeletal elements of the head, and demonstrated that cartilage is more sensitive to thyroxin stimulation than is bone. The hormone stimulation disrupted the normal relationship between cartilage transformation and bone formation, so that Hanken and Summers concluded that endocrine mediation of chondrogenesis is independent of mediation of osteogenesis, and that such treatment effects the dissociation of bone and cartilage during phylogeny. This approach should be made comparative in order to ascertain the patterns of natural transformations. It is important to distinguish between cause and effect, since almost any change in a developmental program will affect the timing of processes. Further, there are suggestions that non-heterochronic mechanisms can cause changes in the timing of developmental events. For example, Collazo (1990) showed that large quantities of yolk in the eggs of direct-developing salamanders provide mechanical constraints that prolong development, with such consequences as simultaneous, rather than sequential, limb development. Also, since heterochronic analysis pertains to modification of existing patterns, such mechanisms may not account for the origin of new structures. Experimental approaches to developmental biology may illuminate mechanisms of change in structure and function, thus addressing the pattern and process of evolution.

6. SYSTEMATICS, PHYLOGENETIC ANALYSIS, AND MORPHOLOGY

Systematics is the study of organismic diversity as that diversity is relevant to a specified relationship thought to exist among groups of populations or taxa (Wiley 1981). Phylogenetic systematics attempts to discover the genealogical (historical, phylogenetic) relationships among organisms and to produce classifications that reflect those relationships. Taxonomy is the theory and practice of describing the diversity of organisms and ordering the diversity in a system that conveys information about the relationships among groups of organisms.

Morphology has provided most of the data used in systematic analysis. Descriptive morphology of both extinct and extant organisms has generated data used in analysis of phylogenetic relationships, the evaluation of character state transformations, and assessment of homologies. Lauder (1981, 1982) stressed the use of phylogenetic information for determining the polarity of evolution and for analyzing organismal design. The recent advent of the use of biochemical and molecular data in systematics has allowed a comparison of relatively unrelated data sets. This has provided a complementarity, for incongruences of patterns of phylogenetic relationships generated from testing the two data sets reveal to morphologists areas that need further investigation, particularly in analysis of homoplasies, and tell biochemical and molecular systematists where to pin-point their research. Kluge (1989) avers that biochemical/molecular data and those from morphology or other bases are not simply complementary, but that *all* available data must be used in phylogenetic analysis – his principle of character congruence. His point is that combinations of characters add confidence to a phylogenetic analysis, for the data are more likely to be independent than those drawn from the same character system. Incongruence can then reveal independent evolution, so that its pattern can be investigated. I, and others, are concerned that a system that uses all data indiscriminately might produce faulty conclusions. For example, large numbers of DNA sequence units can now be generated. Is every base pair position a character? Will the use of large numbers of biochemical and molecular characters in conjunction with small numbers of morphological characters methodologically result in a meaningful analysis, or will the molecular characters swamp out all others? These problems are being examined by some evolutionary biologists, but resolution is not at hand. Until better understanding of the data occurs, I advocate separate analysis of different kinds of data, with comparison of the cladograms generated to examine congruity (and incongruity).

Systematics has been advanced recently not only by acquisition of new techniques from cell and molecular biology, but by the development of

new methods of phylogenetic analysis. The explicit methodology of cladism (see Wiley 1981) has facilitated new approaches to the generation of phylogenetic hypotheses. As demonstrated by most of the examples cited in this essay, morphologists are integrating an evolutionary perspective into their work by using an historical, and whenever possible, a phylogenetic context. In fact, Maynard Smith (1991) considers the advent of the use of cladistics to determine phylogenetic (evolutionary, to Maynard Smith) relationships to have provided major change for the better in the practice of morphology. I will not review the implications of cladism for evolutionary biologists – that has been done by several authors recently (e.g. Greene 1986; Coddington 1988; Funk and Brooks 1990; Baum and Larson 1991; Brooks and McLennan 1991) in cogent arguments for phylogenetic approaches to the study of adaptation, patterns of evolution, and all of comparative biology. Phylogenetic analysis opens new areas for morphologists in that it allows examination of the use of ontogenies as characters, renews questions about the nature of homology, and looks rigorously at morphological data of all sorts.

Many evolutionary biologists recently have independently adopted the technique of mapping structural and functional characters or transitions on cladograms (diagrams of corroborated or generally accepted hypotheses of phylogenetic relationships arrayed in branching sequences of relatedness), as several examples in this essay illustrate. D. Wake (1991) called such diagrams ‘scenariograms.’ They can be used in two ways: first, by mapping characters on cladograms (representations of hypotheses of phylogenetics based on shared derived character states), one can see if phenotypic expressions, morphologically and functionally, match hypotheses of change among groups; second, one can allocate characters onto a cladogram, or phylogenetic hypothesis, and then use patterns of character representation to generate hypotheses of evolutionary change, including adaptation. D. Wake and his collaborators (summarized in 1991) employ ‘scenariograms’ in both ways. They are used in inductive studies that map the data on the cladograms in order to assess such patterns as the number of times a state may have arisen (evolved) independently (webbing of feet, tooth numbers and positions, fusions and/or separations of osteological elements, reproductive mode, tongue morphology, etc.). They also have been used in deductive studies in which evolutionary hypotheses – statements of sequence of change of taxa in a lineage, and their patterns of connectivity – are proposed. Both cladograms and the data used to test the hypotheses must be assessed in terms of possible (probable) homoplasy, or reversals and/or convergences in morphological stages in lineages. Carrier (1987, 1991) used the second technique to evaluate the potential constraint implied by the use of the same elements of the body skeleton and its musculature for both locomotion and respiration (Fig. 8B). For example, he proposes, based

on the functional morphology of large lizards, taken as analogues of members of the amphibian-reptile transition because of their spawling gaits, features of their osteology and myology, etc., that the first terrestrial vertebrates were unable to run and breathe at the same time. This remains the case in most reptiles today, though many have evolved means of compensating for this constraint by having 'sprint' locomotion, in which they run quickly for a short time, then stop – and, among other things – breathe.

Generation of 'evolutionary scenarios' involves rigorous methods, hardly Panglossian adaptive explanations. Evolutionary scenarios provide hypotheses that are testable, directly or indirectly. Scenariograms are probable sequences of evolution that *enable* an interpretation; the interpretation is often, but not necessarily, adaptive. There are reasons other than adaptation for the interpretation of homoplasy, or patterns of convergent morphologies, such as developmental constraints and the stability of ontogenetic trajectories (Wake, D. 1991).

Robust phylogenetic hypotheses provide a framework for analyzing the implications and evolution of structural, functional, and developmental patterns. However, I see two current problems with the application of phylogenetic hypotheses to morphological data. First, so far neither systematists nor evolutionary morphologists have made significant use of functional patterns as characters of utility in generating phylogenetic hypotheses. They characteristically revert to the morphological structure that underlies function in their search for characters, and some doubt that 'functional characters' can be used (see Wake, M. 1991). Cracraft (1981) reviewed this dilemma from the perspective of a systematist, and found functional characters premature at best.

The recent realization by some functional morphologists that functional characters not only can be, but should be, used in order to understand patterns of evolution in an historical context will open new areas of analysis. Lauder (1990) considers the study of function a neglected area in systematic and historical biology. He believes that the form–function dichotomy in biology has been weighted toward the study of form in systematics and evolutionary biology. Direct experimental measurement of function in living organisms provides unique insights into the uses of structural characters, the distribution of characters on a cladogram, interactions and correlations among characters, the nature of organismal diversity, and historical patterns to organismal design. While functional (or physiological) data do not necessarily include such morphometric features as measurements of muscles as force generators, etc., such information may be interpreted in a functional context.

Functional attributes are an important class of characters, for they allow assessment of the way that structures are used by organisms and the way that the use of the structures has changed in evolution. Lauder (1990) asserts that analysis of structural features alone gives an incomplete

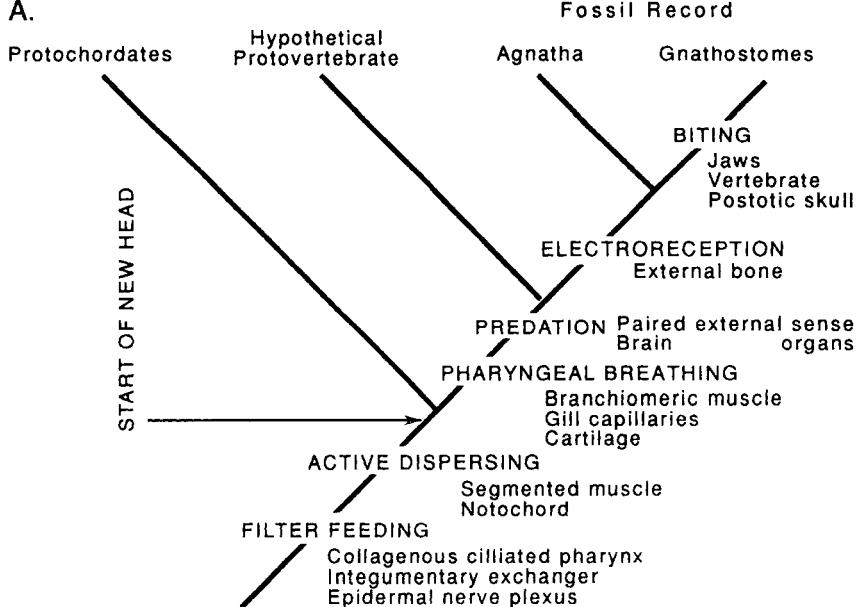
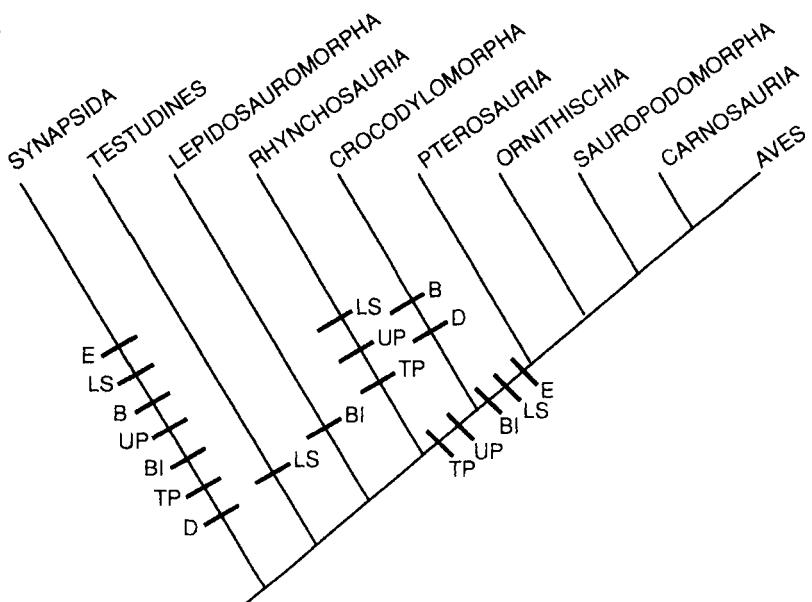
A.**B.**

Fig. 8.

picture of organismal design, though structural characters may generate phylogenetic hypotheses. He illustrates his point with a hypothetical example of a functional analysis of a group of seven species of fishes whose phylogenetic relationships had been determined previously. The structural features of a system of bones and muscles (Fig. 9A) are studied, as are the activity patterns of the muscles. Then both structural and functional attributes of each species are mapped onto the cladogram (Fig. 9B). With both components on the cladogram, one can draw several inferences about the evolution of muscle function. For example, control of bone **d** evolved as ligament 6 arose, followed by ligament 7. Pathway **i** arose first, and species in clades A to E can move bone **d** by pathway **i**. Later, pathway **ii** arose by the addition of ligament 7; only clade A has pathway **ii**. Other features of divergence, retention of the primitive functional pattern, and convergence can be determined as well. Such an analysis allows precise definition of the historical sequence of both structural and functional change, so an understanding of the way a biomechanical system was constructed can occur. Further, one can determine whether structural and functional changes occur congruently, and examine whether functional data tend to be more evolutionarily plastic than structural data. I caution, however, that a circularity must be avoided – that of using the data to erect the phylogeny on which the interpretation of sequence of change is then based. If the data are added to a larger set from which the phylogeny is derived, the problem is reduced.

Functional analyses are useful in assessing how and why characters are correlated with one another in phylogenetic series. Biomechanical and functional analyses may place boundary conditions on hypotheses of form–function–performance relationships. If, for example, the origin of a morphology hypothesized to facilitate a particular performance variable (behavior) occurred before the appearance of the function or performance, then a causal hypothesis among the morphological and the functional characters is not supported. Functional morphology provides an experi-

Fig. 8. Scenariograms in evolutionary morphology. (A) Hypothesized structural and functional transitions in the evolution of vertebrates (after Gans and Northcutt 1983, *Science* **220**, 272. Copyright 1983 by the AAAS). Note that the postulated functional states (capital letters) precede the modified structures (lower-case letters) involved in those states. (B) Analysis of the decoupling of locomotion and ventilation in the evolution of amniotes (after Carrier 1987). Characters (D, diaphragmatic muscles; TP, large transverse processes; BI, bipedal locomotion; UP, upright posture; B, bounding; LS, lateral stability of the vertebral column; E, endothermy) are mapped on a cladogram of the amniotes. Such an analysis can be used to predict the functional states of extinct, as well as extant, taxa.

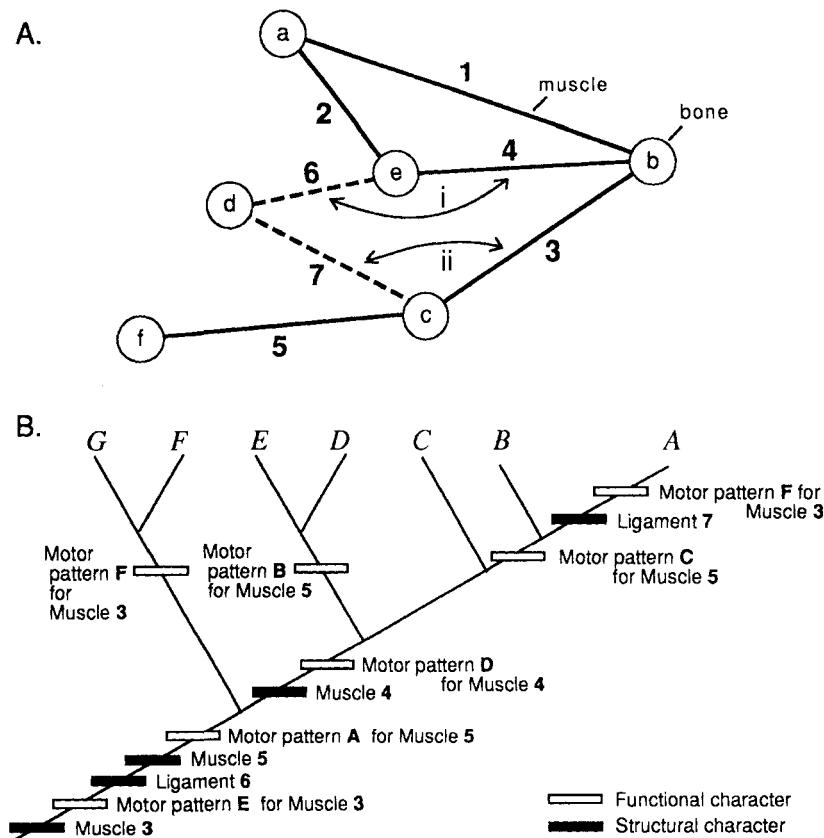


Fig. 9. Diagram of a complex biomechanical system (A) and the cladogram illustrating the historical sequence in which it was constructed (B) (redrawn from Lauder 1990; reproduced with permission from the *Annual Review of Ecology and Systematics*, vol. 21, copyright 1990 by Annual Reviews Inc.). Circles indicate bones, labelled with lower-case letters; ligaments are indicated by dashed lines, muscles by black lines. There are two mechanical pathways for moving bone d: i and ii. Pathway i involves muscle 4 and ligament 6. Pathway ii includes muscle 3 and ligament 7. Characteristics of species A to G are mapped on a cladogram (B). The muscle activity pattern for muscle 3 is convergent between clade A and clade G. The analysis also shows that functional characters (open bars) and structural characters (solid bars) can exhibit incongruent distributions.

mental, theoretical, and mechanistic basis for choosing among correlated characters to determine which are causally related to historical (phylogenetic) changes in function or performance.

Hickman (1988, 1991) carries this concern a step further. She points out that functional characters can be used to construct phylogenies, and then to allow predictions about the evolution of function and of

morphology. She has applied this concept to analysis of recent and fossil functional morphology, and has used a modeling approach to analyze function in fossils in particular. She cogently argues that any character, structural or functional, is only as good as its definition and then its integration in a cladistic methodology, so functional characters should not be rejected out-of-hand. She indicates that fossils often leave only functional characters, such as scraping marks on substrate made by action of snail radulae, so that morphology is inferred from function, rather than function necessarily determined from structure.

Still, many workers have difficulty accepting functional characters, both because they believe that the underlying morphology must be sought, and that morphology then provides the characters appropriate to phylogenetic analysis and because functions are viewed as associations of several potential characters (see discussion in Wake, M. 1991). Further, the causal relationship of function to morphology is not explicit, and must be explored rigorously (Baron 1991).

The second problem that phylogenetic analysis has introduced to evolutionary morphology is an uncritical dependence on phylogenies for the analysis of patterns of evolution. The majority of the cases I have selected as exemplars of evolutionary morphology present data analyzed in terms of phylogenetic hypotheses for the organisms under study. Useful as cladograms are in appropriate contexts, one must understand the *basis* of the phylogenetic hypothesis in order to interpret patterns of structural and/or functional change. Further, it has been implied that work undertaken in the absence of a robust phylogenetic hypothesis cannot be a study of evolution (see section 3.3 for comment on this assertion *vis-à-vis* biomechanics). Morphological and other work is necessary in order to generate phylogenetic hypotheses, and cannot be constrained by absence of hypotheses. Comparative morphology remains an important base for phylogenetic reconstruction. It is part of a program in evolutionary morphology that seeks understanding of pattern.

I am convinced that we are not using morphology to the extent that we might for either studies of structure and function, or for construction of phylogenetic hypotheses. I am exploring this question by analyzing sets of non-traditional morphological characters in gymnophione amphibians (caecilians). I have analyzed my data on developmental and comparative adult morphology of the neuroanatomy of special sensory organs, the cranial and anterior spinal nerves, and the reproductive system to see (1) whether characters of systematic utility are present in these sets, (2) what such characters contribute to analysis of phylogenetic analysis, and (3) what subjecting the data sets independently and together to cladistic analysis reveals about the morphology of these systems as characters. There is a fairly well corroborated phylogenetic hypothesis for the families of caecilians, but not for genera and species. I used characters of eye,

ear, vomeronasal organ, hypoglossal nerve, and sperm morphology to see what clusters of taxa would result from cladistic treatment (neuroanatomical characters) and from principal components analysis (sperm morphology) (Wake, M. 1993). The results met some expectations, and included some surprises. Species within genera usually grouped together; members of the family considered most primitive based on other morphological data were also most primitive based on neuroanatomical characters. The family Caeciliaidae, which I consider paraphyletic based on other characters, is demonstrated to be so using these additional characters. However, members of the family Typhlonectidae, a group of South American aquatic and semi-aquatic animals (other caecilians are terrestrial burrowers), did not cluster together in any of the analyses. One species consistently clustered with an east African scolecomorphid, two others with various taxa in three different families. Only one data set (vomeronasal) gave results comparable to the generally accepted phylogenetic hypothesis. Caeciliaid taxa did not always associate according to the general hypothesis. These results are promising! First, characters from these 'non-traditional' systems are indeed useful in phylogenetic analysis. Secondly, and more importantly to me, these independent and composite analyses give me clues to other important problems in evolutionary morphology. For example, I must investigate the apparent convergence in neuroanatomical morphology of otherwise highly dissimilar animals; since there is so much unexpected but useful variation in the morphology of the structures investigated, I must further explore their developmental morphology in order to understand their bases as characters; I must consider the methodology of phylogenetic analysis in these evaluations. Evolutionary morphologists must not merely use already generated phylogenetic hypotheses as part of their analyses, but they must understand cladistic methodology, use it to explore newly generated morphological data, and use their data to contribute to new phylogenetic hypotheses.

7. INTEGRATIVE EVOLUTIONARY MORPHOLOGY: INVESTIGATION OF THE EVOLUTION OF COMPLEXITY

Evolutionary morphology is more than an umbrella term – its domain is the integration of development, ecological, biomechanical, and phylogenetic analysis (as appropriate) to answer questions of the evolution of organismal complexity. Recently, many workers have successfully integrated various of the facets of evolutionary morphology, and have provided new syntheses and conclusions, and opened new ways of approaching the study of evolution. Some workers have used a wide array of techniques to reach this synthesis, while others have attempted to meld philosophies and

methodologies as well as techniques. Analysis of the evolution of complexity and of complex organisms and integrated systems now can include the effects of morphology on performance, the importance of performance on fitness, and the quantitative genetics of functional attributes.

A recent Dahlem Conference provides an example of the exploration of the integration and evolution of complex organismal functions in vertebrates (Wake and Roth 1989). Discussion focused on the evolution of feeding, locomotion, and reproductive biology to elucidate both conservatism and innovation in the evolution of complexity (Roth and Wake 1989), and developmental, ecological, and phylogenetic (historical) influences on the evolution of complex systems. Transitions in the evolution of complexity were examined at levels from the cellular to the lineage. Particular emphasis was paid to the integration across major phylogenetic boundaries and hierarchical levels of organization (Bramble and Jenkins 1989; Duncker 1989; Vrba 1989). Morphological modification, functional integration, including developmental (Hinchliffe 1989) and ecological and populational (Schluter 1989; Shine 1989; Emerson and Arnold 1989) influences, and such aspects as phenotypic plasticity, its origins and its consequences, were examined. Wake and Roth (1989) reiterated that the processes that modify ontogenies, such as heterochrony, can have major morphological consequences although there is little genetic change. The complex integration of developmental interactions provides a stability to morphology in lineages, but on the other hand, modification of ontogenies can produce surprisingly great morphological divergence. They introduce the concept of ontogenetic repatterning as a significant factor by which evolutionary novelties could be produced (Roth and Wake 1985). The empirical example that Roth and Wake explore is that of feeding in plethodontid salamanders, in which direct development is associated with loss of larval gills. The remaining gill components are involved in a novel force transmission employed in tongue projection for prey capture. Ontogenetic repatterning is the consequence of a major change in morphogenetic processes that yields morphological diversification. Evolutionary and phylogenetic phenomena therefore may be based on changes in developmental systems. New networks of interaction are established that decouple previously established systems, so that novelty is introduced into complex integrated systems.

A significant outcome of the new concern about the evolution of complexity has been the suggestion of methodologies for its investigation, as in other areas of evolutionary morphology. Lauder and Liem (1989) provide a comparative phylogenetic methodology for testing the role of historical factors in the evolution of complex organismal functions (Fig. 10). Two classes of historical factors may regulate the diversity of organismal design: key innovations may be causally related to diversifi-

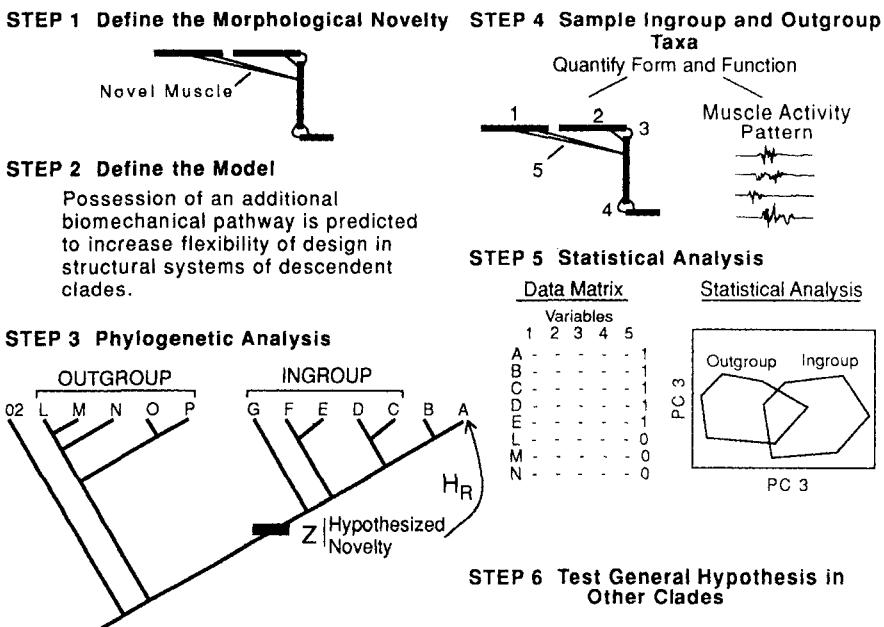


Fig. 10. Diagram showing a six-step phylogenetic procedure for testing historical hypotheses of structural innovation and constraint (after Lauder and Liem 1989). A hypothetical musculoskeletal system that resembles a vertebrate limb is modeled, and a novel muscle connects two of the bones. The steps of the procedure are defined in the diagram. In step 3, the phylogenetic analysis indicates the appearance of the novelty in a clade. In step 4, myelograms representing activity patterns of different muscles are presented as examples of quantifiable data. In step 5, statistical analysis of a taxon data matrix is indicated.

cation of structure and function, and developmental constraints may limit the diversity of phenotypes. Lauder and Liem emphasize that the interactions among different levels of design must be addressed, and hypotheses of the causal historical factors must be formulated and tested. They propose a six-step procedure for the testing of historical hypotheses of structural innovation. It requires (1) definition of the morphological novelty; (2) definition of a causal model, (3) a phylogenetic analysis of the distribution of the novelty, (4) quantification of form and function of the novelty in ingroup and outgroup taxa, (5) a statistical analysis of the variation among taxa, and (6) repetition of the procedure in other clades to test the generality of the pattern of the evolution of the novelty. They illustrate the application of the methodology to hypothetical examples. Liem (1989) proposes a similar approach to the hierarchical examination of morphological, ecological, and phylogenetic data (discussed above).

Zweers (1991) offers a different but not dissimilar methodology for the analysis of the evolution of complexity (Fig. 11A). The difference from previous methodological procedures that he has employed is that the system is one of multiple roles. Therefore a shift in the proportion of *several* functional performances is maximized, rather than a single one. Zweers illustrates the methodology by considering feeding mechanics in ducks (Fig. 11B), based on the extensive database his group has gathered. He illustrates how probing and filter-feeding evolved from pecking mechanisms, through dichotomous pathways that feature decoupled mechanisms, functional shifts, and structure–performance compromises. The feeding system is phenotypically plastic, so that feasible changes can

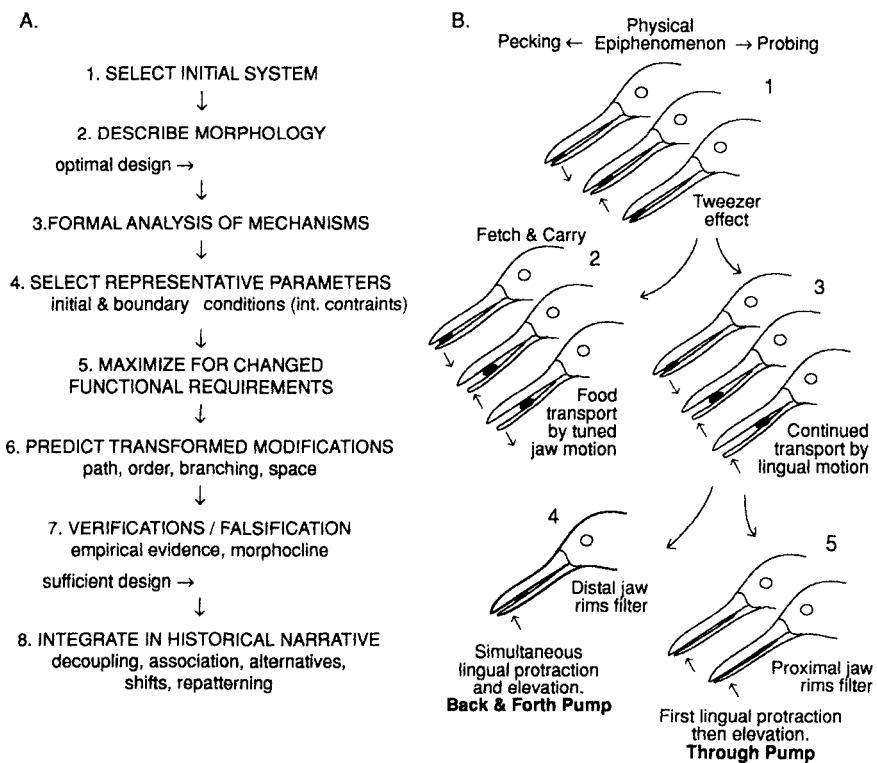


Fig. 11. A methodology for analysis of a complex morphology (after Zweers 1991). From Proceedings of the Fourth International Congress of Systematic and Evolutionary Biology, copyright 1991 by Dioscorides Press (an imprint of Timber Press, Inc. Reprinted by permission.). (A) Diagram of the steps of the methodology. (B) Diagram illustrating the methodology with the branching of feeding mechanisms from a pecking mechanism that is slightly maximized for probing, then further maximized for either fetch and carry transport, back and forth, or through pumping transport.

be predicted. This constitutes a large ‘mechano-space’ in Zweers’ model, which means that filter-feeding, drinking, and pecking can alter without conflicting with each other. The alterations involve ontogenetic modifications and behavioral repatterning, as illustrated in Fig. 11. Zweers characterizes this methodology for analyzing complexity as deductive, revealing feasible paths for evolution of diversity of feeding mechanisms from a set of initial conditions.

I note that there is convergence in several aspects of the methodologies of Lauder and Liem and Zweers, though they are independently derived. The goals of both are elucidation of pattern in morphological evolution, and both use a phylogenetic approach, Lauder and Liem more explicitly than Zweers. These workers advance the study of morphology by codifying their approaches. I remain concerned that because the models are derived from, and applied to, specific lineages thus far, their general applicability is not yet tested. In fact, they likely are points of departure, rather than generalizations. For example, to use Lauder and Liem’s model, one would have to choose taxa for which there are robust phylogenetic hypotheses for both the group in question and for other appropriate clades, or a group so large that there are replicates. This is a major limitation in that it dictates that only well known groups are amenable to analysis. The model is very useful for groups for which the data exist, but cannot be construed as the only approach to the analysis of key innovations (I suspect Lauder and Liem agree).

7.1 Case 1: The integration of studies of morphology, performance, and natural selection

Bock (cited in de Cock Buning *et al.* 1985) suggested that functional morphological studies should be extended to a consideration of the fitness of organisms, rather than emphasizing ‘optimal’ performance. A rigorous framework for such studies has been developing. The relationship of morphological modification and function, or performance, to the fitness of animals and the heritability of performance is a burgeoning field of investigation (Arnold 1983; Bennett 1989; Emerson and Arnold 1989). The application of quantitative genetics theory, models, and methodology allows new insights in evolutionary morphology. Examination of the interaction of morphology and performance yields information about selection. However, as Emerson and Arnold (1989) indicate, intra- and interspecific studies provide different kinds of information. Intraspecific study helps characterize the selection that acts within populations; interspecific studies can resolve aspects of species selection. Combining intra- and interspecific studies could deal with yet other issues, such as the relationship of correlational selection within populations to that among species. Further, Emerson and Arnold suggest that combining intra- and

interspecific studies with phylogenetic and genetic analyses, interspecific covariance, or correlation, in directional selection can be tested. Such studies combine the functional and ecomorphological approaches discussed above, and extend them to specific analyses of selection. The groundwork for such studies is well established (Lande 1979, 1980, 1986; Cheverud 1982; Lande and Arnold 1983; Slatkin and Kirkpatrick 1986, for example). Arnold and Bennett (1988) provide an example of this approach in their study of vertebral numbers and locomotor performance. They sampled 174 newborn garter snakes, and found that numbers of body and tail vertebrae had an interactive effect on burst speed in a 1 m dash. The optimum number of tail vertebrae was an increasing function of the number of body vertebrae, and the interactive effect is presumably part of the selection that acts on newborns. The authors have under way mark and recapture studies of cohorts of newborn snakes of a related species so that they can relate burst speed to survivorship and growth rate in nature, and thereby how important selection on vertebral number might be.

7.2 Case 2: Evolutionary scenarios

Gould and Lewontin (1979) criticized the ‘adaptationist programme’ in part because it facilitated facile ‘evolutionary scenarios’ that were used to provide speculative pathways, or ‘explanations,’ of the course of evolution. They commented that with a good imagination, any body of data could result in a new and plausible scenario to rationalize change from one state to another, and to suggest ‘selection pressures’ that might have effected change, i.e. ‘Just So’ stories. Evolutionary scenarios, then, were seen as end-products of the acquisition of some data that were used to explain the comparative biology of those data. There were many responses to Gould and Lewontin’s criticism, some defenses of the ‘adaptationist programme,’ but more that presented cogent thought and more rigorous analyses of adaptation, often in an historical context (perhaps one of the desired outcomes of Gould and Lewontin’s comments). Several workers recommended broader approaches to the study of adaptation and its complexity (Gans 1974; Bock 1980, 1989; Mayr 1980, 1983, among others). Gans (1988) directly challenged their indictment of adaptationism, and the examples they used to support their argument. He subsequently (1989) presented a thoughtful assessment of the usefulness of evolutionary scenarios that addresses the concerns of Gould and Lewontin. Discussions of the origin of structures and systems are particularly reliant on scenarios, and they have been criticized as something other than scientific endeavor. Gans defined scenarios as outlines of an hypothesized chain of events. Phylogenetic scenarios represent a second set of evolutionary hypotheses that complement the primary cladograms,

or phylogenetic hypotheses, which provided the initial basis for analysis. Phenotypic gain or loss of a character state between phylogenetic stages does not test hypotheses about the process responsible for the observed phenotypes. Gans proposed a set of 'rules' for establishing scenarios, emphasizing (1) derivation from a 'phylogenetic scheme,' (2) use of the best possible information about the biology of the extant members of the group in question, (3) acceptance of several assumptions: natural selection and adaptation as presently characterized, *present* mechanisms of ecology, behavior, physiology, and morphology operating *previously* in the history of the group (this assumption set aside if observations cannot be explained by currently operating mechanisms), (4) any phenotypic pattern may be derived through more than one adaptive pathway, and (5) hypotheses must be framed so that evaluation of their relative merits is permitted.

Gans commented that critics of scenarios are correct that tests of events must go beyond their possibility to their relative probability. Because the most probable path may not necessarily be followed, a range of potential scenarios the stages of which have different probabilities should be constructed. Most importantly, scenarios must be testable hypotheses. However, evolutionary events constitute stages of history and cannot be tested directly. Therefore tests should be made of internally consistent corollaries of the hypotheses. Such tests may be examination of the structures of extant organisms and of available fossils to determine whether their characteristics fit the predictions of the hypothesis. Phylogenetic scenarios allow examination of the evolution of organisms in new ways.

Gans and Northcutt (1983, 1985) and Northcutt and Gans (1983) illustrated these principles by developing a nested set of scenarios for the origin of vertebrates. They associated a presumed sequence of behavioral and environmental changes with observable phenotypic states, and mapped both structural and functional transitions on a cladogram of the relationships of major groups in vertebrate evolution (Fig. 8A). Their analysis integrates morphological, developmental, functional, and ecological data. They emphasize that the evolution of neural crest, and consequently its derivatives, was the singular event in the origin of vertebrates. The vertebrate head, with its neurological, skeletal, and other attributes, associated with a change from passive feeding to active predation, is a new invention, and involves a host of associated functional and morphological changes. Epidermal placodes of the head, interacting with neural crest tissue, give rise to special sensory organs. Bone and cartilage both developed initially in association with electroreceptive and mechanoreceptive organs, respectively, and migrated from their association with the integumental and pharyngeal skeleton. Both may have arisen as spacers or support structures for sensory structures, or perhaps for facilitation of receptor sensitivity. The elements of the facial

skeleton are derivatives of neural crest, but those posteriorly are not, suggesting that a shift in secretory or inductive capacity among cell lineages allowed deposition of hydroxyapatite and proteoglycans. Gans and Northcutt provide hypotheses of the structural and functional transitions associated with the change from filter feeding to predation, acquisition of pharyngeal breathing, and other characteristics of vertebrates (Fig. 8A). They propose alternative scenarios that suggest further research on the biology of selected animals, and tests of the hypotheses are proposed. The use of new molecular techniques in developmental biology, for example, will allow examination of the derivatives of neural crest cells, which will contribute significantly to assessment of the homology of structures. Such assessments of patterns of change are crucial to understanding of origin and evolution of taxa of organisms.

7.3 Case 3: Evolution of brachiopod hinges

Carlson (1989) provides an elegant analysis of morphological and functional variation in the hinge mechanism of articulate brachiopods and the evolution of its complexity. Qualitative differences in five characters of the hinge system are used to establish taxonomic units (genera, families, etc.) within the Articulata, among both Recent forms and those of the extensive fossil record of the group. However, there is not yet a robust phylogenetic hypothesis of the major groups of brachiopods. Little work has been undertaken to analyze hinge components as an integrated functional complex, so there is a limited understanding of the functional implications and the evolution of the differences in the system. The study demonstrates rigorous application of several mathematical and engineering techniques to understand the geometry, morphology, and function of the hinge, and provides new ideas about the nature of the 'functional discontinuity' used to delimit two major groups within the Articulata. Carlson tested the hypothesis that one group developed the ability to resorb and remineralize shell material during growth, particularly in the hinge region; the other group did not acquire this property. Resorption and remineralization allow the teeth and sockets of one group to be interlocking, so that they cannot be disarticulated without breaking. The teeth of the other group are not locked into sockets, but rest in them, so that the hinge line is a fulcrum about which the valve rotates. Therefore the ontogeny, as well as the morphology and function, of the two groups is distinct (literature summarized in Carlson 1989).

Carlson examined species from the Cambrian through the Recent; she notes that only the Paleozoic (each time period) and the Recent faunas are well sampled, so that more work is needed to fill in time and taxonomic gaps. Truss analysis was used to define and compare the geometries of hinge systems, and other techniques (principal components

analysis, major axis analysis, analysis of ratios) were employed to examine shape differences, size-related allometries, and other variables of the morphology (Fig. 12A). The geometry of the taxa indicated major functional differences between the two groups; for example, moment

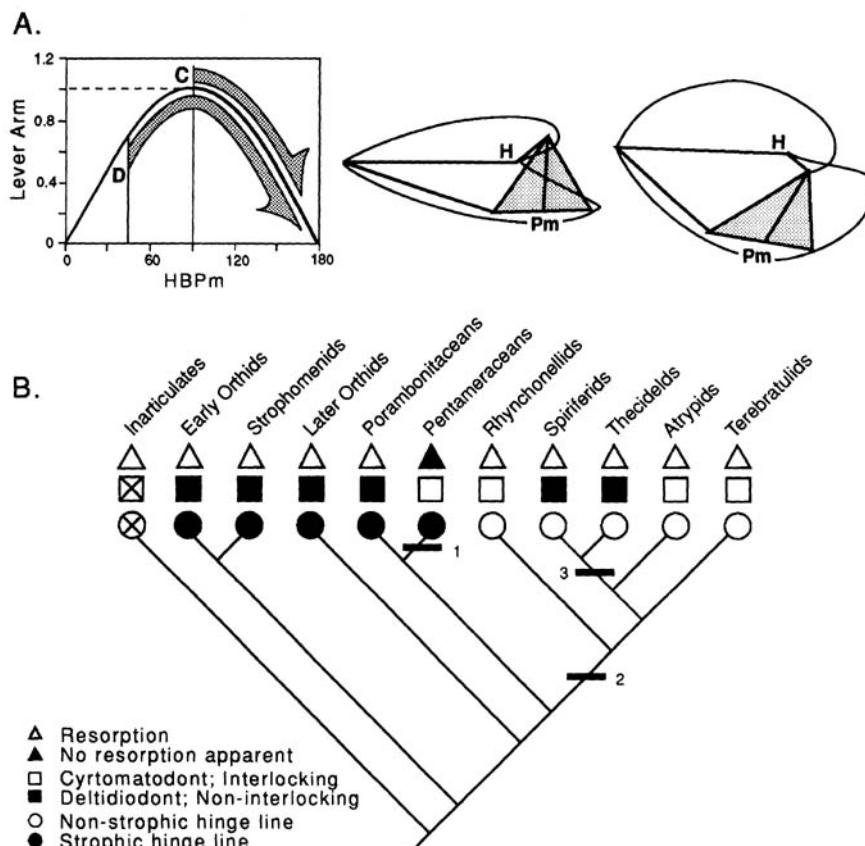


Fig. 12. Evolution of the articulate brachiopod hinge mechanism (after Carlson 1989). (A) The relationship between the length of the lever arm and the gape angle as the animal opens its valves from a closed position. The left diagram illustrates that among deltidiodonts (D), as gape increases, lever arm first increases and then decreases; among cyrtomatodonts (C), lever arm decreases. The middle diagram shows the hinge system geometry of an early deltidiodont, compared with that of a later cyrtomatodont (right diagram). (B) Cladogram of articulate brachiopods with the distribution of three binary characters (two different hinge structures, two types of hinge line, and presence or absence of resorption of any part of the shell) mapped. The analysis reveals that deltidiodonts are a paraphyletic group, that 'functional discontinuities' do not elucidate pattern of evolution (see text), and that the characters used require further investigation. Such an analysis opens new areas of investigation in evolutionary biology.

increases among the deltidiodonts are due largely to an increase in the lever arm relative to other truss elements; moment increases in cyrtomatodonts as a consequence of increase in muscle cross-sectional area. Carlson designed a simple mechanical model of the hinge mechanism that allows prediction of gape angle maxima, initial length of the lever arm, muscle force, and the point in hinge function at which the greatest muscle moment is required for both groups, and formulated predictions for the two groups. Analysis of morphology and function in the two groups indicates that each group meets its particular predictions (Fig. 12A).

The morphological and functional data, analyzed in terms of ability to resorb and reform shell material, revealed the expected differences between the groups. However, the phylogenetic implications of the study counter the proposal that resorption evolved only once, and is a shared derived character of the cyrtomatodonts. Characters were mapped on the most generally accepted tree (Fig. 12B). Accordingly, cyrtomatodonts are a monophyletic group; deltidiodonts, however, are a paraphyletic group. Carlson therefore questioned the definition of 'ability to resorb' – if it is broadly interpreted to include resorption of processes other than those of the hinge, then it occurs in most major groups, including many families, of brachiopods. Given that little is known of the physiology and metabolism of resorption, it is not clear whether 'resorption' has evolved several times and is convergent, or whether it is a developmentally plastic character subject to reversal and loss, all homoplasious conditions. Carlson predicts that further phylogenetic analysis will reveal that resorption has evolved several times because of the increased morphological flexibility that it provides.

The three studies presented in this section illustrate the contribution to evolutionary morphology of multidimensional approaches, using variously biomechanical models and analysis, behavioral (performance) data, and other aspects of the biology of recent and fossil organisms. The studies include those of single species, clades within classes, and whole phyla, and range from proposed tests of morphology, performance, and fitness as properties of pattern and process of evolution to consideration of the origins of major groups of organisms. They emphasize the need to select appropriate methods for analysis, and to provide clear identification of questions being addressed, no matter how specific or general those questions are.

8. THE PHILOSOPHIES AND METHODOLOGIES OF MORPHOLOGY

The examples that I have presented reflect the current focus on explicit methodologies and direct measurement of form and function, often in a

phylogenetic context. Several approaches are similar, but have been independently derived. All begin with careful description of structure. Structure remains fundamental to functional analysis, whether it be biomechanical, physiological, behavioral, or some combination (usually sequential) of diverse approaches. Patterns of change in structure *and* function are most appropriately analyzed by mapping the characters on a cladogram for the group of organisms in question. In addition, similarities and differences in morphology and performance among unrelated taxa can be assessed using the same methodologies, as in the *Clione*-pigeon example. Such analysis gives useful information about convergence, but not about evolution. Those studies that examine structure-function relationships within lineages provide information about, and some explanation of, patterns of evolution. Most of the new methodologies are hierarchical in their assessment of different levels of organismal organization and functions at those levels. I emphasize the advantages of the new approaches, but also indicate their limitations.

Explanations in morphology are being sought at many levels (Dullemeijer 1972; Raup 1972; Fisher 1985; Bock 1988); one is through development of ever more rigorous analytical methods. Explicit methodologies in functional morphology are largely inductive (inferring a general principle from observation of particular instances), as presented by Gans (summarized, in part, in Gans 1974), Liem (1973), and many others; some are deductive (reasoning from the general to the particular), as illustrated by Dullemeijer (1974), Zweers (1985), and Bramble and Wake (1985); yet others (perhaps the majority, at least intuitively) incorporate both components as the worker proceeds from stating the problem, to data gathering, to its analysis and formulation of conclusions (Dullemeijer, personal communication; see the Schluter and Grant example, section 4.1).

Dullemeijer and Barel (1977) compared inductive and deductive methods, characterizing the former as that of comparison of specific cases so that a classification of forms and functions is derived, from which generalization may emerge; it is theoretically independent of historical aspects, though evolution and phylogeny do enter the picture to assess the relation of form and function among related species. They point out that the inductive method has a disadvantage: 'it is theoretically impossible to falsify a supposed relation, because every observation which does not fit, does at best turn the relation into a statistical or stochastic rule. To reach a conclusion by means of the inductive method, 100% agreement is not compulsory. Furthermore any other relationship cannot be excluded . . . The inductive method hides a number of serious pitfalls . . .' I suggest that this problem with the inductive method may be at the root of Gould and Lewontin's (1979) criticism of the 'adaptationist programme.'

The deductive method may be employed following the inductive

comparison and classification, or without that as a preamble. It involves the construction of a general model that allows workers to postulate changes in the model if 'improvements' or changes are to occur in structure and function (Dullemeijer 1980). Then structural and functional analysis of taxa is done to see if in fact structural change and functional change are correlated according to the predictions based on the model. The importance of the method is that it allows the generation of hypotheses that are testable by rigorous examination of living taxa, and often of extinct ones as well. When the deductive method is used to analyze change in taxa for which there is a robust hypothesis of phylogenetic relationship, pattern of evolution can be assessed. In fact, many of the methodologies introduced as new, recently, in functional morphology, biomechanics, ecomorphology, and systematics, are implicitly, if not explicitly, deductive. This may explain some of their similarities. Consequently, we see the use of long-standing logical systems using new tools with new, and old, goals, but it does result in a more synthetic approach to problems in analysis of form–function relationships.

At several points in this essay, I alluded to criticism of the 'adaptationist programme' for its use of natural selection as a facile explanation of virtually any scenario of morphological and functional change, and Gould and Lewontin's (1979) call for a more pluralistic approach, similar to Darwin's own, to assessing the agents of evolutionary change. That pluralism is emerging today, and the integration of structural, functional, ecological, and especially developmental and phylogenetic considerations is a major stimulus for it. The contributions of morphologists are in the reintroduction of concerns about (1) whole-organism approaches (considering all of the aspects of structure and function, their interaction in the organism, and the interaction of the organism with its environment: Dullemeijer 1974), (2) the fundamentals of comparative biology (Rieppel 1988, and others already cited), and (3) architectural or constructional constraints (Seilacher 1970, 1973; Reif *et al.*, 1985, and others cited herein). An approach called structuralism has been championed recently by Goodwin (1984, 1988), Webster (1984), Webster and Goodwin (1982, 1984), and Ho and Saunders (1984). Structuralism is based on the existence of 'laws of form,' or systematic, internal constraints on the morphologies which can exist, and which might explain similarities and differences of form (Webster and Goodwin 1984).

The terms or labels for these philosophies of evolutionary biology have been used in various ways, and have led to some confusion. Reif *et al.* (1985) illustrate the definitional problem of the term and concepts of 'constructional morphology.' They show that there are three major uses of the term: one relates to the emphasis on functional and architectural integration of organic structures as analogues of machines; a second is a typological analysis of morphological similarities used to produce a

hierarchical order of taxa (this is devoid of deduction from theory of descent or causal explanations of evolution); the third is that of Seilacher (1970, 1973), whose premise was that a functional interpretation was a necessary but not sufficient explanation of organismal form. Seilacher emphasizes the importance of principles of morphogenetic fabrication, and that they introduce a component that is independent of the action of selection to produce form. The laws of geometry, materials of composition, and growth processes give rise to patterns that may be non-adaptive. This provides what Seilacher (and others) consider to be a more complete and more flexible paradigm for the analysis of form, similar to that of the approach of the Leiden school (summarized in Dullemeijer 1974) and that of Gutmann (summarized 1981, 1988). The framework integrates functional and constructional morphology with ecology in a context that Reif *et al.* (1985) view as explicitly evolutionary. Constructional morphology holds that evolutionary change is opportunistic, but channelled by historical, functional, constructional, and morphogenetic constraints (see also Gould 1980). Reif *et al.* view constructional morphology as entirely consistent with neo-Darwinian selection theory, and that formal analysis of constraints adds to understanding of pattern and process of evolution. However, the leaders of the 'structuralist' approach believe that neo-Darwinism is of extremely limited explanatory power in assessment of the problem of form, and that genetic, developmental, and environmental concepts render neo-Darwinism outmoded (Ho and Saunders 1984; Webster and Goodwin 1984). They see structuralism as an alternative to atomism, holism, and functionalism (adaptationism). Goodwin (1988) argues for a replacement of neo-Darwinism by a 'rational biology' founded on principles of form and its generation. Reif *et al.* (1985) see structuralism as equivalent to their constructional approach, and state that the approach is compatible with neo-Darwinism, but acts to extend its tenets. I, too (Wake, M. 1986), consider structuralism not a new paradigm, but an extension of the current one. Neo-Darwinists have rejected pan-adaptationism and have become pluralistic in approach. That structuralism makes evolutionary theory more complete, rather than more inclusive, has not yet been demonstrated 5 years after I first made this statement (Wake, M. 1986). Others share this view; the philosopher of science Grene (1990) notes that Kauffman (1985, 1990) uses the same sort of developmental data as Goodwin, Webster, Ho, and Saunders, in addition to other data, to provide for an *expansion* of neo-Darwinism. Grene predicts that 'a Kauffman-like expansion rather than a Goodwin-like replacement . . . will occur.'

Is there a dichotomy between structuralists and functionalists? Obviously, Reif *et al.* (1985), Wake, M. (1986), Kauffman (1985; 1990), and Grene (1990) think not, and perceive a continuum that is extended by

structuralism. On the other hand, Baron (1991) argues that neo-Darwinian adaptationism/functionalism and structuralism have completely separate domains and explanatory capacities. He asserts that functionalism, with its emphasis on selection, is a populational construct, and structuralism, with its emphasis on form and how it is generated, is an organismal construct. The organismal/structural approach is intended to explain the particular forms *realized* in evolution, not how selection has 'chosen' them to alter population composition. Baron sees the possibility of a coexistence of the advocates of organismal and populational approaches, but not a reconciliation, because their 'realms of inquiry are distinct in subject, method and explanatory content.'

Wake and Larson (1987) analyzed the degree to which the 'functionalist' and 'structuralist' perspectives are exclusive or contradictory. They compared the tenets of neo-Darwinism and structuralism, the former viewing form as contingent on many unique events that occur sequentially through time, the latter on time-independent structural properties, such as the 'rules' governing development of the tetrapod limb (Shubin and Alberch 1986). Structuralists emphasize wholeness, while neo-Darwinists search for genic-level bases for variation and then measure population-level forces such as selection. Wake and Larson urged that multi-dimensional analyses that combine elements of structuralism and neo-Darwinism be used to test the generality of evolutionary patterns and processes in other organisms. The explanatory power of both approaches can thereby be combined to produce an understanding of the evolution of lineages of organisms.

My own opinion lies somewhere between Baron's (1991) and the advocates of a strictly structuralist approach to the analysis of form, and that of Wake and Larson (1987) that the approaches can be used interactively with a common goal. On the one hand, I believe that the two approaches can indeed be brought to bear to understand the evolution of lineages of organisms. On the other, I agree with Baron that workers must make explicit the kinds of questions they are asking and the methods they are using, so that the relationship of realized morphologies and their properties of composition and generation (development) to individual- and population-level questions of performance and fitness is clear. Without explicit determinations of goals of study and of methodology, explanations of pattern and process of evolution, especially involving the causation that both morphologists and evolutionists seek, will remain equivocal.

Such debates are healthy for evolutionary morphology. Not only do they stimulate diversity and clarity of approach, but they give credence to the prominent role of morphology in programs of modern integrative and evolutionary biology. The primacy of the organism is coupled with new attention to methodology, new techniques derived from many areas of science, use of direct measurement and quantification (as in many

fields), and cognizance of the contribution of an historical/phylogenetic framework to analysis of questions of form and function. Though there may be redundancy in the new methodologies proposed in the sub-disciplines of morphology, their very similarities will facilitate better communication among workers once isolated.

Morphologists are making major contributions to evolutionary morphology in many ways. They are sorting out the behavior–performance–structure complex (so that we see that behavioral change does not necessarily precede morphological modification, and that components of morphology, e.g. bones and muscles vs. motor patterns, need not alter in concert). They are ascertaining the functional capacities of structures through biomechanical techniques (measurements of composition and performance of structural elements, and of the physical forces generated within an organism, as well as those it faces in nature). They are showing what controls structure and its variation, and what properties are intrinsic and limiting for constructional, developmental, and historical (phylogenetic) reasons (bone is bone; it doesn't do what chitin does). At the same time, they are giving a better explanation of the nature of adaptation through understanding of the relationship of structure to performance to fitness. They are providing a much clearer analysis of the relationships of morphology and function to the ecology of organisms. They are contributing to phylogenetic analysis, as well as using it. A new view of the explanatory power of morphology in all its expressions is emerging, and it is contributing both empiricism and theory to evolutionary biology.

ACKNOWLEDGMENTS

I thank Kipp Baron, David Carrier, Piet Dullemeijer, Douglas Futuyma, Carl Gans, David Wake, Gert Zweers, and an anonymous reviewer for their constructive and critical review of the manuscript. Needless to say, I trust they do not agree with all of the opinions that I present, or the format that I use, but their contributions, in general and to the manuscript, have helped me form my ideas. The National Science Foundation has supported my research, and that of many others, in evolutionary and functional morphology. I hope that this contribution will help the Foundation to see what it has supported, and why it should continue and expand that support. Finally, I thank 25 years' worth of students, undergraduate and graduate, for being both stimulating and provocative about ideas in morphology, and for making the study of evolutionary morphology a joy.

REFERENCES

- Alberch, P. (1980). Ontogenesis and morphological diversification. *Amer. Zool.* **20**, 653–67.
- (1989). The logic of monsters: evidence for internal constraint in development and evolution. *Geobios: Mem. Spec.* **12**, 21–57.
- , Gould, S. J., Oster, G. F. and Wake, D. B. (1979). Size and shape in ontogeny and phylogeny. *Paleobiol.* **5**, 296–317.
- Alexander, R. McN. (1975). *Biomechanics*. Chapman and Hall, London.
- Arnold, S. J. (1983). Morphology, performance, and fitness. *Amer. Zool.* **23**, 347–61.
- and Bennett, A. F. (1988). Behavioural variation in natural populations. V. Morphological correlates of locomotion in the garter snake *Thamnophis radix*. *Biol. J. Linn. Soc.* **34**, 175–90.
- Baron, C. J. (1991). What functional morphology cannot explain: a morphogenetic model of sea urchins and a discussion of the role of morphogenetic explanations in evolutionary morphology. In *The unity of evolutionary biology* (ed. E. C. Dudley), pp. 471–88. Dioscorides Press, Portland, OR.
- Baum, D. A. and Larson, A. (1991). Adaptation reviewed: a phylogenetic methodology for studying character evolution. *Syst. Zool.* **40**, 1–18.
- Bennett, A. F. (1989). Integrated studies of locomotor performance. In *Complex organismal functions: integration and evolution in vertebrates* (ed. D. B. Wake and G. Roth), pp. 191–202. John Wiley, New York.
- Benson, K. R. (1989). Biology's phoenix: historical perspectives on the importance of the organism. *Amer. Zool.* **29**, 1067–74.
- Block, B. A. (1991). Evolutionary novelties: how fish have built a heater out of muscle. *Amer. Zool.* **31**, 726–42.
- Bock, W. J. (1980). The definition and recognition of biological adaptation. *Amer. Zool.* **20**, 217–27.
- (1988). The nature of explanations in morphology. *Amer. Zool.* **28**, 205–15.
- (1989). Organisms as functional machines: a connectivity explanation. *Amer. Zool.* **29**, 1119–32.
- and von Wahlert, G. (1965). Adaptation and the form-function complex. *Evolution* **19**, 269–99.
- Bonner, J. T. (1982). *Evolution and development*. Springer Verlag, Berlin.
- Bramble, D. M. and Jenkins, Jr, F. A. (1989). Structural and functional integration across the reptile-mammal boundary: the locomotor system. In *Complex organismal functions: integration and evolution in vertebrates* (ed. D. B. Wake and G. Roth), pp. 133–46. John Wiley, Chichester.
- and Wake, D. B. (1985). Feeding mechanisms of lower tetrapods. In *Functional vertebrate morphology*, (ed. M. Hildebrand, D. M. Bramble, K. F. Liem, and D. B. Wake), pp. 230–61. Harvard University Press, Cambridge, MA.
- Brooks, D. R. and McLennan, D. A. (1991). *Phylogeny, ecology, and behavior*. University of Chicago Press, Chicago.
- Carlson, S. J. (1989). The articulate brachiopod hinge mechanism: morphological and functional variation. *Paleobiol.* **15**, 364–86.

- Carrier, D. R. (1987). The evolution of locomotor stamina in tetrapods: circumventing a mechanical constraint. *Paleobiol.* **13**, 326–41.
- (1991). Conflict in the hypaxial musculo-skeletal system: documenting an evolutionary constraint. *Amer. Zool.* **31**, 644–54.
- Cheverud, J. M. (1982). Phenotypic, genetic, and environmental morphological integration in the cranium. *Evolution* **36**, 499–516.
- Coddington, J. A. (1988). Cladistic tests of adaptational hypotheses. *Cladistics* **4**, 3–22.
- Coleman, W. (1980). Morphology in the evolutionary synthesis. In *The evolutionary synthesis: perspectives on the unification of biology* (ed. E. Mayr and W. Provine), pp. 174–180. Harvard University Press, Cambridge, MA.
- Collazo, A. (1990). Development and evolution in plethodontid salamanders. Unpublished PhD Dissertation, University of California, Berkeley.
- Cracraft, J. (1981). The use of functional and adaptive criteria in phylogenetic systematics. *Amer. Zool.* **21**, 21–36.
- de Beer, G. R. (1930). *Embryology and evolution*. Clarendon Press, Oxford.
- (1940). *Embryos and ancestors*. Oxford University Press, Oxford.
- de Cock Buning, T., Otten, B. and Dullemeijer, P. (1985). Some methodological remarks. *Acta Biotheor.* **34**, 277–88.
- Denny, M. W. (1990). Terrestrial versus aquatic biology: the medium and its message. *Amer. Zool.* **30**, 111–22.
- Dial, K. P., Kaplan, S. R., Goslow, G. E. and Jenkins, F. A. (1987). Structure and neural control of the pectoralis in pigeons: implications for flight mechanics. *Anat. Rec.* **218**, 284–7.
- Dudley, R. (1991). Comparative biomechanics and the evolutionary diversification of flying insect morphology. In *The unity of evolutionary biology* (ed. E. C. Dudley), pp. 503–14. Dioscorides Press, Portland, OR.
- Dullemeijer, P. (1972). Explanation in morphology. *Acta Biotheor.* **21**, 260–73.
- (1974). *Concepts and approaches in animal morphology*. Van Gorcum and Co., Assen.
- (1980). Functional morphology and evolutionary biology. *Acta Biotheor.* **29**, 151–250.
- and Barel, C. D. N. (1977). Functional morphology and evolution: In *Major patterns in vertebrate evolution*. (ed. M. K. Hecht, P. C. Goody, and B. M. Hecht), pp. 83–117. Plenum, New York.
- Duncker, H. -R. (1989). Structural and functional integration across the reptile–bird transition. In *Complex organismal functions: integration and evolution in vertebrates* (ed. D. B. Wake and G. Roth), pp. 147–70. John Wiley, Chichester.
- Emerson, S. B. (1991). The ecomorphology of Bornean tree frogs (family Rhacophoridae). *Zool. J. Linn. Soc.* **101**, 337–57.
- and Arnold, S. J. (1989). Intra- and interspecific relationships between morphology, performance, and fitness. In *Complex organismal functions: integration and evolution in vertebrates* (ed. D. B. Wake and G. Roth), pp. 295–314. John Wiley, Chichester.
- Fisher, D. C. (1985). Evolutionary morphology: beyond the analogous, anecdotal, and the ad hoc. *Paleobiol.* **11**, 120–38.
- Funk, V. A. and Brooks, D. R. (1990). Phylogenetic systematics as the basis of comparative biology. *Smithsonian Contributions to Botany* **73**, 1–45.

- Gans, C. (1965). Morphology, functional. In *McGraw-Hill Yearbook Science and Technology*, pp. 271–2. McGraw-Hill, New York.
- (1966). Some limitations and approaches to problems in functional anatomy. *Fol. Biotheor.* **6**, 41–50.
- (1974). *Biomechanics, an approach to vertebrate biology*. J. B. Lippencott, Philadelphia.
- (1985). Vertebrate morphology: tale of a phoenix. *Amer. Zool.* **25**, 689–94.
- (1988). Adaptation and the form–function relation. *Amer. Zool.* **28**, 681–97.
- (1989). Stages in the origin of vertebrates: analysis by means of scenarios. *Biol. Rev.* **64**, 221–68.
- and Northcutt, R. G. (1983). Neural crest and the origin of vertebrates: a new head. *Science* **220**, 268–74.
- and Northcutt, R. G. (1985). Neural crest: the implications for comparative anatomy. In *Functional morphology of vertebrates* (ed. H. -R. Duncker and G. Fleischer), *Fortschr. Zool.* **30**, 507–14. G. Fischer Verlag, Stuttgart.
- Ghiselin, M. T. (1980). Failure of morphology to assimilate Darwinism. In *The evolutionary synthesis: perspectives on the unification of biology* (ed. E. Mayr and W. Provine), pp. 181–92. Harvard University Press, Cambridge, MA.
- Goodrich, E. S. (1924). *Living organisms: their origin and evolution*. Oxford University Press, Oxford.
- (1930). *Studies on the structure and development of vertebrates*. Macmillan, London.
- Goodwin, B. C. (1984). Changing from an evolutionary to a generative paradigm in biology. In *Evolutionary theory: paths into the future* (ed. J. W. Pollard), pp. 99–120. John Wiley, New York.
- (1988). Morphogenesis and heredity. In *Evolutionary processes and metaphores* (ed. M. -W. Ho and S. W. Fox), pp. 145–62. John Wiley, New York.
- , Holder, N. and Wylie, C. C. (1982). *Development and evolution*. Cambridge University Press, Cambridge.
- Gould, S. J. (1977). *Ontogeny and phylogeny*. Harvard University Press, Cambridge, MA.
- (1980). The evolutionary biology of constraint. *Daedalus* **109**, 39–52.
- and Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. Roy. Soc. London B* **205**, 581–98.
- Grant, B. R. and Grant, P. R. (1989): *Evolutionary dynamics of a natural population: The large cactus finch of the Galapagos*. University of Chicago Press, Chicago.
- Greene, H. W. (1986). Diet and arboreality in the Emerald Monitor, *Varanus prasinus*, with comments on the study of adaptation. *Fieldiana: Zoology* **1370**, 1–12.
- Grene, M. (1990). Is evolution at a crossroads? In *Evolutionary biology*, vol. 24 (ed. M. K. Hecht, B. Wallace, and R. J. MacIntyre), pp. 51–82. Plenum, New York.
- Gutmann, W. F. (1981). Relationships between invertebrate phyla based on functional–mechanical analysis of the hydrostatic skeleton. *Amer. Zool.* **21**, 63–81.
- (1988). The hydraulic principle. *Amer. Zool.* **28**, 257–66.

- Hanken, J. and Summers, C. H. (1989). Skull development during amphibian metamorphosis: III. Role of thyroid hormone in chondrogenesis. *J. Expl. Zool.* **246**, 156–70.
- Herring, S. W. (1988). Introduction: How to do functional morphology. *Amer. Zool.* **28**, 189–92.
- Hickman, C. S. (1988). Analysis of form and function in fossils. *Amer. Zool.* **28**, 775–93.
- (1991). Functional analysis and the power of the fourth dimension in comparative evolutionary studies. In *The unity of evolutionary biology* (ed. E. C. Dudley), pp. 548–54, Dioscorides Press, Portland, OR.
- Hinchliffe, J. R. (1989). Reconstructing the archetype: innovation and conservatism in the evolution and development of the pentadactyl limb. In *Complex organismal functions: integration and evolution in vertebrates* (ed. D. B. Wake and G. Roth), pp. 171–89. John Wiley, Chichester.
- Ho, M.-W. and Saunders, P. T. (1984). *Beyond neo-Darwinism: an introduction to the new evolutionary paradigm*. Academic Press, New York.
- Homberger, D. G. (1988). Models and tests in functional morphology: the significance of description and integration. *Amer. Zool.* **28**, 217–29.
- Huxley, J. (1954). The evolutionary process. In *Evolution as a process* (ed. J. Huxley, A. C. Hardy, and E. B. Ford), pp. 1–23. George Allen and Unwin, London.
- Kauffman, S. (1985). Self-organization, selective adaptation, and its limits: a new pattern of inference in evolution and development. In *Evolution at a crossroads: the new biology and the new philosophy of science* (ed. D. J. Depew and B. H. Weber), pp. 169–207. MIT Press, Cambridge.
- (1990). *Origins of order: self-organization and selection in evolution*. Oxford University Press, Oxford.
- Kingsolver, J. G. and Koehl, M. A. R. (1985). Aerodynamics, thermoregulation, and the evolution of insect wings: differential scaling and evolutionary change. *Evolution* **39**, 488–504.
- Kluge, A. G. (1989). A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* **38**, 7–25.
- Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* **33**, 402–16.
- (1980). Sexual dimorphism, sexual selection and adaptation in polygenic characters. *Evolution* **34**, 292–305.
- (1986). The dynamics of peak shifts and the pattern of morphological evolution. *Paleobiol.* **12**, 343–54.
- and Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution* **37**, 1210–26.
- Lauder, G. V. (1981). Form and function: structural analysis in evolutionary morphology. *Paleobiol.* **7**, 430–42.
- (1982). Historical biology and the problem of design. *J. Theor. Biol.* **97**, 57–67.
- (1990). Functional morphology and systematics: studying functional patterns in an historical context. *Ann. Rev. Ecol. Syst.* **21**, 317–40.
- (1991). Biomechanics and evolution: integrating physical and historical

- biology in the study of complex systems. In *Biomechanics in evolution* (ed. J. M. V. Rayner), pp. 1–20. Cambridge University Press, Cambridge.
- and Liem, K. F. (1989). The role of historical factors in the evolution of complex organismal functions. In *Complex organismal functions: integration and evolution in vertebrates* (ed. D. B. Wake and G. Roth), pp. 63–78. John Wiley, Chichester.
- Liem, K. F. (1973). Evolutionary strategies and morphological innovations: cichlid pharyngeal jaws. *Syst. Zool.* **22**, 425–41.
- (1989). Functional morphology and phylogenetic testing within the framework of symecomorphosis. *Acta Biotheor.* **27**, 119–31.
- and Wake, D. B. (1985). Morphology: current approaches and concepts. In *Functional vertebrate morphology* (ed. M. Hildebrand, D. M. Bramble, K. F. Liem and D. B. Wake), pp. 366–77. Harvard University Press, Cambridge, MA.
- Maynard Smith, J. (1991). Dinosaur dilemma. *New York Rev. Books* **38**, 5–7.
- , Burian, R., Kauffman, S., Alberch, P., Campbell, J., Goodwin, B., Lande, R., Raup, D., and Wolpert, L. (1985). Developmental constraints and evolution. *Quart. Rev. Biol.* **60**, 265–87.
- Mayr, E. (1980). Introduction (morphology). In *The evolutionary synthesis: perspectives on the unification of biology* (ed. E. Mayr and W. B. Provine), p. 173. Harvard University Press, Cambridge, MA.
- (1983). How to carry out the adaptationist program? *Amer. Nat.* **121**, 324–34.
- and Provine, W. (1980). *The evolutionary synthesis: perspectives on the unification of biology*. Harvard University Press, Cambridge, MA.
- Medawar, P. B. and Medawar, J. S. (1977). *The life sciences*, pp. 19–20. Harper and Row, New York.
- Müller, G. B. (1991). Experimental strategies in evolutionary embryology. *Amer. Zool.* **31**, 605–15.
- , Wagner, G. P. and Hall, B. K. (1989). Experimental vertebrate embryology and the study of evolution: Report of a workshop. In *Trends in vertebrate morphology* (ed. H. Splechtna and H. Hilgers), pp. 299–303. G. Fischer, Stuttgart.
- Niklas, K. J. (1989). Mechanical behavior of plant tissues as inferred from the theory of pressurized cellular solids. *Amer. J. Bot.* **76**, 929–37.
- (1990). Biomechanics of *Psilotum nudum* and some early Paleozoic vascular sporophytes. *Amer. J. Bot.* **77**, 590–606.
- and Kaplan, D. R. (1991). Biomechanics and the adaptive significance of multicellularity in plants. In *The unity of evolutionary biology* (ed. E. C. Dudley), pp. 489–502. Dioscorides Press, Portland, OR.
- Northcutt, R. G. and Gans, C. (1983). The genesis of neural crest and epidermal placodes: a reinterpretation of vertebrate origins. *Quart. Rev. Biol.* **58**, 1–28.
- Parks, A. L., Parr, B. A., Chin, J. -E., Leaf, D. S. and Raff, R. A. (1988). Molecular analysis of heterochronic changes in the evolution of direct developing sea urchins. *J. Evol. Biol.* **1**, 27–44.
- Raff, R. A. (1987). Constraint, flexibility; and phylogenetic history in the evolution of direct development in sea urchins. *Dev. Biol.* **119**, 6–19.
- and Wray, G. A. (1989). Heterochrony: developmental mechanisms and evolutionary results. *J. Evol. Biol.* **2**, 409–34.

- Raup, D. M. (1972). Approaches to morphologic analysis. In *Models in paleobiology* (ed. T. J. Schopf), pp. 28–44. W. H. Freeman, San Francisco.
- Rayner, J. V. M. (1991). *Biomechanics in evolution*. Cambridge University Press, Cambridge.
- Reif, W. -E., Thomas, R. D. K. and Fischer, M. S. (1985). Constructional morphology: the analysis of constraints in evolution. *Acta Biotheor.* **34**, 233–48.
- Rieppel, O. (1988). *Fundamentals of comparative biology*. Birkhauser, Basel.
- Roth, G. and Wake, D. B. (1985). Trends in the functional morphology and sensorimotor control of feeding behavior in salamanders: an example of internal dynamics in evolution. *Acta Biotheor.* **34**, 175–92.
- (1989). Conservatism and innovation in the evolution of feeding in vertebrates. In *Complex organismal functions: integration and evolution in vertebrates* (ed. D. B. Wake and G. Roth), pp. 7–21. John Wiley, Chichester.
- Satterlie, R. A., Goslow, G. E. and Reyes, A. (1990). Two types of striated muscle are related to two-ganged swimming in the pteropod mollusc *Clione limacina*. *J. Expt. Zool.* **255**, 131–40.
- Schlüter, D. (1989). Bridging population and phylogenetic approaches to the evolution of complex traits. In *Complex organismal functions: integration and evolution in vertebrates* (ed. D. B. Wake and G. Roth), pp. 79–95. John Wiley, Chichester.
- and Grant, P. R. (1984). Determinants of morphological patterns in communities of Darwin's finches. *Amer. Nat.* **123**, 175–96.
- Seilacher, A. (1970). Arbeitskonzept zur Konstruktionsmorphologie. *Lethaia* **3**, 393–5.
- (1973). Fabricational noise in adaptive morphology. *Syst. Zool.* **22**, 451–65.
- Shine, R. (1989). Ecological influences on the evolution of vertebrate viviparity. In *Complex organismal functions: integration and evolution in vertebrates* (ed. D. B. Wake and G. Roth), pp. 263–78. John Wiley, Chichester.
- Shubin, N. H. and Alberch, P. (1986). A morphogenetic approach to the origin and basic organization of the tetrapod limb. *Evol. Biol.* **20**, 319–87.
- Slatkin, M. and Kirkpatrick, M. (1986). Extrapolating quantitative genetic theory to evolutionary problems. In *Evolutionary genetics of invertebrate behavior* (ed. M. Huettel), pp. 283–94. Plenum, New York.
- Vincent, J. F. V. and Currey, J. D. (1981). *Mechanical properties of biological materials*. Cambridge University Press, Cambridge.
- Vogel, S. (1981). *Life in moving fluids*. Princeton University Press, Princeton.
- (1988). *Life's devices: the physical world of animals and plants*. Princeton University Press, Princeton.
- Vrba, E. S. (1989). What are the biotic hierarchies of integration and linkage? In *Complex organismal functions: integration and evolution in vertebrates* (ed. D. B. Wake and G. Roth), pp. 379–401. John Wiley, Chichester.
- Wainwright, P. C. (1987). Biomechanical limits to ecological performance: mollusc-crushing by the Caribbean hogfish, *Lachnolaimus maximus* (Labridae). *J. Zool. London* **213**, 283–97.
- (1991). Ecomorphology: experimental functional anatomy for ecological problems. *Amer. Zool.* **31**, 680–93.
- , Lauder, G. V., Osenberg, C. W. and Mittelbach, G. G. (1991). The functional basis of intraspecific trophic diversification in sunfishes. In *The unity*

- of evolutionary biology.* (ed. E. C. Dudley), pp. 515–29. Dioscorides Press, Portland, OR.
- Wainwright, S. A., Biggs, W. D., Currey, J. D. and Gosline, J. M. (1976). *Mechanical design in organisms.* Edward Arnold, London.
- (1988). *Axis and circumference: the cylindrical shape of plants and animals.* Harvard University Press, Cambridge, MA.
- Waisbren, S. J. (1988). The importance of morphology in the evolutionary synthesis as demonstrated by the contributions of the Oxford group: Goodrich, Huxley, and de Beer. *J. Hist. Biol.* **21**, 291–330.
- Wake, D. B. (1982a). Functional and evolutionary biology. *Persp. Biol. Med.* **25**, 603–20.
- (1982b). Functional and developmental constraints and opportunities in the evolution of feeding systems in urodeles. In *Environmental adaptation and evolution* (ed. D. Mossakowski and G. Roth), pp. 51–66. Gustav Fischer Verlag, Heidelberg.
- (1991). Homoplasy: the result of natural selection, or evidence of design limitations? *Am. Nat.* **138**, 543–67.
- and Larson, A. (1987). Multidimensional analysis of an evolving lineage. *Science* **238**, 42–8.
- and Roth, G. (1989). The linkage between ontogeny and phylogeny in the evolution of complex systems. In *Complex organismal functions: integration and evolution in vertebrates* (ed. D. B. Wake and G. Roth), pp. 361–77. John Wiley, Chichester.
- Hanken, J., Mabee, P., and Wagner, G. (1991). Developmental biology and evolution. In *The unity of evolutionary biology* (ed. E. C. Dudley), pp. 582–8. Dioscorides Press, Portland, OR.
- Wake, M. H. (1986). Book review of *Beyond neo-Darwinism. An introduction to the new evolutionary paradigm.* By M. -W. Ho and P. T. Saunders (eds.) *Amer. Zool.* **26**, 289–90.
- (1990). The evolution of integration of biological systems: an evolutionary perspective through studies of cells, tissues, and organs. *Amer. Zool.* **30**, 897–906.
- (1991). The impact of research in functional morphology and biomechanics on studies of evolution. In *The unity of evolutionary biology* (ed. E. C. Dudley), pp. 555–7. Dioscorides Press, Portland, OR.
- (1993). Non-traditional characters in the assessment of caecilian phylogenetic relationships. *Herpet. Monogr.* MS subm.
- Webster, G. (1984). The relations of natural forms. In *Beyond neo-Darwinism: an introduction to the new evolutionary paradigm* (ed. M. -W. Ho and P. T. Saunders), pp. 193–217. Academic Press, London.
- and Goodwin, B. C. (1982). The origin of species: a structuralist approach. *J. Soc. Biol. Struc.* **5**, 15–42.
- (1984). A structuralist approach to morphology. *Rev. Biol.* **77**, 503–31.
- Welsford, I. G., Meyers, R. A., Wilson, D. S., Satterlie, R. A. and Goslow, G. E. (1991). Neuromuscular organization for “wing” control in a mollusc (*Clione limacina*) and a bird (*Columba livia*): parallels in design. *Amer. Zool.* **31**, 670–9.

- Wiley, E. O. (1981). *Phylogenetics: The theory and practice of phylogenetic systematics*. John Wiley, New York.
- Zweers, G. (1985). Greek classicism in living structure? Some deductive pathways in animal morphology. *Acta Biotheor.* **34**, 249–76.
- (1991). Path and space for evolution of feeding mechanisms in birds. In *The unity of evolutionary biology* (ed. E. C. Dudley), pp. 530–47. Dioscorides Press, Portland, OR.

Symbiosis in evolution

A. E. DOUGLAS

1. INTRODUCTION

This article explores the significance of symbiosis as a source of novel capabilities. It concentrates on one class of symbiosis, the endosymbioses, which have been defined as associations between species of unequal size in which the whole body of the smaller (known as the symbiont) is enclosed within the larger (the host) (Douglas and Smith 1989). The symbionts are microorganisms (prokaryotes and protists) and the hosts are eukaryotes (protists, fungi, plants, and animals). Two types of endosymbiosis are recognized: the intracellular symbioses, in which the symbionts are within host cells; and the extracellular symbioses, where the symbionts are located between closely apposed host cells, in a cavity or (in some animals) the gut lumen. The associations persist for a large part of the life-span of most hosts and, in many, the microbial symbionts are transmitted from one host generation to the next.

The evolutionary significance of endosymbioses is two-fold. Firstly, the microbial symbionts in many (but not all) endosymbioses possess a metabolic capability that their hosts lack. Symbiosis can therefore be viewed as a means by which eukaryotes may acquire metabolic capabilities from phylogenetically distant organisms (i.e. their microbial symbionts), or, less commonly, as a means to enhance pre-existing capabilities. Secondly, intracellular symbionts have probably given rise to key organelles of eukaryotes, the mitochondria and plastids. Symbionts and symbiont-derived organelles have thus dramatically expanded the metabolic repertoire of eukaryotes. However, not all eukaryotes possess symbionts or symbiont-derived organelles, and one aim of this article is to establish the factors that facilitate and constrain the evolution of symbionts and organelles.

2. SYMBIONTS AS A SOURCE OF METABOLIC CAPABILITIES

This section reviews seven metabolic capabilities acquired by eukaryotes through symbiosis (Table 1). These capabilities are novel for virtually all the hosts. For each capability, the characteristics of the symbioses and phylogenetic range of the organisms involved are discussed in the context

Table 1
A survey of endosymbioses in which the host acquires a novel metabolic capability from its symbionts

Metabolic capability	Symbiont taxa	Host taxa	Comments
Photosynthesis	<i>Chlorella</i>	Freshwater protists and invertebrates	Most hosts have high surface area/volume relationship to trap sufficient light to maintain high photosynthetic rates per unit total biomass
	<i>Trebouxia/ Pseudotrebouxia</i>	Lichenized fungi	
	<i>Symbiodinium</i>	Marine protists and invertebrates	
	Various cyanobacteria (e.g. <i>Nostoc, Prochloron</i> , sponges, ascidians 'cyanelles')	Lichenized fungi, protists, marine sponges, ascidians	
Chemosynthesis	Gamma-Proteobacteria	All Pogonophora, some bivalve mollusks, and nematodes	Symbioses restricted to zones of mixing between oxic and anoxic environments. Most hosts are large and have low metabolic rate because chemosynthesis is inefficient
Nitrogen fixation	<i>Rhizobia</i>	Many legumes	Symbioses not widespread because nitrogen fixation is energetically costly and requires protection from oxygen
	<i>Frankia</i>	Various dicots	
	<i>Nostoc/ Anabaena</i>	Few plants and lichens	
	<i>Teredinibacter turneri</i>	Teredinid bivalves	
	Various bacteria (e.g. <i>Enterobacter, Citrobacter</i>)	Few termites	
Methanogenesis	<i>Methanobacterium</i>	Sapropelic protists	
Synthesis of essential nutrients	Usually bacteria	Various insects, protists	Hosts have nutrient-poor or unbalanced diets

Table 1
Continued

Metabolic capability	Symbiont taxa	Host taxa	Comments
Luminescence	<i>Vibrio fisheri</i> , <i>Photobacterium phosphoreum</i> and <i>P. leiognathi</i>	Teleost fish, cephalopod mollusks	Symbioses less widespread than intrinsic luminescence. Complex morphological adaptations in host to control and direct light emission
Cellulolysis	Various bacteria, ciliate protists and phycomycete fungi	Herbivorous vertebrates	Slow process requires large volume and so most hosts are large
	Protists	'Lower' termites	Invertebrates have intrinsic cellulases

of the properties of the metabolic pathway acquired. One major group of endosymbioses, the mycorrhizal associations between fungi and plant roots, is not considered here because the fungi do not provide a novel capability but enhance the pre-existing ability of plant roots to absorb minerals and water from soils.

2.1 Provision of organic carbon: photosynthetic and chemosynthetic symbionts

The key enzyme in this capability is ribulose-1, 5-bisphosphate-carboxylase, which catalyses the first step in the incorporation of carbon from carbon dioxide into sugars. The energy and reductant for the assimilation of carbon derives from either light (photosynthesis) or the oxidation of inorganic compounds (chemosynthesis).

2.1.1 Photosynthetic symbionts

All known photosynthetic symbionts have oxygenic photosynthesis, i.e. they use water as a reductant, generating oxygen. The bacteria with non-oxygenic photosynthesis are anaerobes and are presumably excluded from the aerobic environment of most eukaryotic cells by their intolerance of molecular oxygen. Oxygenic photosynthesis is characteristic of all

cyanobacteria and prochlorons among the prokaryotes and eukaryotic plastids. Both the prochloron and plastid lineages lie within the cyanobacterial radiation (Giovannoni *et al.* 1988; van den Eynde *et al.* 1988; Turner *et al.* 1989) and it is very likely that oxygenic photosynthesis evolved once in the ancestor of the cyanobacteria (Woese 1987).

An important factor determining the rate of photosynthetic carbon fixation in symbiosis is the surface area available for light capture. Consequently, photosynthetic symbionts are generally found in hosts with a high surface area/volume relationship, especially protists and small multicellular eukaryotes with thin tissue layers. In the aquatic environment, hosts include foraminiferans, radiolarians, and ciliates among the protists, and various invertebrates, e.g. cnidarians, flatworms, sponges, and ascidians. The only large and structurally complex hosts are the mollusks; for example, in the tridacnids (giant clams), a large surface area is afforded by hypertrophied siphonal tissue extending along the dorsal (uppermost) surface. The giant reef-building rudist bivalves in the Cretaceous and some Permian brachiopods also probably had photosynthetic symbionts (Cowen 1970; Vogel 1975). The dominant associations in the terrestrial environment are the lichens, which also have a high surface area/volume relationship. Lichenized fungi include nearly half of all species of ascomycete fungi and a few species of other fungi (Hawksworth 1988a).

Within each major taxon of hosts, symbioses have probably evolved independently several times. For example, symbiotic foraminiferans today include members of 11 families, and the fossil record indicates that these associations arose at least four times in the Mesozoic and Cenozoic (Cowen 1983). Among the Cnidaria, symbionts are documented in more than 100 genera from all three classes, and taxa with and without symbionts are interspersed, indicating a complex history of multiple acquisitions (and possibly loss) of symbiosis. The taxonomic distribution of lichens among ascomycete fungi point to a 'complex network of evolving and devolving lichen associations' (Hawksworth 1988b).

A variety of algae and cyanobacteria has been documented as symbionts (Table 1), but the great majority of hosts possess one of three taxa, all of which are algae: the closely related genera *Trebouxia/Pseudotrebouxia* in 50–70 per cent of lichen species, *Chlorella vulgaris* in virtually all freshwater protists and invertebrates, and the dinoflagellate *Symbiodinium* in many marine hosts. However, in a few host groups, these algae are not the dominant symbionts. The usual symbionts in most marine sponges are cyanobacteria, and in didemnid ascidians they are prochlorons; the foraminiferans have a wide variety of symbionts, including dinoflagellates, diatoms, prasinophytes, and rhodophytes.

The distribution of certain photosynthetic symbionts can be accounted for by horizontal transfer between host taxa. Cowen (1983) has proposed

that algae (or cyanobacteria) capable of forming a symbiosis evolve very rarely, but once an association is established in one host taxon, the symbionts may spread to other organisms. Consistent with this suggestion of the evolutionary importance of horizontal transfer, several hosts are known to acquire their symbionts from other associations at each generation. For example, the widespread lichen *Xanthoria parietina* obtains its symbionts by invading the lichen *Physcia tenella* (Ott 1987), and some nudibranch mollusks acquire *Symbiodinium* by feeding on symbiotic cnidarians (Rudman 1981).

In all associations investigated, the host derives photosynthetically fixed organic compounds from living cells of the symbionts. Transfer is substantial, frequently representing more than half of the carbon fixed by the symbiont (reviewed in Douglas 1988a). In several invertebrate systems, it is sufficient to fuel the host respiratory needs, and this enables hosts to tolerate a low availability of particulate food (e.g. Frost and Williamson 1980; Wellington 1982). In lichens, symbiont-derived carbon contributes to the high internal concentration of polyols in the fungal host, essential for the lichens' tolerance of osmotic stress during frequent dessication and rewetting (Farrar 1976).

2.1.2 Chemosynthetic symbionts

Chemosynthetic symbioses were first recognized in 1980, in members of the fauna associated with deep-sea hydrothermal vents (Cavanaugh *et al.* 1981). Intensive research over the last decade has revealed further associations in benthic and interstitial invertebrates, primarily of two taxa: the pogonophorans and the bivalve mollusks. All members of the phylum Pogonophora (which comprises two groups, the Vestimentifera associated with the hydrothermal vents, and the Perviata in sediments of the continental shelf) have chemosynthetic symbionts, located in cells of the 'trophosome' tissue which fills most of the trunk. The associations in bivalves have probably evolved independently at least four times, once in the subclass Protobranchia (e.g. *Solemya*) and three times in the subclass Lamellibranchia (Southward 1987). In all bivalves, the bacteria are retained, usually intracellularly, at the periphery of the gills. Chemosynthetic bacteria have also been demonstrated between the epidermis and cuticle of the oligochaete annelid *Phalodrilus* (Giere *et al.* 1988), and on the surface of interstitial nematodes of the subfamily Stilbonematinae (Schiemer *et al.* 1990). Only one protist host has been documented, the interstitial ciliate *Kentrophoros*, which bears the symbionts externally on its dorsal surface, probably embedded in mucus (Fenchel and Finlay 1989).

The fossil record of chemosynthetic symbioses is very poor, but both vestimentiferan pogonophorans and symbiotic bivalves were probably present in the Cretaceous (Hamyon *et al.* 1984; Beauchamp *et al.* 1989).

Very little is known about the chemosynthetic symbionts because they have not been brought into axenic culture. Most symbionts examined are sulphur oxidizers. Many, for example in *Riftia*, use sulphide (Wilmot and Vetter 1990); others, for example in the bivalve *Solemya*, use thiosulphate (derived from the oxidation of sulphide by the host mitochondria) (Powell and Somero 1986; O'Brien and Vetter 1990). A few bivalves have methane-oxidizing bacteria (Cavanaugh *et al.* 1987). The taxonomic position of most symbionts is obscure, but the sulphur-oxidizing bacteria in the pogonophoran *Riftia pachyptila* and several bivalves have been assigned to the gamma subdivision of Proteobacteria from their 16S rRNA sequences (Distel *et al.* 1988). Although the symbionts are more closely related to each other than any free-living bacteria, the analyses indicate that more than one lineage have been acquired by bivalves and that *Riftia* may have obtained its symbionts by horizontal transfer from bivalves.

Many animal hosts have greatly reduced (and apparently non-functional) digestive systems, and presumably do not feed. It is widely agreed that the chemosynthetic symbionts provide most, or all, of their carbon requirements. This view is supported by the stable carbon isotope ratios in most associations (Spiro *et al.* 1986), the transfer of carbon compounds fixed by the symbionts to the host tissues in the bivalve *Solemya* (Fisher and Childress 1986), and the net growth of an un-named bivalve symbiosis in the absence of particulate food (Cary *et al.* 1988).

The incidence of chemosynthetic symbioses is determined largely by two characteristics of chemosynthesis.

(1) Chemosynthetic bacteria are limited to habitats containing both the reduced inorganic substrate (such as sulphide or methane) and molecular oxygen. Consequently, both free-living chemosynthetic bacteria and the symbioses are restricted to regions of contact between anoxic and oxic environments. These include the zone of discontinuity between oxygen-rich, upper layers and deeper, anoxic layers of marine sediments; regions of mixing in the marine water column between upwelling reduced waters and the surrounding oxic sea water, as in hydrothermal vents or hydrocarbon seeps; and waste disposal sites, such as sewage outfalls and pulp mill effluents.

(2) The oxidation of inorganic compounds generates energy (ATP) very inefficiently, and consequently fuels considerably lower rates of carbon dioxide fixation than in photosynthetic organisms. Chemosynthetic symbionts would be expected to be restricted to hosts with a low requirement for carbon, i.e. a low metabolic rate. The determinants of metabolic rate in invertebrates and protists are not clearly understood (Schmidt-Nielsen 1984), but there is a general trend of decreasing metabolic rate per unit body weight with increasing body weight. Fenchel and Finlay (1989) have argued that most hosts of chemosynthetic symbionts

are large, slow-moving or sessile forms and that protist hosts are rare, for this reason. Consistent with their interpretation, the symbionts occupy 50 per cent of the biomass in the only known protist host, *Kentrophoros*, up to 35 per cent of the biomass of stilbonematid nematodes, and only 1–10 per cent of the biomass in bivalves and pogonophorans. For comparison, photosynthetic symbionts represent less than 10 per cent of the biomass of most protist and invertebrate hosts.

2.2 Nitrogen-fixing symbioses

Nitrogen fixation, or diazotrophy, refers to the reduction of nitrogen to ammonia. It has probably evolved once (Postgate and Eady 1988), before the advent of oxygenic photosynthesis (Raven and Sprent 1989), and it is widely distributed among prokaryotes, but apparently absent from eukaryotes (Sprent and Sprent 1990). The reaction is relatively uniform at the enzymological level, catalysed by the nitrogenase complex, comprising a nitrogenase and nitrogenase reductase.

Very few groups of eukaryotes have acquired the capacity to fix nitrogen by symbiosis. Most are terrestrial plants; diazotrophic associations are rare in animals and unknown in non-photosynthetic protists (Table 1).

The best known diazotrophic symbionts are the rhizobia, including the genera *Rhizobium* and *Bradyrhizobium*. They are housed in nodules, usually on the roots of members of the dicot family Leguminosae. The legumes are undoubtedly predisposed to form associations with rhizobia. The symbiosis has probably evolved several times since the family evolved in the late Cretaceous (Young and Johnston 1989), but is known in only one genus of non-legumes, *Parasponia* (= *Trema*) of the family Ulmaceae (Trinick 1973). Furthermore, the capacity to form the symbiosis is a characteristic of advanced legumes. Less than a third of genera of the primitive subfamily Caesalpinoideae are symbiotic, but more than 75 per cent and 95 per cent of the advanced subfamilies Mimosoideae and Papilionoideae, respectively, have rhizobia.

A second major diazotrophic symbiont of angiosperms is the actinomycete *Frankia*, located in clusters of lateral roots, known as actinorhizas. The symbiosis with *Frankia* has been reported in at least 220 species in 23 genera and eight families of dicots, including *Casuarina*, *Alnus*, *Myrica*, and *Eleagnus* (Newcomb and Wood 1987). The associations are restricted to woody perennials and primarily species that thrive in nitrogen-poor and stressed soils, such as early successional stages of forests, sand dunes, and wetlands. Thus, unlike the rhizobial associations, the distribution of this symbiosis is more closely linked to the habit than taxonomic position of the plant host. Normand and Bousquet (1989) have suggested that the association evolved in the Myricaceae, an ancient

and completely actinorhizal family, and the symbiotic *Frankia* were subsequently transferred to other dicot taxa. Some temperate hosts, such as *Alnus* and *Eleagnus*, may have acquired *Frankia* as they colonized nitrogen-poor soils following the retreat of the glaciers in the Pleistocene (Baker and Miller 1980).

Many angiosperms have diazotrophic bacteria (e.g. *Azospirillum*, *Azotobacter*) on the surface, and frequently in the cortex, of their roots, but the significance of nitrogen fixation by these bacteria is unclear (Sprent and Sprent 1990).

Symbiotic nitrogen-fixing cyanobacteria of the order Nostocales (e.g. *Nostoc*, *Anabaena*) have a considerably broader host range than both rhizobia and *Frankia* (see Table 1). They are known in members of all major taxa of plants from bryophytes to angiosperms, the diatom *Rhizosolenia* and some lichens (Sprent and Sprent 1990). Approximately 30 per cent of cycad species bear cyanobacteria in their lateral 'coralloid' roots, but in other host groups the associations are not widespread. For example, in angiosperms and pteridophytes, only one genus, *Gunnera* and *Azolla*, respectively, are infected, and less than 5 per cent of lichen species have cyanobacterial symbionts.

Diazotrophic bacteria are significant to the nitrogen nutrition of only two groups of animals, both of which feed exclusively on wood: all 'shipworms' (bivalve molluscs of the order Teredinidae), and some of the termites. The bacteria in shipworms are located in an organ, known as the gland of Deshayes, arising from the oesophagus. They can be cultured readily, and can both fix nitrogen and degrade cellulose (Waterbury *et al.* 1983). The 16S rRNA sequences of the bacteria in 10 different genera of shipworms are indistinguishable; and they have been assigned to a single species in a new genus, *Teredinibacter turneri*, within the gamma-Proteobacteria (Distel 1990). Nitrogen-fixing bacteria have been demonstrated in the hindgut of a variety of termites, but their significance to the insects' nitrogen nutrition varies greatly between species and with composition of the diet (Breznak 1984).

The low incidence of nitrogen-fixing symbioses probably arises from two characteristics of nitrogen fixation: first, it is an energetically costly reaction, requiring at least 16 molecules of ATP to reduce each nitrogen molecule to ammonia; and second, the nitrogenase complex is irreversibly inactivated by molecular oxygen.

The high energy requirement of nitrogen fixation undoubtedly represents a major cost of diazotrophy. For example, in both legume and actinorhizal nodules, 11–13 per cent of the photosynthetic carbon fixed by the plant is used in nodule respiration (Minchin and Pate 1973; Tjepkema 1985). The energetic cost of diazotrophic symbionts in other hosts is not known, but it is relevant that, among animals, the only known hosts, shipworms and termites, thrive on wood, which has an exceptionally low nitrogen

content (Mattson 1980). Many other animals on low-nitrogen diets have evolved alternative strategies, including the acquisition of non-diazotrophic symbionts which may increase the efficiency of nitrogen use by recycling host waste nitrogen (Smith and Douglas 1987).

Nitrogen-fixing microorganisms maintain very low intracellular oxygen tensions because of the sensitivity of nitrogenase to oxygen. This can most readily be obtained by living in anoxic environments. However, anaerobic respiration is very inefficient, and generates sufficient ATP to support only low levels of nitrogen fixation. The nitrogen fixation rates per unit of microbial biomass are generally higher in symbiosis than in free-living systems because the symbionts fix sufficient nitrogen to support the nitrogen requirements of both themselves and their hosts. It is probably as a consequence of this that nitrogen fixation fueled by anaerobic respiration is widespread in free-living diazotrophs but not in symbiosis. Thus, no anaerobic eukaryote with nitrogen-fixing symbionts is known; nitrogen fixation in the plant symbioses is abolished under low external oxygen conditions, such as in waterlogged soils (Appleby 1984; Tjepkema *et al.* 1986); and the only well-documented anaerobic diazotrophic symbionts are those in the anoxic hindgut of termites.

In plant symbioses, both host and symbiont may contribute to the protection of nitrogenase from molecular oxygen, while providing sufficient oxygen for aerobic respiration to fuel nitrogen fixation. In the cyanobacterial and *Frankia* symbioses, the protection is provided primarily by the symbionts, and in rhizobial associations, it is afforded by the host. The nitrogenase in cyanobacteria is restricted to specialized cells, heterocysts, and in *Frankia* it is located in 'symbiotic vesicles'; which are terminal swellings of the hyphae. Heterocysts and symbiotic vesicles can fix nitrogen under atmospheric oxygen tensions because they have thick, impermeable cell walls and high rates of respiratory oxygen consumption; the heterocysts, unlike other cyanobacterial cell types, additionally lack the oxygen-evolving photosystem II (Sprent and Sprent 1990). Rhizobia have no comparable protection, and the activity of nitrogenase depends on very low oxygen tensions in the nodule. These conditions are maintained largely by the paucity of air spaces in the root cortex and, in some legumes, by suberized cell walls. The high rates of aerobic respiration by the rhizobia are dependent on a plant pigment, leghemoglobin, which facilitates oxygen transport at low oxygen tensions (Appleby 1984). A few hosts of *Frankia*, e.g. *Casuarina* and *Myrica*, also have impermeable suberized or lignified cell walls and low intranodule oxygen tensions; these actinorhizas also have leghemoglobin (Tjepkema *et al.* 1986).

As a consequence of the ability of *Frankia* and cyanobacteria to fix nitrogen at atmospheric oxygen tensions, potential hosts could more readily provide a suitable environment for nitrogen fixation. This may contribute to the greater host range of these symbionts than the rhizobia.

2.3 Methanogens: symbionts as electron sinks

This section considers symbionts that contribute directly to the energy metabolism of their hosts in a fashion analogous to the mitochondria in aerobic eukaryotic cells. Mitochondria contain various metabolic pathways, including oxidative phosphorylation, by which electrons are transferred from NADH and FADH₂ to oxygen with the concomitant synthesis of ATP. The advantage to the cell is two-fold: the mitochondria act as an electron sink, regenerating NAD and FAD for further oxidative reactions; and ATP is transported to the cytoplasm and used in endergonic reactions.

There are no documented symbionts that provide ATP to their hosts. The only known symbionts that act as an electron sink are methanogens in a small number of anaerobic protists, most of which possess hydrogenosomes. (Hydrogenosomes are organelles that metabolize pyruvate (the end product of glycolysis) to acetate, with the synthesis of ATP by substrate level phosphorylation and transfer of electrons to protons, producing hydrogen (Muller 1988.) The methanogenic bacteria use the hydrogen generated by the hydrogenosomes as a substrate for methanogenesis. These symbionts have been documented in several ciliates, including *Metopus*, which live in anaerobic carbon-rich ('sapropelic') habitats (Van Bruggen *et al.* 1984).

However, the possession of methanogens and hydrogenosomes are not invariably linked. A few hydrogenosome-containing protists lack methanogens (Muller 1988), and the amoeboflagellates *Pelomyxa* and *Mastigella* have methanogens but not hydrogenosomes (Van Bruggen *et al.* 1985, 1988). Van Bruggen *et al.* (1988) have suggested that a further symbiont (a 'thick' Gram-positive rod, distinct from the 'slender' Gram-negative methanogens) in these amoeboflagellates may parallel the hydrogenosome, by providing hydrogen for the methanogens.

Significantly, there are no known symbionts whose capacity for oxidative phosphorylation is used directly by a host, as in the relationship between mitochondria and aerobic eukaryotes. The suggestions in the early literature that the symbionts in *Pelomyxa* are aerobic analogues of mitochondria were mistaken.

2.4 Symbionts that provide essential nutrients

Animals and certain protists are relatively impoverished in terms of their metabolic capabilities. For example, they cannot synthesize 10 of the constituent amino acids of proteins (the essential amino acids) and many cofactors (vitamins). Although most of these eukaryotes derive these compounds from their diet, some may obtain them from microbial symbionts.

The best-known of this type of symbiosis are the 'mycetocyte symbioses'

between insects and intracellular bacteria (the insect cells containing the symbionts are called mycetocytes). These associations are widespread or universal in four insect groups, the cockroaches, homopterans, lice, and beetles, and they are also recorded in a few dipterans and ants (Douglas 1989). The distribution of mycetocyte symbioses is well (but not perfectly) correlated with poor quality of the insect diet (Buchner 1965). For example, most Homoptera (e.g. aphids, planthoppers, whitefly) feed on plant sap (of unbalanced amino acid composition) and their symbionts probably provide essential amino acids (Douglas 1988b), and the sucking lice, which feed on vertebrate blood (deficient in B vitamins), may obtain vitamins from their symbionts (Buchner 1965). The mycetocyte symbioses have undoubtedly evolved independently many times (Douglas 1989), and the great morphological diversity of the symbionts suggests that many different bacterial taxa are involved. However, with the exception of aphid symbionts, which have been assigned to the gamma-Proteobacteria (Unterman *et al.* 1989), the phylogenetic position of the symbionts is obscure.

The bacterial symbionts of certain protists, e.g. the trypanosome *Cryptosida*, may also compensate for the loss of certain biosynthetic pathways in their hosts, by provision of amino acids, purines, vitamins, and other nutrients (Lee *et al.* 1985). Associations between phagotrophic protists and bacteria are very widespread, but poorly studied, and it is possible that many protists derive essential nutrients from their symbionts.

2.5 Luminescent symbionts

Luminescence is known in only four genera of bacteria, but it is widespread in eukaryotes, including many protists, some fungi, and animals (Campbell 1989). Biochemically, there are no substantial barriers to the evolution of luminescence. Low levels of light are emitted in many oxidative reactions, and luminescence has evolved by amplification of five distinct reactions, at least 35 times (Hastings 1983; Selinger 1987). Correlated with this, most luminescence systems in eukaryotes are intrinsic. Symbiotic light production is documented in very few animals (Table 1) most notably teleost fish and cephalopod mollusks but, even within these groups, it is less common than intrinsic systems. Among cephalopods, 19 families have luminescent members, but only two, the Loliginidae and Sepiolidae, have bacterial light production (Herring 1988). In fish, less than half of the families with luminescent members have species with bacterial luminescence, and these are in the most advanced group, the Euteleostei (Herring and Morin 1978). However, in those families/genera where bacterial luminescence is reported, it is very widespread or universal. For example, among ceratioid anglerfish, comprising 34 genera, the females of all species in 32 genera have bacterial light organs, and among

perciform fish all members of the family Leiognathidae have bacterial luminescence.

The microbial partners in all luminescence symbioses are the bacteria *Vibrio* and *Photobacterium*. These genera are widespread in the marine environment, including the surface and guts of animals, and the symbioses probably evolved from such casual associations. However, only a single symbiont species occurs in each host species, and the identity of the bacterial symbiont is linked to host taxonomy. For example, all monocentrid fish and sepiolid cephalopods have *V. fisheri*, all perciform fish and loliginids have *P. leiognathi*, and the gadiform fish have *P. phosphoreum*.

The evolution of the symbioses has involved great anatomical changes in the host, to form the light organ housing the bacterial symbionts, and accessory structures to control light emission. In both fish and cephalopods, the light organs comprise a mass of tubules in which the bacteria are densely packed. The light organs in fish are usually elaborations of gut diverticula or, less commonly, on the surface (e.g. in the orbit of the eyes of anomalopids) and, in all cephalopods, they are modified nidamental glands opening into the mantle cavity (Herring 1985, 1988). The accessory structures associated with bacterial light organs are frequently very complex for two reasons. First, bacterial light production is continuous and therefore the accessory structures are the sole means by which light emission from the animal is controlled; and, second, the maximum number of bacterial light organs is three (it is not known why) and therefore complex systems are required to generate diffuse illumination (Herring and Morin 1978; Herring 1985). As an example, light emission by leiognathid fish is controlled by a shutter linked to the single circumesophageal light organ, and a ventral glow is obtained by reflection of light from the light organ off the gas bladder into transparent muscle and through precisely oriented guanine platelets in the skin (McFall-Ngai 1983). The accessory structures for intrinsic light organs are generally simple because intrinsic light can be switched on and off rapidly, its intensity can be modulated, and multiple light organs are common.

2.6 Cellulolytic symbionts

Cellulolysis, the breakdown of cellulose to glucose, is widely regarded as central to the ability of many herbivorous animals to use structural plant material (although other enzymatic capabilities are also important, e.g. breakdown of hemicelluloses and, in high fiber diets, lignolysis to free the cellulose from lignocellulose complexes). In virtually all animals that use microorganisms to degrade plant material, the symbionts are extracellular in an anaerobic portion of the digestive tract. The microorganisms hydrolyse the cellulose and other polysaccharides to their constituent

sugars, which are then used in anaerobic respiration with short chain aliphatic compounds, chiefly the volatile fatty acids (VFAs) acetic acid, propionic acid, and butyric acid, as microbial waste products. The VFAs are absorbed into the host tissues, where they are used as substrates for aerobic respiration and various biosynthetic pathways. The region of the host gut containing the symbionts is usually greatly enlarged, representing up to 15 per cent of total body weight. This anatomical modification is related to the fact that cellulose breakdown is a very slow process, requiring large amounts of plant material to be retained in the gut for long periods.

No cellulases of vertebrate origin are known. Cellulolysis in vertebrates is mediated by microbial symbionts, which may meet up to 30 per cent of host energy requirements (McBee 1989). The associations are general among herbivorous mammals, birds, and reptiles (McBee 1977; Zimmerman and Tracey 1989), and many herbivorous dinosaurs had capacious guts, indicative of a gut microflora (Farlow 1987). In most hosts, the fermentation chamber is in the hindgut, usually in caeca (e.g. rodents, lagomorphs, grouse) or, less commonly, in the colon (e.g. horses, emu, iguanid lizards). Foregut symbioses, in enlarged regions of the oesophagus and/or stomach, are restricted to mammals. They have evolved independently at least six times; in the kangaroos, sloths, colobine monkeys, and (among the artiodactyls) the hippos, camels, and ruminants (Janis 1976; Bauchop 1977). Janis (1976) has argued that the foregut symbionts in early artiodactyls of the Eocene were not cellulolytic. They may have detoxified plant secondary compounds or recycled nitrogen, and foregut cellulolysis was acquired secondarily, possibly in the Oligocene.

Cellulolytic symbionts are more advantageous to large than small animals. Because cellulose is degraded slowly, the host has to allocate considerable space to the fermentation chamber containing the symbionts and decomposing plant material. The relationship between gut capacity and body size is isometric, but the energy required per unit of body weight decreases with increasing body size, and therefore small herbivores are less able to process sufficient plant material to support their energy demands. Consequently, small mammals such as rabbits and voles pass food through the gut more rapidly and select more digestible foods than large species, such as horses and elephants.

The gut symbioses in vertebrates are more complex than other types of symbiosis discussed in this article because they involve a wide diversity of microorganisms which contribute to host nutrition in different ways. For example, in mammals, many taxa of bacteria, ciliate protists, and a number of phycomycete fungi are involved in the transformation of cellulose and related polysaccharides to VFAs (Coleman 1980; Bauchop 1989). These and other symbiont taxa also contribute indirectly to cellulolysis by maintaining an appropriate environment in the gut,

for example by scavenging trace oxygen and methanogenesis. Some microorganisms are significant in processes unrelated to cellulolysis, e.g. water conservation and provision of essential amino acids and vitamins, and in certain hosts, such as chickens, these functions predominate and cellulose breakdown is insignificant (Mead 1989).

Cellulolytic symbioses are less widespread in invertebrates than vertebrates. Among invertebrates they are restricted largely to taxa (such as termites, cockroaches, and shipworms) feeding on very poor diets with a high fiber content; they make little or no contribution to the nutrition of most phytophagous insects, including the Orthoptera and Lepidoptera. This difference between vertebrates and insects can be linked to their difference in size (see above), and also to the ability of some invertebrates to synthesize cellulases. Thus, shipworms and termites have intrinsic cellulases, and the symbionts amplify their cellulolytic capability. The correlation between diet quality and possession of symbionts is elegantly illustrated by the symbiosis in termites. Lower termites have both intrinsic cellulases in their salivary glands and cellulolytic protists in their hindgut (Yamin and Trager 1979; Hogan *et al.*, 1988a). 'Higher' termites rely entirely on endogenous cellulases; they lack protist symbionts, and the bacteria in their hindguts are not cellulolytic (Hogan *et al.* 1988b).

2.7 Conclusions

The survey of endosymbioses indicates that eukaryotes of all major groups can form symbioses, but with a very restricted range of microbial symbionts.

The principal constraint on the acquisition of symbionts by eukaryotic hosts is that symbionts represent a cost, requiring space and nutrients. In general, it is therefore advantageous for a host to acquire microorganisms if they possess a metabolic capability that the host lacks. This is why plants lack photosynthetic symbionts, luminescence has rarely evolved intrinsically and been acquired by symbiosis in the same organism, and cellulolytic symbionts are largely restricted to vertebrates (which lack cellulases). More specific constraints relate to particular symbioses (see Table 1), and many are linked to host size. For example, small organisms with a high specific metabolic rate cannot readily accommodate diazotrophs, which are energetically expensive, or chemosynthetic symbionts, which fix carbon at very low rates; but small heterotrophs (protists and animals) with a high surface area/volume relationship are better suited than large animals to use photosynthetic symbionts.

Most eukaryotic groups include both symbiont-bearing and symbiont-free species. In many host taxa, the associations have evolved several times and some groups have lost symbionts secondarily (e.g. photosynthetic symbionts in lichens, cellulolytic and mycetocyte symbionts in insects).

This suggests that, in an evolutionary sense, the ability of hosts to form symbioses is not an important determinant of the incidence of symbiosis. The evolution and persistence of symbioses is determined primarily by the selection pressure to acquire and retain the metabolic capability that the symbiont possesses, and by the availability of appropriate symbionts.

Very few taxa of microorganisms with a certain metabolic capability are acquired as symbionts of eukaryotes, and in many cases closely similar or indistinguishable symbiont taxa have been acquired by phylogenetically diverse hosts. For example, *Chlorella vulgaris* is the dominant photosynthetic symbiont of a variety of freshwater protists and invertebrates, only three of the many species of luminescent bacteria in the oceans are known as light organ symbionts in teleost fish and cephalopods, and a very limited range of heterocystous cyanobacteria (*Nostoc/Anabaena*) are known in a variety of plants, from bryophytes to angiosperms. This suggests that very few microorganisms are *able* to form a symbiosis.

3. THE EVOLUTIONARY CONSEQUENCES OF SYMBIOSIS

The expansion of the metabolic repertoire of many eukaryotes by symbiosis has had three linked consequences: substantive morphological and physiological changes in the host to accommodate the symbionts and exploit their metabolic capabilities; the invasion of novel habitats or (for animals) the use of novel diets; and the radiation of host taxa. These three effects are considered in this section.

3.1 Morphological and other changes in the host

Many associations described in section 2 illustrate the role of symbiosis as a selective pressure for dramatic morphological and physiological changes in eukaryotes. As examples, the cellulolytic microflora and incipient herbivorous habits of many vertebrate lineages promoted the expansion and modifications in gut anatomy, diversification in dentition, and, at the biochemical level, a switch from glucose to VFAs as precursors of many metabolic pathways (Morris and Rogers 1982; Langer 1984). The acquisition of chemosynthetic bacteria by the Pogonophora (or their annelid-like progenitors) led the loss of feeding, transformation of the adult gut into the trophosome housing the bacteria, and changes in the circulatory system to enhance the transport of oxygen and sulphide (Southward 1987). However, the unusual morphological features of some hosts are probably not a consequence of symbiosis. The asymmetric growth and hypertrophied siphonal tissues of tridacnid mollusks promote light capture by the algal symbionts, but it evolved in the symbiont-free

ancestral genus *Avicularia*, probably as a means to process large volumes of water for filter-feeding (Cowen 1983).

3.2 The exploitation of novel habitats and diets

The habitats (or diets) of many eukaryotes are linked, presumably causally, to their possession of symbionts. Various invertebrates derive organic carbon from their algal symbionts, so reducing their dependence on holozoic feeding; many plants are able to live in nitrogen-poor soils through symbiosis with nitrogen-fixing bacteria, such as *Frankia* and rhizobia. However, most detailed discussion in the literature on the role of symbiosis in the exploitation of novel habitats by eukaryotes has concerned the invasion of land by fungi and plants.

An important route by which ascomycete fungi have invaded the terrestrial environment is through lichenization with algae or cyanobacteria. Although some non-symbiotic ascomycetes thrive in dry terrestrial environments, e.g. xerophilic *Aspergillus* in house dust, the lichenized fungi's tolerance of extreme desiccation and rapid changes in water content is unique, enabling them to dominate terrestrial environments in which vascular plants do not thrive (e.g. high altitudes and latitudes, bare rock and bark, some deserts). The underlying mechanisms are intimately linked to the symbiosis. The fungus is believed to use organic carbon compounds, notably polyols, derived from the photosynthetic symbionts as an osmoprotectant under desiccating conditions and to fuel high respiratory rates on rewetting. The complex morphology of lichens and some 'lichen substances,' generated only in symbiosis, also contribute to the water relations of lichens.

Much discussion of the evolution of terrestrial plants has concentrated on the difficulties early plants probably encountered in the acquisition of mineral nutrients. These plants lacked absorptive roots and the substratum, although probably biologically active, was of lower quality than modern soils (Wright 1985). Several authors, especially Pirozynski and Malloch (1975), have argued that these early plants had fungal symbionts, comparable to vesicular-arbuscular mycorrhizal (VAM) fungi, which extended into the substratum and extracted minerals from a greater volume of soil than uninfected roots. Fungi assigned to the genus *Palaeomyces* have been identified in permineralized fossils *Rhynia* and *Aglaphyton* of the early Devonian (Edwards 1986). Pirozynski and Malloch (1975) argue that these fungi were mycorrhizal symbionts. A more plausible interpretation is that they were pathogens or saprophytes. They extend into the aerial parts of the early plants, they are absent from many well-preserved specimens (Taylor and White 1989) and plant cell walls abutting the fungal hyphae appear to be breaking down

(Edwards 1986). The first unambiguous VAM fungus in the fossil record is in the Triassic cycad *Antarticycas* (Stubblefield *et al.* 1987).

3.3 The generation of novel host taxa

Bermudes and Margulis (1987) have argued that the evolution of symbiosis has been the catalyst for many of the adaptive radiations in the eukaryotes. The generation of novel taxa arises indirectly from the acquisition of a novel metabolic capability. One may argue, for example, that mycetocyte symbionts enabled early homopteran insects to thrive on plant sap, and this was followed by an adaptive radiation of taxa to exploit different plant species and locations (phloem/xylem, root/shoot, etc.). Similarly, the ability of ascomycete fungi to thrive in terrestrial environments, as a consequence of lichenization, led to the proliferation of taxa adapted to use different substrata (rock types, bark, etc.) and environments.

The fossil record of very few symbioses is adequate to explore directly the relation between acquisition of symbionts and radiation of host taxa. One example is the scleractinian corals, a group of cnidarians that produces an external skeleton of calcium carbonate. Only species with algal symbionts have a sufficiently high calcification rate to generate reefs. The first scleractinians in the mid Triassic did not contribute significantly to reefbuilding, but, about 25 million years later in the Jurassic, they rapidly came to dominate reef habitats and underwent a massive radiation of new families and genera. Stanley (1981) has argued that the early scleractinians in the Triassic lacked algal symbionts, and that both the onset of reefbuilding forms and proliferation of new scleractinian lineages arose directly from the formation of the symbiosis in the late Triassic/early Jurassic. However, recent advances in methods to distinguish alga-bearing and alga-free scleractinian fossils do not support Stanley's hypothesis unequivocally (Coates and Jackson 1987).

4. SYMBIONT-DERIVED ORGANELLES

It is widely believed that certain intracellular symbionts in eukaryotes may evolve into organelles. Symbiotic origins are accepted for mitochondria and plastids, and have also been proposed for peroxisomes and hydrogenosomes (Muller 1988; Cavalier-Smith 1990).

The literature on the acquisition of symbiont-derived organelles and origin of eukaryotes is somewhat confused because Margulis (1970) mistakenly considered that the acquisition of mitochondria played a central role in the evolution of eukaryotes. As discussed in more detail in section 4.2.2, eukaryotes evolved before mitochondria were acquired, and there is a substantial radiation of mitochondrion-free protists. The

eukaryotes are defined by their possession of a double-membrane-enclosed nucleus, and they differ from prokaryotes (archaeabacteria and eubacteria) by a variety of other features, including their endocytotically-active cell membrane, by which intracellular symbionts are acquired (Smith and Douglas 1987).

In this section, the factors influencing the transformation of symbionts into organelles are considered. In section 4.1, the differences between organelles and symbionts are reviewed, to establish the principal 'hurdles' that must be overcome in the evolution of organelles. In section 4.2, current hypotheses on the evolutionary origins of plastids and mitochondria are described, with the aim to examine the number of times organelles may have arisen and the identity of their symbiont precursors.

By the original 'serial endosymbiosis theory' of the origin of eukaryotic organelles, Margulis (1970) proposed that eukaryotic flagella and microtubules are derived from spirochaete symbionts. The mechanisms underlying the putative transformation of spirochaetes into microtubular systems differ fundamentally from the proposed origin of mitochondria and plastids, and therefore the status of this system is considered separately in section 4.3.

4.1 Characteristics of symbiont-derived organelles

The various endosymbionts discussed in section 2 are analogous to the mitochondria and plastids, which provide eukaryotic cells with complex metabolic capabilities, aerobic respiration, and photosynthesis, respectively. However, the organelles differ from microbial symbionts in that they have small genomes (14–2400 kbp for mitochondria and 120–200 kbp for plastids), and that 80–90 per cent of the organelle-specific polypeptides are coded by the nucleus. (Douce and Neuberger 1989; Gray 1989). By the symbiotic theory of the origin of mitochondria and plastids, the condition of the organelles' genomes has arisen by the transfer of many genes to the nucleus and loss of organellar DNA coding for functions that duplicated the nuclear-coded capabilities (Margulis 1970, 1981).

Most discussions about the evolution of eukaryotic organelles are vague about the difference between symbionts and symbiont-derived organelles. I propose that a symbiont can be considered to have evolved into an organelle if genes essential to its function are transferred to the host nucleus. The resultant organelle is irrevocably dependent on the host lineage that acquired its genes. Microbial symbionts with a very small genome size cannot be described as organelles because they are not necessarily dependent on one particular host lineage. The only well-documented microbial symbionts with a low DNA content are the 'cyanelles' in *Cyanophora paradoxa*, with a genome size of 127 kbp (Wassman *et al.* 1987). Gene transfer to the host has not been

demonstrated; the small subunit of ribulose-1, 5-bisphosphate carboxylase (coded in the nucleus of chlorophytes and plants) is coded by the cyanelles (Starnes *et al.* 1985).

Three factors that may influence the probability that symbionts evolve into organelles are considered in this section: the probability that genes are transferred from symbionts to the host nucleus; the structural organization of the host; and the selective transport of the products of nucleus-encoded genes to the symbiont/organelle.

4.1.1 The transfer of genes from symbiont/organelle to host nucleus

The movement of DNA sequences between nucleus, mitochondria, and plastids of eukaryotes is well documented (Timmis & Scott 1983; Gray 1989; Levings and Brown 1989), suggesting that this is not a major barrier to the evolution of organelles. Central to the symbiotic theory is that net transfer of functional genes is asymmetric, with the nucleus as the usual recipient and symbionts/organelles as donors. Several factors may contribute to this asymmetry. The nucleus may 'scavenge' free cytoplasmic DNA arising from transient breakdown of organelle membranes or organelle lysis (Thorsness and Fox 1990). Also, the nuclear genome is much less compact than bacterial and organellar genomes, and therefore incoming DNA is less likely to cause disruption by integrating into a region coding for essential functions in the nucleus than organelles. Further, since most eukaryotic cells have one nucleus but many organelles, a gene acquired by the nucleus is far more likely to be transmitted to daughter cells than one acquired by a single organelle.

The validity of the proposed asymmetry of gene movement has recently been verified experimentally by Thorsness and Fox (1990), who demonstrated that a functional gene is transferred from mitochondria to the nucleus of yeast at a frequency of 2×10^{-5} per cell per generation, and from the nucleus to mitochondria at an undetectable rate (less than 10^{-10} per cell per generation). It is also consistent that no nuclear functions are known to be coded by an organelle or symbiont genome. The proposed transfer of the gene for superoxide dismutase from leiognathid fish to their luminescent symbionts has been discredited (Steffens *et al.* 1983) and, contrary to early reports, the eukaryote-like glutamine synthetase in rhizobia is probably of bacterial origin, and not derived from the legume host (Shatters and Kahn 1989).

4.1.2 Structural organization of the eukaryotic host

One consequence of the transfer of symbiont genes to the host nucleus (and associated loss of those genes from the symbiont) is that the symbiont/organelle is irrevocably dependent on that nucleus and its progeny. The nucleus most likely to acquire symbiont DNA is that of the cell housing the symbionts, and therefore organelles can be expected

to evolve in organisms in which the cell containing the symbionts also gives rise to host progeny. This is the norm in unicellular protists. In most multicellular eukaryotes, the cells containing symbionts are developmentally distinct from the reproductive propagules, and therefore organelles are less likely to evolve in multicellular than unicellular hosts.

The conclusion that symbiont-derived organelles are most likely to evolve in protist hosts is consistent with the facts; none are known to have evolved in multicellular eukaryotes. It also has implications for the functions that organelles may possess. Symbionts with certain metabolic capabilities, notably nitrogen fixation and chemosynthesis, are largely restricted to multicellular hosts, and this may account for the absence of diazotrophic and chemosynthetic organelles.

4.1.3 Targeting of nucleus-encoded polypeptides to the organelle

Nucleus-encoded genes are translated in the eukaryotic cytoplasm, and therefore the products of organelle-specific genes have to be transported specifically to the organelle. Cavalier-Smith (1987a) has argued that the greatest barrier to the evolution of organelles is the development of mechanisms by which polypeptides are targeted faithfully to the appropriate compartment. Consistent with this hypothesis, the selective sequestration of host polypeptides by symbionts has not been demonstrated unambiguously in any symbiosis.

4.2 The evolutionary origins of organelles

4.2.1 Plastids

The plastids in eukaryotes can be assigned to three groups: those with chlorophyll *a* and *b* in the Chlorophyta, Euglenophyta, and plants; those with chlorophyll *a* and phycobilisomes in the Rhodophyta; and those with chlorophyll *a* and *c* in Chromophyta (including Phaeophyceae, Bacillariophyceae, and Chrysophyceae). This has been interpreted as evidence for three prokaryotic ancestors of plastids: the prochlorons with chlorophyll *b*, cyanobacteria with phycobilisomes, and an unknown prokaryote with chlorophyll *c* (Whatley and Whatley 1981). However, various DNA sequence data do not support this simple interpretation, but indicate considerable complexity in the relationships among the various plastids. For example, Kuhsel and Kowallik (1989) conclude that the plastids in the Chromophyta and Chlorophyta are derived from the same prokaryotic symbiont, but the analyses of van den Eynde *et al.* (1988) suggest that the plastids in the Chlorophyta and plants (both containing chlorophyll *b*) may be derived from different prokaryotic lineages; the phycobilisome-bearing 'cyanelle' symbionts, but not the prochloron *Prochlorothrix*, can be allied specifically with chlorophyll-*b*.

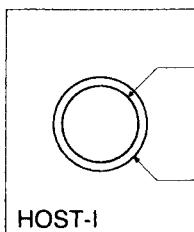
bearing plastids of plants (Giovannoni *et al.* 1988; Turner *et al.* 1989). The molecular data have not yet resolved the number of prokaryotic ancestors of plastids, but they do support the suggestion of Taylor (1983) that photosynthetic pigments are not a reliable character because they may have evolved independently more than once, including in the symbiotic condition.

Further complexity in plastid origins is indicated by some plastids with more than two membranes. These 'complex' plastids have probably arisen from two successive symbioses (Whatley and Whatley 1981; Gibbs 1981). Specifically, it is proposed that a eukaryote (host-I in Fig. 1) containing a prokaryotic photosynthetic symbiont/plastid has been acquired by a second eukaryote (host-II), and that host-I was subsequently reduced. The space between membrane-2 and membrane-3 of complex plastids is known as the periplastid space, and it is the proposed location of host-I cell contents. In most taxa, this space is apparently 'empty,' but in the Cryptophyceae the periplastid space has ribosomes and a 'nucleomorph,' a DNA-containing structure, bounded by two membranes, which may be the vestigial nucleus of host-I (Ludwig and Gibbs 1985). Consistent with this interpretation, the nucleomorph and ribosomes in the periplastid space of the cryptophyte *Chroomonas caudata* contain eukaryotic rRNA (McFadden 1990).

One possible scenario that would account for the proposed retention of host-I nucleus (the nucleomorph) in some but not all complex plastids is that the former host-I contained plastids and the latter contained autonomous prokaryotic symbionts, when acquired by host-II. In the first group (which includes the Cryptophyta), the host-I nucleus may have contained plastid-specific genes and was retained as the nucleomorph. In the latter, the symbionts would have evolved into organelles in the 'double' symbiosis, with the transfer of genes directly to the host-II nucleus, while the host-I nucleus, and all the cytoplasmic components for which it coded, were redundant and therefore eliminated. Consistent with this interpretation, many intracellular algal symbionts of non-photosynthetic protists are known (see section 2.1), but the only ones in which the algal nucleus is eliminated are those with plastids containing a nucleomorph. For example, the dinoflagellate *Gymnodinium acidotum* contains a cryptomonad symbiont, but in 60 per cent of the specimens examined by Farmer and Roberts (1990), the cryptomonad nucleus had been lost.

Turning to the eukaryotic hosts of plastids, both structural and molecular data demonstrate unambiguously that a wide diversity of protists have acquired these organelles. 28S rRNA sequence analyses indicate the nucleocytoplasm of the three major 'algal' groups, Rhodophyta, Chlorophyta, and Chromophyta, are distinct, probably monophyletic assemblages (Perasso *et al.* 1989). The Euglenophyta and Dinophyta are allied to the

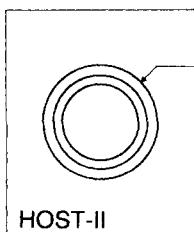
(a) Two membranes (e.g., Chlorophyta, Rhodophyta)



Membrane-1: cell membrane of prokaryote symbiont

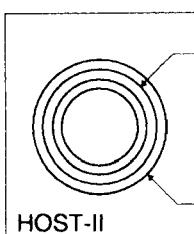
Membrane-2: outer membrane of prokaryote symbiont
or perisymbiont membrane of host-I

(b) Three membranes (e.g., Euglenophyta, many Dinophyta)



Membrane-3: cell membrane of host-I
or perisymbiont membrane of
host-II

(c) Four membranes (e.g., Chromophyta)



Membrane-3: cell membrane of host-I

Membrane-4: perisymbiont membrane of host-II

Fig. 1. Possible origins of membranes enclosing plastids. The bounding membranes of plastids are shown as circles, with the innermost membrane as membrane-1, and the host cell as a square (a). In (b) and (c), the eukaryotic host-I (or its symbionts/plastids) have been acquired by host-II. Two points of uncertainty are: (1) whether membrane-2 and membrane-3 are the perisymbiont membranes of the host (Whatley and Whatley 1981) or cell membranes of the symbiont (Keegstra *et al.* 1984). The latter is more probable; the properties of membrane-2 of organelles are more similar to the outer membrane of Gram-negative bacteria than perisymbiont membranes of intracellular symbioses (Keegstra *et al.* 1984), and among intracellular symbioses perisymbiont membranes are more readily lost, leaving the symbionts in direct contact with the host cytoplasm, e.g. the mycetocyte symbionts of weevils and tsetse flies (Douglas 1989), the chrysophyte algal symbionts of the dinoflagellates *Peridinium balticum*

Kinetoplastida and ciliates, respectively (Hori and Osawa 1987). All these protist hosts of plastids are advanced groups, and Perasso *et al.* (1989) have concluded that the plastids were probably acquired in the late Precambrian, just before the radiation of metazoa. The reasons for the long interval between the evolution of prokaryotes with oxygenic photosynthesis (about 2×10^9 years ago) and the evolution of eukaryotic algae (i.e. protists with plastids) are obscure. Perhaps, prokaryotic lineage(s) predisposed to symbiosis evolve very rarely (see section 2.7), and the transformation of symbionts to organelles is improbable (see section 4.1.3).

4.2.2 Mitochondria

Most early ideas on the symbiotic origin of mitochondria assumed that these organelles are monophyletic, and evolved from an aerobic Gram-negative eubacterium (Margulis 1981). The mitochondria have been allied specifically with the alpha-Proteobacterium *Paracoccus denitrificans*, whose electron transport chain is closely similar to that of mitochondria (John and Whatley 1975). Although the genetic organization of mitochondria is now known to differ dramatically from eubacteria (e.g. mitochondria possess introns and diverge from the universal code) (Gray 1989), this view has been supported by some rRNA sequence studies (Yang *et al.* 1985; Cedergren *et al.* 1988). Other studies indicate that more than one prokaryotic lineage may have given rise to mitochondria. Firstly, as recognized by Manton (1959), the mitochondria of some protists (e.g. the Chromophyta) have tubular cristae and not the more widespread lamellar cristae, but the interpretation that this structural difference reflects separate evolutionary origins of the two types of mitochondria (Stewart and Mattox 1984) has not been investigated further. Secondly, recent rRNA sequence data suggest that plant mitochondria may have different origins from animal and fungal mitochondria. Gray *et al.* (1989) assign the animal/fungal and plant mitochondria to different lineages within the Proteobacteria, accounting for the differences in biochemistry and genetic organization between plant and other mitochondria (Douce

Fig. 1. Continued

and *Glenodinium foliaceum* (Eschbach *et al.* 1990), and bacterial symbionts of the trypanosome *Blastocrithidia* (Soares and de Souza 1988), (2) whether the three-membrane plastids were acquired as plastids (Whatley and Whatley, 1981) or as intact cells (Gibbs 1981) of host-I. The first interpretation is unlikely because plastids require nucleus-encoded gene products and do not persist indefinitely in cells of a heterotroph (Smith and Douglas 1987). The three-membrane plastids are probably derived from four-membrane plastids, by the loss of host-II perisymbiont membrane.

and Neuburger 1989; Levings and Brown 1989). More radically, van den Peer *et al.* (1990) have concluded that only the plant mitochondria belong to the eubacteria; the mitochondria of animals, fungi and protists could not be assigned to eubacteria, archaeabacteria, or eukaryotes. This is difficult to reconcile with the near uniformity in mitochondrial-specific genes coded by the nucleus of all eukaryotes (Gray 1989) but it is consistent with the 'archigenetic' theory of Mikelsaar (1987), by which the mitochondria of many organisms are the sole surviving representatives of the most ancient life forms, from which the three primary kingdoms arose.

Ideas about the identity of the host cell that acquired mitochondria have changed dramatically over the last decade. The serial endosymbiosis theory of Margulis, as originally formulated, explicitly viewed the acquisition of mitochondria as central to the origin of eukaryotes, with the implication that the first host of mitochondria was not a eukaryote – archaeabacteria were considered as likely candidates (Searcy *et al.* 1978). It is now appreciated that up to 1000 species of protists, including the metamonads, parabasalians, microsporidians, and some amebae, primitively lack mitochondria, but few ideas have emerged about the relationship between these protists and the first host(s) of mitochondria. The suggestion that the ameboflagellate *Pelomyxa* is allied to the ancestor of protists with mitochondria (Margulis 1981) is based on the erroneous assumptions that *Pelomyxa* is archaic (see Griffin 1988) and its symbionts respire aerobically (see section 2.3). The proposed significance of diplomonads such as *Giardia* as hosts of mitochondria (Cavalier-Smith 1987b) can be discounted, because sequence analyses indicate a very early divergence of *Giardia* from other eukaryotes (Sogin *et al.* 1989).

Modern phylogenies of protists based on sequence data, especially of 28S rRNA, indicate that eukaryotes acquired mitochondria long before plastids (Beroim *et al.* 1988). Most authorities follow the view of Margulis (1981) that protists with mitochondria evolved after the evolution of oxygenic photosynthesis and rise in atmospheric oxygen, about 2×10^9 years ago (Raven and Sprent 1989). However, Finlay *et al.* (1983) and Mikelsaar (1987) have suggested an even more ancient origin of mitochondria, under anaerobic conditions. They propose that the symbiont progenitor of mitochondria may have used terminal electron acceptors other than oxygen, and oxidative phosphorylation evolved in symbiosis. As an analogue, the mitochondria in a facultatively anaerobic ciliate *Loxodes* can use nitrate as terminal acceptor (Finlay *et al.* 1983).

Pertinent to the origins of mitochondria is the apparent absence of analogous symbionts (see section 2.3). Perhaps such symbioses are intrinsically improbable (although it is not clear why). Alternatively, mitochondrion-like symbionts may have been widespread in early eukaryotes, but are now absent, either because most organisms predisposed

to form the association are extinct (consistent with the archigenetic theory of Mikelsaar) or because no symbiosis could compete with the eukaryotes containing mitochondria.

In summary, the symbiotic origin of mitochondria is generally accepted, but considerable uncertainty remains concerning the number of times mitochondria have evolved and the identity of the host and symbiont taxa involved. Mitochondria differ from plastids in two important respects: the lack of analogues of mitochondria among extant symbioses, and absence of 'complex' mitochondria derived from eukaryotic symbionts. The reasons are obscure, but may be linked to the probable antiquity of mitochondria.

4.3 Eukaryotic flagella and microtubules

All eukaryotes have cytoplasmic microtubules, and the majority have the most complex microtubule-based organelle, the flagellum (=undulipodium). The flagellum has probably evolved once (Dodge 1980) very early in the history of eukaryotes (Beroin *et al.* 1988); most aflagellate eukaryotes (including the rhodophytes and fungi) have lost flagella secondarily, but the microsporidia may be primitively aflagellate (i.e. they had no flagellate ancestors). The fossil record of eukaryotes extends to $0.8\text{--}1.5 \times 10^9$ years ago, but recent sequence data suggest a more ancient origin, with the divergence of microsporidia at 2.8×10^9 years (Vossbrinck *et al.* 1987) and *Giardia* possibly even earlier (Sogin *et al.* 1989). Thus, if microtubules are derived from prokaryotic symbionts, the symbiosis must have been very ancient and (contrary to the original serial endosymbiosis theory of Margulis 1970) preceded the acquisition of mitochondria.

The symbiotic theory of the evolution of flagella and microtubules is based largely on the *analogy* between flagellate eukaryotic cells and certain protists with ectosymbiotic spirochaetes, some of which are attached to their hosts by structures strikingly similar to the basal bodies of eukaryotic flagella (Margulis 1981). It is an analogy because the protist hosts of spirochaetes have flagella.

Spirochaete motility is mediated by axial filaments of a protein known as (bacterial) flagellin. Margulis (1981) has proposed that spirochaetes additionally have intracellular filaments composed of tubulin. She envisages the direct transformation of putative ectosymbiotic spirochaetes of early eukaryotes into flagella, accompanied by a switch from axial filaments to microtubule-mediated motility. The attachment sites between spirochaetes and host evolved into the basal bodies of the flagella, and gave rise to other microtubule organizing centres (MTOCs), from which the cell's other microtubule systems were generated.

Several alternative scenarios for the symbiotic origin of microtubules

and flagella have been proposed. Notably, Szthmary (1987) has suggested that, after the generation of attachment sites/basal bodies, the spirochaete symbionts were phagocytosed and maintained as intracellular symbionts; their tubulin genes were transferred to the eukaryotic nucleus and used to generate flagella and microtubules independently. The primitively aflagellate microsporidia are consistent with this hypothesis but not with the earlier scenario of Margulis, in which the flagellum is viewed as the progenitor of microtubules.

A long-standing approach to test the validity of the symbiotic theory has been to search for tubulin in spirochaetes. Preliminary positive evidence is published (Margulis *et al.* 1978), but the putative tubulin-like proteins and their genes have yet to be characterized fully. Interest in the symbiotic theory has been renewed recently by the demonstration of substantial DNA (6900 kbp) coding for flagellum-specific functions in the basal bodies of the green alga *Chlamydomonas* (Hall *et al.* 1989). However, the basal body DNA is transferred to the nucleus at meiosis, indicating that it may be of nuclear origin. Further data are required to establish whether the basal bodies of flagella have an independent translation machinery and, if appropriate sequences (rDNA?) are present, to investigate the phylogenetic relationship between this DNA, spirochaetes, and eukaryotes.

5. SYMBIOSIS AND HORIZONTAL GENE TRANSFER

One approach to assess the significance of symbiosis in the evolutionary history of eukaryotes is to consider the alternative means by which eukaryotes may have acquired complex metabolic capabilities. Syvanen (1985) has argued that eukaryotes may obtain novel properties from phylogenetically distant organisms via isolated genes, i.e. by horizontal gene transfer. This is widespread among prokaryotes, especially for functions, such as multiple antibiotic resistance, coded on plasmids (Trieu-Cuot *et al.* 1987; Eberhard 1990). There are also well-documented examples of transfer from bacteria to eukaryotes, notably of plasmid-coded T-DNA from *Agrobacterium* to dicot plants (Zambryski *et al.* 1989) and of genes on conjugative plasmids of *E. coli* to yeast (Heinemann and Sprague 1989), but these are *rare* instances. Horizontal gene transfer is believed to be of little general significance in the evolution of eukaryotes (Kreiber and Rose 1986; Stachel and Zambryski 1989), and trivial as a source of novel metabolic capabilities. For example, many eukaryotes have acquired nitrogen fixation by symbiosis, but none by horizontal gene transfer, even though the genes are frequently plasmid-borne and have probably been transmitted laterally among prokaryotes (Ruvken and Ausubel 1980; Postgate and Eady 1988).

However, symbiosis and horizontal gene transfer are not exclusively alternatives, because symbiosis has led to the transfer of many genes from prokaryotes to protists in the evolution of organelles (section 4). The promotion of gene transfer across the prokaryote-eukaryote border can be viewed as a further evolutionary consequence of symbiosis.

6. CONCLUSIONS

The central proposition of this article is that symbiosis with microorganisms is a key means by which eukaryotes gain genetically and biochemically complex metabolic capabilities. In particular, symbiosis is the sole route by which eukaryotes have acquired photosynthesis, chemosynthesis, and nitrogen fixation.

Photosynthetic symbionts (both algae and cyanobacteria) are widely distributed among protists, ascomycete fungi (as lichens), and structurally simple invertebrates, but chemosynthesis and nitrogen fixation are largely restricted to multicellular eukaryotes because their inefficient carbon output and high energy requirements, respectively, would be very costly for small organisms with a high specific metabolic rate. Some heterotrophic protists and animals derive essential nutrients (e.g. vitamins, essential amino acids) from bacterial symbionts, but, apart from the mycetocyte symbionts in insects, these associations are poorly studied and their prevalence is uncertain. The capacity to generate light has evolved intrinsically many times in eukaryotes, but a minority of marine teleost fish and cephalopod mollusks have luminescent bacterial symbionts, although rarely in conjunction with intrinsic luminescence. Vertebrates lack cellulases, and virtually all herbivorous vertebrates use cellulolytic members of their gut microflora to degrade plant material. The lower incidence of cellulolytic symbionts in herbivorous insects can be linked, in part, to the possession of intrinsic cellulases in these invertebrates.

In general, relatively few microorganisms with a given metabolic capability enter into symbiosis, but they are acquired by a broad taxonomic range of eukaryotic hosts. It is argued that, given the selective advantage for the host to acquire a certain metabolic capability, the principal constraint on the evolution of an association is the availability of suitable microorganisms capable of forming a symbiosis.

The acquisition of symbionts has represented strong selection pressure for substantive morphological and biochemical changes in the eukaryotic hosts. Examples include the changes in gut anatomy in herbivorous vertebrates to accommodate cellulolytic microorganisms, and the transformation of the gut of pogonophorans to the trophosome in which the chemosynthetic symbionts are located. The acquisition of novel metabolic capabilities has enabled eukaryotes to exploit novel niches, e.g. plant sap

by homopteran insects, the terrestrial environment by lichenized fungi. The expansion of the metabolic repertoire of eukaryotes by symbiosis may be the foundation of many radiations, including lichens, herbivores, and reefbuilding corals.

Symbiont-derived organelles can be distinguished from their microbial precursors by the transfer of genes for symbiont/organelle-specific functions to the host nucleus. The widespread movement of DNA between organelles in eukaryotes indicates that this is not a major barrier to the evolution of organelles; the greatest barrier is probably the requirement for specific targeting of polypeptides, coded by genes transferred to the nucleus, to the nascent organelle. Symbiont-derived organelles are most likely to evolve in protists because the most probable recipient of symbiont-derived genes is the cell housing the symbionts, but this cell would not give rise to the next host generation in most multicellular hosts.

Plastids have been acquired multiply by a wide variety of protists, but the phylogenetic range of their prokaryotic precursors is not certain. The variation in photosynthetic pigments of plastids suggests that three distinct cyanobacterial lineages may be involved, but some molecular data indicate that plastids have a more complex ancestry. The evolutionary origins of mitochondria are uncertain. They may have evolved from eubacteria (specifically alpha-Proteobacteria), but, by some rRNA sequence analyses, the mitochondria of animals, fungi, and ciliates can be assigned to no extant kingdom of organisms. The first hosts of mitochondria were eukaryotes, but their relation to modern anaerobic protists is not known.

Interest in the proposed symbiotic origin of eukaryotic flagella and microtubules from spirochaetes has recently been heightened by the demonstration of DNA in the basal body of *Chlamydomonas*, but further study is required.

ACKNOWLEDGMENTS

I thank Dr B. Okamura, Sir David Smith and especially Dr J. B. Searle for their helpful comments on the manuscript. This article was written while in receipt of a University Research Fellowship of the Royal Society of London.

REFERENCES

- Appleby, C. A. (1984). Leghaemoglobin and *Rhizobium* respiration. *Annu. Rev. Plant Physiol.* **35**, 443–78.
Baker, D. and Miller, N. G. (1980). Ultrastructural evidence for the existence of actinorrhizal symbioses in the late Pleistocene. *Can. J. Bot.* **58**, 1612–20.

- Bauchop, T. (1977). Foregut fermentation. In *Microbial ecology of the gut*. (ed. R. T. H. Clarke and T. Bauchop), pp. 223–51. Academic Press, New York.
- Bauchop, T. (1989). Biology of gut anaerobic fungi. *BioSystems* **23**, 53–64.
- Beauchamp, B., Krouse, H. R., Harrison, J. C., Nassichu, K. W. W. and Elink, L. S. (1989). Cretaceous cold-seep communities and methane derived carbonates in the Canadian Arctic. *Science*, **244**, 53–6.
- Bermudes, D. and Margulis, L. (1987). Symbiont acquisition as neosome: origin of species and higher taxa. *Symbiosis* **4**, 379–97.
- Beroir, A., Perasso, R., Qu, L-H., Brugerolle, G., Bachellerie, J. P. and Adoutte, A. (1988). Partial phylogeny of unicellular eukaryotes based on rapid sequencing of a portion of 28S ribosomal RNA. *Proc. Natl. Acad. Sci. USA* **85**, 3474–8.
- Breznak, J. A. (1984). Biochemical aspects of symbiosis between termites and their intestinal microbiota. In *Invertebrate-microbial interactions* (ed. J. M. Anderson, A. D. M. Rayner and D. W. H. Walton), pp. 173–204. Cambridge University Press, Cambridge.
- Buchner, P. (1965). *Endosymbioses of animals with plant microorganisms* Wiley, New York.
- Campbell, A. K. (1989). Living light: biochemistry, function and biomedical applications. *Essays Biochem.* **24**, 41–81.
- Cary, S. C., Fisher C. R. and Felbeck, H. (1988). Mussel growth supported by methane as sole carbon and energy source. *Science*, **240**, 78–80.
- Cavalier-Smith, T. (1987a). The origin of eukaryote and archaeabacterial cells. *Ann. N. Y. Acad. Sci.* **503**, 17–54.
- (1987b). Eukaryotes with no mitochondria. *Nature* **326**, 332–3.
- (1990). Symbiotic origin of peroxisomes. In *Endocytobiology IV* (ed. P. Nardon, V. Gianinazzi-Pearson, A. M. Grenier, L. Margulis and D. C. Smith), pp. 515–21. INRA, Paris.
- Cavanaugh, C. M., Gardiner, S. L., Jones, M. L., Jannasch, H. W. and Waterbury, J. B. (1981). Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* Jones: possible chemoaautotrophic symbionts. *Science* **213**, 340–2.
- Cavanaugh, C. M., Levering, P. R., Maki, J. S., Mitchell, R. and Lidstrom, M. E. (1987). Symbiosis of methylotrophic bacteria and deep sea mussels. *Science* **325**, 346–8.
- Cedergren, R., Gray, M. W., Abel, Y. and Sankoff, D. (1988). The evolutionary relationships among known life forms. *J. Mol. Evol.* **28**, 98–112.
- Coates, A. G. and Jackson, J. B. C. (1987). Colonial growth, algal symbiosis and reef formation by corals. *Paleobiology* **13**, 363–78.
- Coleman, G. S. (1980). Rumen ciliate protozoa. *Adv. Parasitol.* **18**, 121–73.
- Cowen, R. (1970). Analogies between the recent bivalve *Tridacna* and the fossil brachiopods Lyttonicaea and Richthofenicea. *Palaeogeography, Palaeoclimatol. Palaeoecol.* **8**, 329–44.
- (1983). Algal symbiosis and its recognition in the fossil record. In *Biotic interactions in recent and fossil benthic communities* (ed. M. J. S. Tevesz and P. L. McCall), pp. 431–78. Plenum Press, New York.
- Distel, D. L. (1990). Detection, identification and phylogenetic analysis of endosymbiotic bacteria using ribosomal RNA sequences In *Endocytobiology*

- IV* (ed. P. Nardon, V. Gianinazzi-Pearson, A. M. Grenier, L. Margulis and D. C. Smith), pp. 339–42. INRA, Paris.
- , Lane, D. J., Olsen, G. J., Giovannoni, S. J., Pace, B., Pace, N. R., Stahl, D. A. and Felbeck, H. (1988). Sulfur-oxidising bacterial endosymbionts: analysis of phylogeny and specificity by 16SrRNA sequences. *J. Bacteriol.* **170**, 2506–10.
- Dodge, J. D. (1980). Morphology and phylogeny of flagellated protists. In *Endocytobiology I* (ed. W. Schwemmler and H. E. A. Schenk), pp. 33–50. Walter de Gruyter, Berlin.
- Douce, R. and Neuburger, M. (1989). The uniqueness of plant mitochondria. *Annu. Rev. Plant Physiol. Mol. Biol.* **40**, 371–414.
- Douglas, A. E. (1988a). Algal-invertebrate symbiosis. In *The Biochemistry of algae and cyanobacteria* (ed. J. Gallon and J. G. Rogers), pp. 297–309. Oxford University Press, Oxford.
- (1988b). Sulphate utilisation in an aphid symbiosis. *Insect Biochem.* **18**, 599–605.
- (1989). Mycetocyte symbiosis in insects. *Biol. Rev. Camb. Philos. Soc.* **64**, 409–434.
- and Smith D. C. (1989). Are endosymbioses mutualistic? *Trends Ecol. Evol.* **4**, 350–2.
- Eberhard, W. G. (1990). Evolution in bacterial plasmids and levels of selection. *Q. Rev. Biol.* **65**, 3–22.
- Edwards, D. S. (1986). *Aglaophyton major*, a non-vascular land plant from the Devonian Rhynie chert. *Bot. J. Linn. Soc.* **93**, 173–204.
- Eschbach, S., Speth, V., Hausmann, P. and Sitte, P. (1990). Freeze-fracture study of the single membrane between host cell and endosymbiont in the dinoflagellates *Glenodinium foliaceum* and *Peridinium balticum*. *J. Phycol.* **26**, 324–8.
- Farlow, J. O. (1987). Speculations about the diet and digestive physiology of herbivorous dinosaurs. *Paleobiology* **13**, 60–72.
- Farmer, M. A. and Roberts, K. R. (1990). Organelle loss in the endosymbiont of *Gymnodinium acidotum* (Dinophyceae). *Protoplasma* **153**, 178–85.
- Farrar, J. F. (1976). The lichen as an ecosystem: observation and experiment. In *Progress and problems in lichenology* (ed. D. H. Brown, D. Hawksworth and R. H. Bailey), pp. 385–406. Academic Press, New York.
- Fenchel, T. and Finlay, B. J. (1989). *Kentrophoros*: a mouthless ciliate with a symbiotic kitchen garden. *Ophelia* **30**, 75–93.
- Finlay, B. J., Span, A. S. W. and Harman, J. M. P. (1983). Nitrate respiration in primitive eukaryotes. *Nature* **303**, 333–5.
- Fisher, C. R. and Childress, J. J. (1986). Translocation of fixed carbon from symbiotic bacteria to host tissues in the gutless bivalve *Solemya reidi*. *Mar. Biol.* **93**, 59–68.
- Frost, T. M. and Williamson, C. E. (1980). *In situ* determination of the effect of symbiotic algae on the growth of the freshwater sponges *Spongilla lacustris*. *Ecology* **6**, 1361–70.
- Gibbs, S. P. (1981). The chloroplasts of some algal groups may have evolved from eukaryotic algae. *Ann. N. Y. Acad. Sci.* **361**, 193–208.
- Giere, O., Wirsén, C. O., Schmidt, C. and Jannasch, H. W. (1988). Contrasting

- effects of sulphide and thiosulphate on symbiotic CO₂ assimilation of *Phallodrilus leukodermatus* (Annelida). *Mar. Biol.* **97**, 413–9.
- Giovannoni, S. J., Turner, S., Olsen, G. J., Barns, S., Lane, D. L. and Pace, N. R. (1988). Evolutionary relationships among cyanobacteria and green chloroplasts. *J. Bacteriol.* **170**, 3584–8.
- Gray, M. W. (1989). Origin and evolution of mitochondrial DNA. *Annu. Rev. Cell Biol.* **5**, 25–50.
- , Cedergren, R., Abel, Y. and Sankoff, D. (1989). On the evolutionary origin of the plant mitochondrion and its genome. *Proc. Natl. Acad. Sci. USA* **86**, 2267–71.
- Griffin, J. L. (1988). Fine structure and taxonomic position of the giant amoeboid flagellate *Pelomyxa palustris*. *J. Protozool.* **35**, 300–15.
- Hall, J. L., Ramanis, Z. and Luck, D. J. L. (1989). Basal body/centriolar DNA: molecular genetic studies in *Chlamydomonas*. *Cell* **59**, 121–32.
- Hamyon, R. M., Koski, R. A. and Sinclair, C. (1984). Fossils of hydrothermal vent worms from Cretaceous sulfide ores of the Samail ophiolite, Oman. *Science* **233**, 1047–9.
- Hastings, J. W. (1983). Biological diversity, chemical mechanisms and the evolutionary origins of bioluminescent systems. *J. Mol. Evol.* **19**, 309–21.
- Hawksworth, D. L. (1988a). The variety of fungal-algal symbioses, their evolutionary significance and the nature of lichens. *Bot. J. Linn. Soc.* **96**, 3–20.
- (1988b). Coevolution of fungi with algae and cyanobacteria in lichen symbiosis. In *Coevolution of fungi with plants and animals* (ed. K. A. Pirozynski and D. L. Hawksworth), pp. 125–48. Academic Press, New York.
- Heinemann, J. A. and Sprague, G. F. (1989). Bacterial conjugative plasmids mobilise DNA transfer between bacteria and yeast. *Nature* **340**, 205–9.
- Herring, P. J. (1985). Physiological adaptations of marine animals. *Symp. Soc. Exp. Biol.* **39**, 323–50.
- (1988). Luminescent organs. In *The mollusca* Vol. II (ed. E. R. Trueman and M. R. Clarke), Academic Press, New York.
- and Morin, J. G. (1978). Bioluminescence in fishes. In *Bioluminescence in action* (ed. P. J. Herring), pp. 273–330. Academic Press, New York.
- Hogan, M. E., Schulz, M. W., Slaytor, M., Czolij, R. T. and O'Brien, R. W. (1988a). Components of termite and protozoal cellulases from the lower termite *Coptotermes lacteus* Froggatt. *Insect Biochem.* **18**, 45–51.
- Hogan, M. E., Veivers, P. C., Slaytor, M. and Czolij, R. T. (1988b). The site of cellulose breakdown in higher termites (*Nasutitermes walkeri* and *Nasutitermes exitiosus*). *J. Insect Physiol.* **34**, 891–9.
- Hori, H. and Osawa, S. (1987). Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. *Mol. Biol. Evol.* **4**, 445–72.
- Janis, C. (1976). The evolutionary strategy of the Equidae and the origins of rumen and cecal digestion. *Evolution* **30**, 757–74.
- John, P. and Whatley, F. R. (1975). *Paracoccus denitrificans* and the evolutionary origin of the mitochondrion. *Nature* **254**, 495–8.
- Keegstra, K., Werner-Washburne, M., Clue, K. and Andrew, J. (1984). The chloroplast envelope: is it homologous with the double membranes of mitochondria and Gram-negative bacteria? *J. Cell Biochem.* **24**, 55–68.

- Kreiber, M. and Rose, M. R. (1986). Molecular aspects of the species barrier. *Annu. Rev. Ecol. Syst.* **17**, 465–85.
- Kuhsel, M. and Kowallik, K. W. (1989). The plastome of a brown algae, *Dictyota dichotoma*. II. Genes coding for polypeptides of the photosynthetic electron transport chain. *Mol. Gen. Genet.*
- Langer, P. (1984). Comparative anatomy of the stomach in mammalian herbivores. *Q. J. Expl Physiol.* **69**, 615–25.
- Lee, J. J., Soldo, A. T., Reisser, W., Lee, M. J., Jeon, K. W. and Gortz, H. D. (1985). The extent of algal and bacterial endosymbioses in Protozoa. *J. Protozool.* **32**, 391–403.
- Levings, C. S. and Brown, C. G. (1989) Molecular biology of plant mitochondria. *Cell* **56**, 171–9.
- Ludwig, M. and Gibbs, J. P. (1985). DNA is present in the nucleomorph of *Cryptomonas* demonstrated by DAPI fluorescence. *Z. Naturforsch. Sect. C* **40**, 933–5.
- Manton, I. (1959). Electron microscopical observations of a very small flagellate: the problem of *Chromulina pusilla* Butcher. *J. Mar. Biol. Assoc. UK.* **38**, 319–33.
- Margulis, L. (1970). *Origin of eukaryotic cells*. Yale University Press, New Haven.
- Margulis, L. (1981). *Symbiosis in cell evolution*. Freeman, San Francisco.
- Margulis, L., To, L. P. and Chase, D. (1978). Microtubules in prokaryotes. *Science* **200**, 1118–24.
- Mattson, W. J. (1980). Herbivory in relation to plant nitrogen content. *Annu. Rev. Ecol. Syst.* **11**, 119–61.
- McBee, R. H. (1977). Fermentation in the hindgut. In *Microbial ecology of the gut* (ed. R. T. H. Clarke and T. Bauchop), pp. 185–222. Academic Press, New York.
- McBee, R. H. (1989). Hindgut fermentations in nonavian species. *J. Exp. Zool. (Suppl.)* **3**, 55–60.
- McFadden, G. I. (1990). Evidence that cryptomonad chloroplasts evolved photosynthetic eukaryotic endosymbionts. *J. Cell Sci.* **95**, 303–8.
- McFall-Ngai, M. J. (1983). Adaptations for reflection of bioluminescent light in the gas bladder of *Leiognathus equalus* (Perciformes: Leiognathidae). *J. Exp. Zool.* **227**, 23–33.
- Mead, G. C. (1989). Microbes of the avian caecum: types present and substrates utilised. *J. Exp. Zool. (Suppl.)* **3**, 48–54.
- Mikelsaar, R. (1987). A view of early cellular evolution. *J. Mol. Evol.* **25**, 168–83.
- Minchin, F. R. and Pate, J. S. (1973). The carbon balance of a legume and the functional economy of its root nodules. *J. Exp. Bot.* **24**, 259–71.
- Morris, J. G. and Rogers, Q. R. (1982). Nutritionally-related adaptations of carnivores and ruminants. In *Adaptations to terrestrial environments* (ed. N. S. Margaris, K. Arianoutsou-Faraggitaki and R. J. Reiter), pp. 165–80. Plenum Press, New York.
- Muller, H. (1988). Energy metabolism of protozoa without mitochondria. *Annu. Rev. Microbiol.* **42**, 465–88.
- Newcomb, W. and Wood, S. M. (1987). Morphogenesis and fine structure of *Frankia* (Actinomycetales): the microsymbiont of nitrogen-fixing actinorhizal root nodules. *Internat. Rev. Cytol.* **109**, 1–58.

- Normand, P. and Bousquet, J. (1989). Phylogeny of nitrogenase sequences in *Frankia* and other nitrogen fixing microorganisms. *J. Mol. Evol.* **29**, 436–47.
- O'Brien, J. and Vetter, R. D. (1990). Production of thiosulphate during sulphide oxidation by mitochondria of the symbiont-containing bivalve *Solemya reidi*. *J. Exp. Biol.* **149**, 133–48.
- Ott, S. (1987). Reproductive strategies in lichens. *Bibliotheca Lichenol.* **25**, 81–93.
- Perasso, R., Baroni, A., Qu, L. H., Bachellerie, J. P. and Adoutte, A. (1989). Origin of the algae. *Nature* **339**, 142–4.
- Pirozynski, K. A. and Malloch, D. W. (1975). The origin of land plants: a matter of mycotrophism. *BioSystems* **6**, 153–64.
- Postgate, J. R. and Eady, R. B. (1988). The evolution of nitrogen fixation: In *Nitrogen fixation: 100 years after* (ed. H. Bothe, F. J. de Brujn and W. E. Newton), pp. 31–40. Gustav Fischer, Stuttgart.
- Powell, M. A. and Somero, G. N. (1986). Adaptations to sulphide by hydrothermal vent animals: sites and mechanisms of detoxification and metabolism. *Biol. Bull.* **171**, 274–90.
- Raven, J. A. and Sprent, J. I. (1989). Phototrophy, diazotrophy and palaeoatmospheres: biological catalysis and the H,C,N and O cycles. *J. Geol. Soc.* **146**, 161–70.
- Rudman, W. D. (1981). The anatomy and biology of alcyonarian-feeding aeolid opisthobranch molluses and their development of symbiosis with zooxanthellae. *Zool. J. Linn. Soc.* **72**, 219–62.
- Ruvken, G. B. and Ausubel, F. M. (1980). Interspecies homology of nitrogenase genes. *Proc. Natl. Acad. Sci. USA* **77**, 191–5.
- Schiemer, F., Novak, R. and Ott, J. (1990). Metabolic studies on thiobiotic free-living nematodes and their symbiotic microorganisms. *Mar. Biol.* **106**, 129–37.
- Schmidt-Nielsen, K. (1984). *Scaling: why is animal size so important?* Cambridge University Press, Cambridge.
- Searcy, D. G., Stein, D. B. and Green, G. R. (1978). Phylogenetic affinities between eukaryotic cells and a thermophilic mycoplasma. *BioSystems* **10**, 19–28.
- Selinger, H. H. (1987). The evolution of bioluminescence in bacteria. *Photochem. Photobiol.* **45**, 291–7.
- Shatters, R. G. and Kahn, M. L. (1989). Glutamine synthetase II in Rhizobium: reexamination of the proposed horizontal transfer of DNA from eukaryotes to prokaryotes. *J. Mol. Evol.* **29**, 422–8.
- Smith, D. C. and Douglas, A. E. (1987). *The biology of symbiosis*. Edward Arnold, London.
- Soares, M. J. and de Souza, W. (1988). Freeze-fracture study of the endosymbiont of *Blastocerithidium culiculus*. *J. Protozool.* **35**, 370–4.
- Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A. and Peatti, D. A. (1989). Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science* **243**, 75–7.
- Southward, E. C. (1987). Contribution of symbiotic chemoautotrophs to the nutrition of benthic invertebrates. In *Microbes in the sea* (ed. M. L. Sleigh), pp. 83–118. Ellis Horwood, Chichester UK.
- Spiro, B., Greenwood, P. B., Southward, A. J. and Dando, P. R. (1986). $^{13}\text{C}/^{12}\text{C}$ ratios in marine invertebrates from reducing sediments: confirmation of

- nutritional importance of chemoautotrophic endosymbiotic bacteria. *Mar. Ecol. (Prog. Ser.)* **28**, 233–40.
- Sprent, J. I. and Sprent, P. (1990). *Nitrogen-fixing organisms*. Chapman Hall, London.
- Stachel, S. E. and Zambryski, P. C. (1989). Generic trans-kingdom sex? *Nature* **340**, 190–1.
- Stanley, G. D. (1981). Early history of scleractinian corals and its geological consequences. *Geology* **9**, 507–11.
- Starnes, S. M., Lambert, D. H., Maxwell, E. S., Stevens, S. E., Porter, R. D. and Stively, J. M. (1985). Cotranscription of the large and small genes of ribulose-1,5-bisphosphate carboxylase/oxygenase in *Cyanophora paradoxa*. *FEMS Microbiol. Lett.* **28**, 165–9.
- Steffens, G. J., Bannister, J. V., Bannister, W. H., Folohe, L., Gunzler, W. A., Kim, S-M.A. and Ottung, F. (1983). The primary structure of Cu-Zn superoxide dismutase from *Photobacterium leiognathi*: evidence for a separate evolution of Cu-Zn superoxide dismutase in bacteria. *Hoppe-Seyler's Z. Physiol. Chem.* **364**, 675–90.
- Stewart, K. D. and Mattox, K. R. (1984). The case for a polyphyletic origin of mitochondria: morphological and molecular comparisons. *J. Mol. Evol.* **21**, 54–57.
- Stubblefield, S. P., Taylor, T. N. and Trappe, J. M. (1987). Vesicular-arbuscular mycorrhizae from the Triassic of Antarctica. *Am. J. Bot.* **74**, 1904–11.
- Syvanen, M. (1985). Cross-species gene transfer; implications for a new theory of evolution. *J. Theor. Biol.* **112**, 333–43.
- Szthmary, E. (1987). Early evolution of microtubules and undulipodia. *BioSystems* **20**, 115–31.
- Taylor, F. J. R. (1983). Some eco-evolutionary aspects of intracellular symbioses. *Internat. Rev. Cytol. (Suppl.)* **14**, 1–28.
- Taylor, T. N. and White, J. F. (1989). Fossil fungi (Endogonaceae) from the Triassic of Antarctica. *Am. J. Bot.* **76**, 389–96.
- Thorsness, P. E. and Fox, T. D. (1990). Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. *Nature* **346**, 376–9.
- Timmis, J. N. and Scott, N. S. (1983). Sequence homology between spinach nuclear and chloroplast genes. *Nature* **305**, 65–7.
- Tjepkema, J. D. (1985). Utilisation of photosynthate for nitrogen fixation in seedlings of *Myrica gale* and *Alnus rubra*. In *Nitrogen fixation and carbon dioxide metabolism* (ed. P. W. Ludden and J. E. Burris), pp. 183–92. Elsevier, New York.
- _____, Schwintzer, C. R. and Benson, D. R. (1986). Physiology of actinorhizal nodules. *Annu. Rev. Plant Physiol.* **37**, 209–32.
- Trieu-Cuot, P., Arthur, M. and Courvalin, P. (1987). Origin, evolution and dissemination of antibiotic resistance genes. *Microbiol. Sci.* **4**, 263–6.
- Trinick, M. J. (1973). Symbiosis between *Rhizobium* and the non-legume *Trema aspera*. *Nature* **244**, 459–60.
- Turner, S., Burger-Wiersma, T., Giovannoni, S. J., Mur, L. R. and Pace, N. R. (1989). The relationship of a prochlorophyte *Prochlorothrix hollandica* to green chloroplasts. *Nature* **337**, 380–2.
- Unterman, B. M., Baumann, P. and McLean, D. L. (1989). Pea aphid symbiont

- relationships established by analysis of 16S rRNA ribosomal RNAs. *J. Bacteriol.* **171**, 2970–4.
- Van Bruggen, J. J. A., Zwart, K. B., van Assema, R. M., Stumm, C. K. and Vogels, C. D. (1984). *Methanobacterium formicum*, an endosymbiont of the anaerobic ciliate *Metopus stratus* McMurrich. *Arch. Microbiol.* **139**, 1–7.
- Van Bruggen, J. J. A., Stumm, C. K., Zwart, K. B. and Vogels, G. D. (1985). Endosymbiotic methanogenic bacteria of the sapropelic amoeba *Mastigella*. *FEMS Microbiol. Ecol.* **31**, 187–192.
- Van Bruggen, J. J. A., van Rens, G. L. M., Geertman, E. J. M., Zwart, K. B., Stumm, C. K. and Vogels, G. D. (1988). Isolation of a methanogenic endosymbiont of the sapropelic amoeba *Pelomyxa palustris* Greeff. *J. Protozool.* **35**, 20–3.
- Van den Eynde, H., de Baere, R., de Roeck, E., van de Peer, Y., Vandenberghe, A., Willekens, P. and de Wachter, R. (1988). The 5S ribosomal RNA sequences of a red algal rhodoplast and a gymnosperm chloroplast. Implications for the evolution of plastids and cyanobacteria. *J. Mol. Evol.* **27**, 126–32.
- Van den Peer, Y., Neefs, J. M. and Wachter, R. de (1990). Small ribosomal subunit RNA sequences, evolutionary relationships among different life forms and mitochondrial origins. *J. Mol. Evol.* **30**, 463–76.
- Vogel, G. K. (1975). Endosymbiotic algae in rudists? *Palaeogeography, Palaeoclimatol., Palaeoecol.* **17**, 327–32.
- Vossbrinck, C. R., Maddox, J. V., Friedman, S., Debrunner-Vossbrinck, B. A. and Woese, C. R. (1987). Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature* **326**, 411–4.
- Wassman, C. C., Loffelhardt, W. and Bohnert, H. J. (1987). Cyanelles: organisation and molecular biology. In *The cyanobacteria – A comprehensive review* (ed. P. Fay and C. van Baleen), pp. 303–24. Elsevier, Amsterdam.
- Waterbury, J. B., Calloway, C. B. and Turner, R. D. (1983). A cellulolytic nitrogen-fixing bacterium cultured from the gland of Deshayes in shipworms (Bivalvia: Teredinidae). *Science* **221**, 1401–3.
- Wellington, G. M. (1982). An experimental analysis of the effect of light and zooplankton on coral zonation. *Oecologia* **52**, 311–20.
- Whatley, J. M. and Whatley, F. R. (1981). Chloroplast evolution. *New Phytol.* **87**, 233–47.
- Wilmot, D. B. and Vetter R. D. (1990). The bacterial symbiont from the hydrothermal vent tubeworm *Riftia pachyptila* is a sulphide specialist. *Mar. Biol.* **106**, 273–83.
- Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* **51**, 221–71.
- Wright, V. P. (1985). The precursor environment for vascular plant colonisation. *Philos. Trans. R. Soc. B.* **309B**, 143–5.
- Yamin, M. A. and Trager, W. (1979). Cellulolytic activity of an axenically-cultivated termite flagellate, *Trichomitopsis termopsidis*. *J. Gen. Microbiol.* **113**, 417–20.
- Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J. and Woese, C. R. (1985). Mitochondrial origins. *Proc. Natl Acad. Sci. USA* **82**, 4443–7.
- Young, J. P. W. and Johnston, A. W. B. (1989). The evolution of specificity in the legume-rhizobium symbiosis. *Trends Ecol. Evol.* **4**, 341–9.

- Zambryski, P., Tempe, J. and Schell, J. (1989). Transfer and function of T-DNA genes from *Agrobacterium* Ti and Ri plasmids in plants. *Cell* **56**, 193-201.
- Zimmerman, L. C. and Tracey, C. R. (1989). Interactions between the environment and ectothermy and herbivory in reptiles. *Physiol. Zoo.* **62**, 374-409.

Index

- aborted species 221
acetic acid 359
achondroplasia 274, 276
adaptation 226–7, 292
and Shifting Balance Theory 57–9
and speciation 242
adaptationist programme 292, 329, 335
adaptive evolution
epistasis in 50–9
gene interactions in 45–50
adaptive radiation 76, 230–3
examples of 232
additive effect on fitness 55–6
additive genetic variance and population subdivision 51–4
adipose tissue 263
Aglaophyton 362–3
Agrobacterium 372
algae 351
allelic diversity 209, 210
allopatric populations, polytypic species of 233
allopatric speciation 16, 219–57
dispersal effects 223
environmental opportunities for 240
multiplicity of effects of single biological phenomenon 244
three-stage 219, 243–5
allopatry 69, 70, 73
allopolyploid inheritance 200
selection toward 200–1
allopolyploidization 200, 210
allopolyploids
backcrossed to parent 189
models for 189
segregation of 186
allotetraploids, quantitative genetics 203
Alnus 353, 354
amebae 370
amphibian–reptile transition 319
Anabaena 354, 361
Animal Species and Evolution 14
animal taxonomy 6
Anolis 85, 86
Anolis bimaculatus leachi 80
ANOVA 42–4, 47, 48, 50
Antarcticycas 363
apoC-II deficiency 126, 127, 129
apoC-II gene 126
apoC-II protein 129
apomixis 9
archigenetic theory 370, 371
Artiodactyls 239
Aspergillus 362
asynchronous flight muscle 302–3
ATP 352, 354–6
hydrolysis 298
autopolyploids
meiosis, 186
segregation of 186
autopolyploidy 9
quantitative genetics 185
theoretical population genetics 185–217
autosomal dominant 274
autotetraploids
quantitative genetics 203
segregation of 187
autotetraploidy, early models 189
Avicularia 362
Azolla 354
Azospirillum 354
Azotobacter 354
bacterial symbionts 358
beanbag genetics 20–2
behavior–performance–structure complex 338
Bergmann's rule 263
biological races 12
biological species concept 8–10
biomechanical analysis 321
biomechanical models 314, 333
biomechanical system 321
biomechanics, domain of 299
birds, species richness of 237
Blastocerithidium 369
blood group distribution 26
body mass, and life history 268–73
body size
ecological consequences of 261
effects 236, 259–63, 272, 278
evolution 98
optimum 262
body weights 272
Bornean tree frogs, ecomorphology of 311–14
bosne formation 316
brachiopod hinges, evolution of 331–3

- Bradyrhizobium* 353
 branching process method 127
 bush elephant 264
 butterfly diversity 238
 butyric acid 359
- Caeciliaidae 324
California Channel Islands 260, 271, 280
Callosobruchus chinensis 82–4
Callosobruchus maculatus 82–3
 Caribbean hogfish 310
 carnivores, number of species of 238
 cartilage transformation 316
 cDNA 112
 cell lineages 316
 cellulolytic symbionts 358–60
Cepaea 26, 221
 cephalopods 361
 character congruence 317
 character convergence 65, 72
 apparent 72
 models for evolution of 72–3
 character displacement 63, 65, 66
 coevolutionary stable community (or CSC) approach 85, 89–90, 93–8
 detection of 69–71
 ecological mechanisms 66–8
 ecological setting 86–9
 ecology of 66–77
 evolutionarily stable strategies (ESS) 85, 90–1, 93–8
 further theoretical studies 99–100
 further directions 100
 inferences about 76–7
 modeling 85–100
 quantitative genetic recursions (QGR) 85, 91–8
 character divergence 63–5
 among species 70
 environmental effects 70–1
 experimental demonstration 82–5
 in guilds 70
 rapid 78
 rapid evolution on islands 78–81
 rates of 77–81
 character expression 68
 character patterns, community-level 74–7
 character shifts 63
 coevolved 63
 chemosynthetic symbiosis 351–3
 chiasma formation 189
ChLA-A108 128
ChLA-A126 128
Chlamydomonas 372, 374
Chlorella vulgaris 350, 361
 chlorophyll 366
Chlorophyta 366, 367
 chloroplast DNA 141
 chondrodystrophic anomalies 274
- chloroplast DNA 140
 chromatids 186
Chromophyta 366, 367
 chromosomal changes 17
 chromosomal rearrangements 228, 229
 chromosomal speciation 11–12
Chroomonas caudata 367
Chrysocolaptes lucidus 72
Chrysophyceae 366
 cladism 318
 cladistic methodology 323, 324
 cladograms 318, 321, 329, 330
 climate change 234
Clione limacina 294–6
Cnemidophorus 70, 85
Cnemidophorus tigris 70, 78
 coevolution 63–109
 environmental effects 100
 further theoretical studies 99–100
 coevolutionary dynamics, modeling 85–6
 coevolutionary stable community (or CSC)
 approach 85, 89–90, 93–8
 Cold Spring Harbor 13, 14, 20
Columba livia 294–6
 Columbian mammoth 268
 community assembly, modeling 85–100
 community population dynamics 89
 comparative anatomy 291
 comparative biology 335
 competition
 avoidance 70
 function 87
 competitive ability, laboratory studies 81
 competitive asymmetry 87
 competitive character displacement,
 conditions for 68–9
 competitive character divergence 67
 complex organismal functions 325
 complexity, evolution of 324–33
 constructional morphology 335, 336
 correlational selection 328
Corvus cornix 16
Corvus corone 16
 Cretaceous 239
Critidinia 357
 critical migration rate 58
 crop evolution 135
Cryptophyceae 367
Cryptophyta 367
 ctDNA 140, 141, 172
 cyanobacteria 350, 351, 355, 366
Cyanophora paradoxa 364
 cyrtomatodonts 333
- Daphne system 77
 Darwinism 4, 5, 7, 10, 15, 16, 19, 20, 25,
 26, 28–30, 242
Darwin's Finches 8
 daughter species 233

- deductive method 334–5
 degenerate sites, two-fold 115–16
 deleterious recessive 209
 demes
 combination of selection within and among 57
 genetic differentiation among 54, 58
 individual selection within 59
 descriptive morphology 317
 development and evolution 315
 developmental biology 28
 advances in 314
 domain of 314
 developmental morphology 314–24
 diazotrophic bacteria 354
 diazotrophic symbionts 353
Dicaeum tristrami 19
 dichopatric speciation 14
Dicrurus megarhynchus 19
 dietary generalists 262
 dietary selectivity 271
Dinophyta 367
Dinopium 72
 dinosaurs, numbers of genera in major taxa of 238
 diploidization 200
 diploids, types of zygotes 191
 disturbance and diversity 240–1
 diversity
 and disturbance 240–1
 and trophic resources 235–9
 DNA 229, 317, 364, 372
 sequences 140, 365
 variation in humans 111–34
 dominance variance for fitness 54
 dosage effect hypothesis 209
 double reduction phenomenon 186–9
 driver 82–3
Drosophila 13, 15, 29, 111, 222
Drosophila melanogaster 2, 36, 111, 118, 119
Drosophila nuclear DNA 117–19
Drosophila pseudoobscura 25, 118, 119
Drosophila simulans 2, 118
 ducks 327
 dumbbell (dichopatric) model 16
 dwarf mammals 275
 dwarfed elephants, evolutionary patterns and processes 276–81
 dwarfing syndromes 276
 dwarfism 260
 ecological amplitude 224–6
 ecological segregation 69
 ecology and morphology 314
 ecomorphology 308–14
 domain of 308
Eleagnus 353, 354
 elephants, dwarfed 260, 264–8
Elephas antiquus 268, 269, 276
Elephas falconeri 268, 269, 272–4, 276
Elephas maximus 273, 274, 278
Elephas namadicus 268
Eleutherodactylus 224
 elytron length 83
Embryos and Evolution 291
 empty ecospace 230
 endosymbioses 347
 energy survival crisis 270
 enhanced mating segregation 68
 enhanced pleiotropy 228
 environmental change 240
 environmental interactions 259
 environmental rigor 226
 enzyme alleles 26
 epistasis
 and Shifting Balance Theory 37, 57–9
 definitions of 36
 in adaptive evolution 50–9
 in evolution 35
 in genetics 36
 epistatic inheritance 21
 equilibrium genetic relationships 228
 essential nutrients in symbiosis 356–7
Euglenophyta 366, 367
 eukaryotes, metabolic capabilities acquired by symbiosis 347–61
 eukaryotic flagella and microtubules 371–2
 eutrophic conditions 239
 evolution
 and development 315
 of complexity 324–33
 synthetic theory of 291
Evolution (journal) 29
 evolutionarily stable strategies (ESS) 85, 90–1, 93–8
 evolutionary biology
 changes in 1–34
 philosophies of 335
 evolutionary dynamics 89, 90, 91
 evolutionary embryology 314
 evolutionary morphology 289–93, 323, 328, 337, 338
 current focus in 294
 multidimensional approaches 333
 evolutionary replacement 239
 evolutionary scenarios 319, 329–31
 evolutionary synthesis 1, 4–5, 7, 11, 17, 24, 27–30
 evolutionary systematics 6
 evolutionary theory, development of 290
 experimentation 292
 factorial designs 42
 FAD 356
 FAD₂ 356
 fasting endurance 263, 269
 feeding mechanics 327

- finches, morphology of 308–10
 fish feeding mechanisms 310–11
 fish heads, heaters in 296–8
 fitness
 additive effect on 55–6
 and gene interactions 48–50
 dominance variance for 54
 evaluation of 46
 function 87
 genetic variance in 52
 of organisms 328
 fixation 228
 flapping locomotion 294–6
 flock aggregation 72
 food
 shortages 263
 supply 309–10
 Foraminifera 239
 forest elephant 264
 form–function–performance relationships
 321
 fossil record 69
 founder effect 228
 four-fold degenerate sites 115–16
Frankia 353–5, 362
 functional analysis 321, 334
 functional attributes 319–21
 functional characters 319, 322, 323
 functional morphology 293–8, 323
 new orientation in 293
 functionalism 336–7
 Galapagos Islands 308, 309
 gamma-Proteobacteria 354
 garter snakes 329
 gene combinations 37
 gene dosage effects 210
 gene flow, across ploidy levels 198–200
 gene frequency
 analysis of 193–4
 dynamics 185
 gene interactions 35–62
 and fitness 48–50
 fitness component argument 49–50
 in adaptive evolution 45–50
 gene transfer
 from symbiont-organelle to host nucleus
 365
 horizontal 372–3
 genetic algebra 188
 genetic differentiation among demes 54, 58
 genetic diversity, inferred from phenotypic
 and molecular evidence 140–9
 genetic reorganization 228
 genetic revolutions 15, 228, 229
 genetic syndromes 274
 genetic transilience 228, 229
 genetic variability 111, 112
 genetic variance 36
 in fitness 52
 genetic variation 2, 277
Genetical Theory of Natural Selection, The
 46
 genetics and geneticists 1, 3, 4, 7, 13–16,
 20–2
Genetics and the Origin of Species 2, 3, 18
 genome
 analysis 189
 in speciation events 29
 genotype 39, 291
 genotype-by-environment interactions 46
 geographic speciation 2, 4, 10–12, 14, 18,
 20, 220, 228
Geospiza difficilis 73
Geospiza fortis 77
Geospiza fuliginosa 73, 77
 gestation time 272
Giardia 370, 371
 gigantism 260
Glenodinium foliaceum 369
 gradual sympatric speciation 12
 graviportal animals 273–6
 group selection 24–5, 262
 guild patterns 74–7
Gunnera 354
Gymnodinium acidotum 367
 heaters in fish heads 296–8
 hemoglobin α chain locus 127
 herbivores, number of species of 238
 heritability 92
 heterochronic evolution 273–6
 heterochrony 315–16
 analysis of 316
 heterogeneous genetic effect of
 homogeneous selection 57
 heterosis 207–10
 heterozygosity, rate of decline for different
 mating systems 192
 heterozygotes 191
 historical factors, testing role of 325
 HLA-A locus 128
 HLA-A11E allele 128
 HLA-B locus 128
Homo erectus 120, 121, 125, 126
Homo sapiens 128
 origins and evolution 119–22
 homogeneous selection, heterogeneous
 genetic effect of 57
 homozygotes 191
 human evolution, bottleneck in 127–9
 hummingbirds 224
 Hutchinsonian ratio 74
 hybridization 9, 67
Hydrobia ulvae 78
Hydrobia ventrosa 78
 hydrogenosomes 363–72

- immigrant selection 261
 inbreeding 191–4
 coefficient 191–3, 208
 of finite populations of polyploids 193
 depression 207–10
 generation matrix analysis 191, 192
 path analysis 191
 index of separation 188
Indica, *see under* Rice
 inductive method 334
 inheritance 5
 innovations 233
 insect flight
 biomechanical analysis 302
 evolution of 302–4
 insect wings
 aerodynamic characteristics 300
 evolution of 299–302
 loading of 303
 thermoregulatory effects 300
 instantaneous (sympatric) speciation 12
 insular mammals, trends for 261
 integrative evolutionary morphology 324–33
 interaction effects 41, 43
 interactions, parallels between Neyman and Wright 47–8
 interdemic dispersion 53
 interdemic selection 37
 intermediate disturbance hypothesis 241
 interpopulational gene differentiation (G_{ST}) 146
 interspecific territoriality 72
 intrinsic molecular processes 228
 invasion scenario 100
 invertebrates, cellulolytic symbioses 360
 island faunas 261
 island rule for mammalian body size 259–63
 islands
 extinction on 243–4
 predation on 261
 rapid evolution on 78–81
 isolate differentiation 227–30
 isolate formation 221–4
 extrinsic mechanisms 222–3
 intrinsic mechanisms 223–4
 macroevolutionary significance of
 relative importance of 243
 probability of 223, 224
 rates of 222
 specialized species 226
 isolate persistence 224–7, 243
 rates of 236
 tropical habitats 234
 isolated populations 220
 environmental factors 222–3
 extinction probability 244
 vulnerability to extinction 225
 Isthmus of Panama 223
 Japonica, *see under* Rice
Kentrophorus 351, 353
 key adjustments 231
 key innovations 231, 232
 key inventions 231
 key modifications 231
Lachnolaimus maximus 310
 Lamarckian theory 23–4
 Lamellibranchia 351
 land bridge islands 79
 latitudinal diversity gradients 233–5, 238
Leersia hexandra 151
 Leguminosae 353
 leiognathid fish 358
Lens culinaris 82
 Lepidoptera 360
Lepomis gibbosus 311
 lichens, symbiont-derived carbon 351
 limb
 bones 274–75
 posture 273
 linkage disequilibrium level 196
 linkage equilibrium frequency 195
Litoria 67
 litter size 272
 lizards 319
 local naturalists 7
 loci, linkage between 194–6
 loliginids 358
 Lotka-Volterra competition 87, 88
Loxodes 370
Loxodonta africana africana 264–5
Loxodonta africana cyclotis 264–5
Loxodonta pumilio 264–5
 luminescent symbionts 357–8
 macroevolution 11, 15, 18–20, 241–2
 main effects 40–1
 major histocompatibility complex (MHC)
 class I genes 128
 Malta 269
 mammalian body size, island rule for 259–63
 mammalian quadrupeds 268
 mammals, species richness of 237
Mammuthus columbi 268, 273, 274, 280
Mammuthus exilis 268, 273, 274, 281
 mass selection within demes 37
Mastigella 356
Material Basis of Evolution 27
 mathematical genetics 22
 meiosis, pairing behavior at 186
 meiotic configurations 188–90

Mendelian inheritance 274
 Mendelian Laws 1
 Mendelism 1
 mesenchymal cell lineages 316
 mesenchymal skeletogenesis 316
 metabolic rate 259
 metamonads 370
 metapopulations 220–1
 methanogens, symbionts as electron sinks 356
Metopus 356
 microbial symbionts 359
 microbial waste products 359
 microsporidians 370
 microtubule organizing centres (MTOCs) 371
 Miocene 239
 mitochondria 356, 363–72
 mitochondrial DNA (mtDNA) 112, 117, 122, 123, 125, 128
 molecular biology 22, 29, 298, 315
 molecular data 122–7
 molecular genetic techniques 69, 101
 monocentrid fish 358
 monomorphic species 90
 monomorphism 127
 morphogenetic fabrication 336
 morphological differentiation 76
 morphological evolution 77, 79
 morphological species 9
 morphologists 28
 morphology 289–345
 and ecology 314
 contribution to development of evolutionary theory 290
 in systematic analysis 317
 non-contributory 291
 philosophies and methodologies of 333–8
 renaissance in 289
 sub-disciplines 289
 synthetic work in 291
 motoneuron excitation 298
 multicellularity
 biomechanical properties conferred by 305
 in plants 305–8
 multimodal resource peaks 77
 multiplication of species 11
 multiplicity of genic effects 45
 multiregional models 121, 127
 mutation-selection balance 210
 mutations 5
 mycetocyte symbioses 356–7
Myrica 353, 355
 Myricaceae 353
 NADH 356
 natural selection 16, 17, 22–5, 49, 328–9
 naturalists 20

naturalists–systematists 1
 nature
 hypothetical model 45
 investigating and characterizing 42
 Neyman's model 45
 neo-Darwinism 336–7
 neural crest 330–1
 neutral evolution 25–30
 neutral mutation hypothesis 127
 New Guinea 8
New Systematics, The 17
 Neyman model of nature 40
 niche diversification 239
 niche space 226
Nicotiana 160
 nitrogen-fixing symbiosis 353–5
 nitrogenase 353, 355
 reductase 353
 ‘Noah's Ark’ hypothesis 119–20, 122
 non-dimensional species concept 9
 non-geographic speciation 12
 non-heritable factors 47
 non-heterochronic mechanisms 316
 non-random species assortment 77
Nostoc 354, 361
Nostocales 354
Notechis ater niger 80
 nuclear DNA 172
 nucleotide diversity 112–19
 nutrient availability 239
 nutritional deficiency 262
 oceanic islands 79
 oligotrophic conditions 239
 ontogenetic repatterning 325
 organellar DNA 172
 organelles
 evolutionary origins 366–71
 symbiont-derived 363–72
 targeting of nucleus-encoded polypeptides 366
 organismal biology 298
 organismal complexity 324
 organic form, theory of 314–15
Origin of Species 8
Oriolus 72
 Orthoptera 360
Oryza 136
Oryza barthii 137, 138, 140; 142, 143, 145, 151, 152, 160
Oryza barthii A. Chev. 136
Oryza brachyantha A. Chev. et Roehr. 137
Oryza breviligulata 152
Oryza breviligulata A. Chev. et Roehr. 136
Oryza cubensis Ekman 136
Oryza glaberrima 137, 138, 140, 142, 145, 151, 160, 162, 163, 167

- Oryza glaberrima* Steud. 136
Oryza glumaepatula 152, 164
Oryza glumaepatula Steud. 136
Oryza longistaminata 138, 141, 142, 145, 152, 160, 164
Oryza longistaminata A. Chev. et Roehr. 136
Oryza meridionalis 164
Oryza meridionalis Ng 136
Oryza meyeriana Baill 137
Oryza nivara 138
Oryza nivara Sharma et Shastry 136
Oryza officinalis 137
Oryza officinalis Wall. ex Watt 136
Oryza perennis 136
Oryza punctata 137
Oryza ridleyi Hook 137
Oryza rufipogon 136–8, 140–5, 147–57, 162–4, 168
Oryza rufipogon Griff. 136
Oryza sativa 137–42, 144–7, 149–51, 160–4, 170, 175
Oryza sativa f. *spontanea* 138
Oryza sativa L. 136
Oryza sativa subsp. *indica* 166
Oryza sativa subsp. *japonica* 166
‘Out of Africa’ hypothesis 120, 121, 122, 127
overdominance 207–8, 210
oviposition preference 83
pairwise character divergence 74–6
Palaeomyces 362–3
pan-adaptationism 336
parabasalians 370
Paracoccus denitrificans 369
parallel character shifts 65
parallel clinal variation 69
parapatric speciation 12, 16
Parasponia 353
Pelomyxa 356, 370
perciform fish 358
performance 328–9
Peridinium balticum 368
peripatric speciation 14–16
Perissodactyls 239
Peromyscus 6
peroxisomes 363
Perviata 351
Phaethorninae 224
Phalodrilus 351
phenotype 291
 distribution function 87
phenotypic diversity 145
phenotypic position vector 89
phenotypic variation 46, 275
Philemon moluccensis 72
Photobacterium 358
Photobacterium leiognathi 358
Photobacterium phosphoreum 358
photosynthetic symbionts 349–51
phycobilisomes 366
Phylloscopus amoenus 19
phylogenetic analysis 69, 98, 317–24
phylogenetic hypothesis 318, 319, 321, 328, 330
phylogenetic methodology 69
phylogenetic scenarios 329
Physcia tenella 351
pigeon 294–6
plant bodies, hydrostatic 308
plant symbiosis 355
plant tissue, biophysics of 306
plants
 character convergence 73
 multicellularity in 305–8
 species concept 9
plastids 350, 363–72
 complex 367
 eukaryotic hosts of 367
 possible origins of membranes enclosing 368
pleiotropy 45
Pleistocene 234, 260, 261, 265–9, 280
Pliocene 239
pluralism 25–30
pluteus larvae 316
Pogonophora 351
polypeptides, nucleus-encoded 366
polyploidization 205–10
polyploids
 determination of recombination frequencies 194
 establishment of 198–200
 hypothesis of genetic mechanisms for success 205–10
inbreeding coefficients of finite populations of 193
inbreeding depression 208, 209
meiosis in 186
overdominance in 207–8
population genetics of 185, 204
quantitative genetics of 201–4
recombination within chromosomes and recombination resulting from reassortment of chromosomes into gametes 195
segregation of 186–90
selection 196–8
 sib mating in 191
polyploidy 185
 reviews 185
polysaccharides 359
polytocous species 262
polytypic species 9
population constriction 128
population dynamics 90, 91

- population formation 220–1
 population genetics 6, 11, 14, 15, 20, 21
 autopolyploidy 185–217
 population size
 and stability 225
 attained at ecological equilibrium 89
 population structure 223
 population subdivision 38, 56, 57
 and additive genetic variance 51–4
 populational persistence 277
 populations, fates of 221
 postcranial elements 274
 Precambrian 369
 predation on islands 261
 predators
 and body size 261
 mobbing 72
 proboscideans 265–8
 dwarfed 259
 prochloron 350
Prochlorothrix 366
 propionic acid 359
 Proteobacteria 369
 protists 370
Protobranchia 351
Pseudotrebouxia 350
 pteropod mollusc 294
 punctuated equilibria 19
 pygmy elephant 264
- quantitative genetic recursions (QGR) 85, 91–8
 quantitative trait loci (QTL) 144
- random chromatid segregation (RCdS)
 186, 188, 190–1, 193
 random chromosome segregation (RCeS)
 186, 188, 190–1, 193
 random drift 228
 random genetic drift 37, 55–6
 random mating equilibrium (RME) 190–1
 rDNA 140, 141, 169, 172
 recessive–deleterious hypothesis 208, 209
 recombination 37
 between loci 195
 regional continuity model 121
 reinforcement 68
 reproductive character divergence 66
 reproductive isolation 16–18
 reptiles, species richness of 236
 resource
 capture trait 83
 competition 63
 limitation 262
 utilization 68, 83, 92
 rhacophorid frogs 311
Rhizobia 355
Rhizobium 353
Rhizosolenia 354
- Rhodophyta* 366, 367
Rhynia 362–3
 ribulose-1, 5-bisphosphate carboxylase 349, 365
- rice
 association between propagating system and mating system 153–4
 association between sterility genes and photoperiod sensitivity genes 166
 cytoplasmic male sterility 163–4
 delimitation of wild ancestors of cultigens 137–9
 differences between wild and domesticated forms 143–4
 diversification by human selection 144–5
 domestication and differentiation of Indica and Japonica types 170–3
 dynamics of natural populations 154–6
 evolutionary studies 135–84
 F statistics 155–6
 F_1 sterility 159–63, 175
 formation of profound barriers through gene complexes 164–6
 gametic disequilibrium 149
 gametic selection in distant crosses 175
 gene pool concept 136–7
 genetic structure 154–6
 genic sterility 159–63
 geographical distribution of genetic diversity 146–8
 hybrid sterility as internal barrier 159
 Indica–Japonica differentiation of cultivars 166–75
 intermediate wild-cultivated plants 139–40
 intrapopulational gene diversity 155
 intrapopulational genetic variation 145–6
 intrapopulational structure on fine spatial scale 157–9
 KC1O₃ resistance 166–7, 172
 life history traits 149–51
 multilocus system 148–9
 parapatric subpopulations at Bhubaneswar and Chiangrai 157–9
 partitioning of total variation into inter- and intrapopulational components 155
 pollinating allocation 154
 putative internal mechanisms of Indica–Japonica differentiation 173–5
 rate of selfing vs. outbreeding 152
 reproductive allocation 154
 sterility barriers developed between and within taxa 159–66
 taxonomic status 136–7
 upland culture role in evolution of Japonica type 170

- variation among and within taxa having genome AA 140–3
 variation in isozymes and DNA of Indica and Japonica types 167–70
 variation in mating system 151–3
- Riftia* 352
Riftia pachyptila 352
 rodents 262
 giant 259
 rRNA sequence 369
- saltationism 20
 San Clemente 280
 San Miguel 281
 Santa Barbara 260
 Santa Catalina Islands 281
 Santa Rosa 281
 scenariograms 318, 321
 sea urchins 316
 segregation effects 47
 selection
 polyploids 196–8
 pressure 229
 processes 228
 theory 336
 towards allopolyploid inheritance 200–1
 selfed diploids 191
 selfed tetraploids 191
 semi-geographic speciation 12, 16
 serial endosymbiosis theory 364, 370
 sexual dimorphism 275
 sexual selection 17, 20
 shifting balance experiment 54
 Shifting Balance Theory 35–62
 three-phase process 37
 shipworms 360
 Sicily 261, 269; 271, 281
 Silurian 239
 simple effects 41
 single-cell marking techniques 316
 small experiments 48–9
 snail-eating sunfish 311
 solar energy 238
Solemya 352
 Solomon Islands 8
Sorex araneus 78
Sorex minutus 78
 speciation 10–13
 and adaptation 242
 rate, categories of cause for 242
 speciation evolution 15, 19
 species concept 9–10
 species-energy hypothesis 238
 species index 89
 spirochaetes
 motility 371
 tubulin in 372
 statistical analysis 292
 sterility 16
- barriers developed between and within taxa 159–66
 stilbonematid nematodes 353
 structural tissue 263
 structuralism 335–7
 structure–function relationship 294, 298, 308, 313, 314
Studies in the Structure and Development of Vertebrates 291
 stunting 262
Symbiodinium 350, 351
 symbiosis 347
 and horizontal gene transfer 372–3
 as source of metabolic capabilities 347–61
 cellulolysis in 358–60
 chemosynthetic 351
 essential nutrients in 356–7
 evolutionary consequences of 361–3
 evolutionary origins of organelles 366–71
 exploitation of novel habitats and diets 362
 generation of novel host taxa 363
 luminescence in 357–8
 methanogens as electron sinks 356
 morphological and other changes in host 361–4
 nitrogen-fixing 353–5
 organelles 363–72
 photosynthetic 350–1
 photosynthetic carbon fixation 350–1
 provision of organic carbon 349–53
 structural organization of eukaryotic host 365–6
 targeting of nucleus-encoded 366
 transfer of genes from symbiont/organelle to host nucleus 365
- symbiotic cnidarians 351
 symbiotic foraminiferans 350
 symecomorphosis 309
 sympatric generalists 226
 sympatric specialists 226
 sympatric speciation 12, 13
 sympathy 69, 70
 systematics 8, 317–24
Systematics and the Origin of Species 4, 5, 6, 10, 13, 18
- target species 82–3
 taxon cycles 98
 teeth 275–6
 teleost fish 361
 temperate vs. tropical organisms 234–5
Teredinibacter turneri 354
 termites 360
 tetraploids
 case of two alleles in 203
 cross-over per chromosome rate 195

- establishment of 199
inbreeding depression 208, 209
maximum recombination rate 195
meiotic configuration 189
mixed mating population 193
segregation patterns 186
 different possible genotypes of zygotes for 188
thyroxin stimulation 316
Trebouxia 350
Tribolium 53
triploids, meiotic configuration 189
trochilines 224
trophic pyramid 236
trophic resources
 and diversity 235–9
 increased levels of 236
 variable or fluctuating 236
tropical diversity 234–5
tropical vs. temperate organisms 234–5
truss analysis 331
trypanosomes 357
tubulin in spirochaetes 372
two-fold degenerate sites 115–16
Type II error 43
Typhlonectidae 324
typological–morphological species 8
typological species concept 9
unit character 291
unit of selection 24
universal pleiotropy 46
3' untranslated (UT) region 112, 116
5' untranslated (UT) region 112, 116
upwelling zones 239
- Varanus rosenbergi* 80
Variation and Evolution in Plants 9
Veronica peregrina 159
vertebrates
 cellulolysis in 359
 evolution in 330
 gut symbioses in 359
vesicular–arbuscular mycorrhizal (VAM)
 fungi 362–3
Vestimentifera 351
Vibrio 358
Vibrio fisheri 358
vicariance
 abiotic factors 222
 biotic factors 222
Vigna radiata 82
volatile fatty acids (VFAs) 359, 361
whole-organism approaches 335
Xanthoria parientina 351

OXFORD SURVEYS IN EVOLUTIONARY BIOLOGY

Volume 1: 1984

Edited by R. Dawkins and M. Ridley

Contents

THE EVOLUTION OF CHEMICAL SIGNAL-RECEPTOR SYSTEMS: FROM SLIME MOULDS TO MAN

John Tyler Bonner

EVOLUTIONARY ORIGINS OF SEX

Lynn Margulis and Dorion Sagan

ON DECEPTION, A WAY OF ALL FLESH, AND FIREFLY SIGNALLING AND SYSTEMATICS

James E. Lloyd

TWO WAYS TO BE A TROPICAL BIG MOTH: SANTA ROSA SATURNIIDS AND SPHINGIDS

Daniel H. Janzen

SIZE INEQUALITIES AND FITNESS IN PLANT POPULATIONS

Ottot T. Solbrig and Dorothy J. Solbrig

DYNAMICS OF MULTILOCUS GENETIC SYSTEMS

Michael T. Clegg

HIERARCHY AND EVOLUTION

Niles Eldredge and Stanley N. Salthe

NARROW APPROACHES TO PHYLOGENY: A REVIEW OF NINE BOOKS ON CLADISM

Michael T. Ghiselin

OXFORD SURVEYS IN EVOLUTIONARY BIOLOGY

Volume 2: 1985

Edited by R. Dawkins and M. Ridley

Contents

A DEFENSE OF REDUCTIONISM IN EVOLUTIONARY BIOLOGY

George C. Williams

A GEOMETRIC VIEW OF RELATEDNESS

Alan Grafen

MEASURES OF SEXUAL SELECTION

William J. Sutherland

ISSUES IN BRAIN EVOLUTION

Harry J. Jerison

MODELS OF DIVERSITY AND PHYLOGENETIC RECONSTRUCTION

T. S. Kemp

FISHER'S EVOLUTIONARY FAITH AND THE CHALLENGE OF MIMICRY

John R. G. Turner

THE R. A. FISHER-SEWALL WRIGHT CONTROVERSY AND ITS INFLUENCE UPON MODERN EVOLUTIONARY BIOLOGY

William B. Provine

ESSAY REVIEW: THE RELATIONSHIP BETWEEN DEVELOPMENT AND EVOLUTION

Keith Stewart Thomson

OXFORD SURVEYS IN EVOLUTIONARY BIOLOGY

Volume 3: 1986

Edited by R. Dawkins and M. Ridley

Contents

WHY NOT NEO-DARWINISM? A CRITIQUE OF PALEO-BIOLOGICAL CHALLENGES

Max K. Hecht and Antoni Hoffman

PATTERNS IN CODON USAGE OF DIFFERENT KINDS OF SPECIES

Richard Grantham, Pascale Perrin, and Dominique Mouchiroud

THE COST OF REPRODUCTION

Graham Bell and Vassiliki Koufopanou

THE CAUSAL LOGIC OF NATURAL SELECTION: A GENERAL THEORY

F. T. Cloak, Jr

RONALD FISHER AND THE DEVELOPMENT OF EVOLUTIONARY THEORY: I. THE ROLE OF SELECTION

Egbert Giles Leigh, Jr

THE NEWER SYNTHESIS? SOME CONCEPTUAL PROBLEMS IN EVOLUTIONARY BIOLOGY

John A. Endler

OXFORD SURVEYS IN EVOLUTIONARY BIOLOGY

Volume 5: 1988

Edited by P. H. Harvey and L. Partridge

Contents

EVOLUTIONARY ASPECTS OF PROTEIN SYNTHESIS

Michael Bulmer

MULTIGENE AND SUPERGENE FAMILIES

Tomoko Ohta

THE ORIGIN OF MOLLUSCS IN THE LIGHT OF MOLECULAR EVIDENCE

Michael T. Ghiselin

HOW FUNDAMENTAL ARE FISHERIAN SEX RATIOS?

J. J. Bull and Eric L. Charnov

THE EVOLUTION OF FEMALE MATE PREFERENCES FOR MALE GENETIC QUALITY

Andrew Neil Pomiankowski

PARENTAL EFFORT IN BIRDS AND MAMMALS: THEORY AND MEASUREMENT

David W. Winkler and Gerald S. Wilkinson

CULTURAL VECTORS, VIRULENCE, AND THE EMERGENCE OF EVOLUTIONARY EPIDEMIOLOGY

Paul W. Ewald

A CRITICISM OF J. R. G. TURNER'S ARTICLE 'FISHER'S EVOLUTIONARY FAITH AND THE CHALLENGE OF MIMICRY'

A. J. Cain

REPLY: MEN OF FISHER'S?

John R. G. Turner

OXFORD SURVEYS IN EVOLUTIONARY BIOLOGY

Volume 4: 1987

Edited by P. H. Harvey and L. Partridge

Contents

J. B. S. HALDANE

John Maynard Smith

MOLECULAR EVOLUTION AND THE NEUTRAL ALLELE
THEORY

John H. Gillespie

MOLECULAR POPULATION GENETICS

Martin Kreitman

MODELS OF EVOLUTIONARY CHANGE AND THE
ESTIMATION OF EVOLUTIONARY TREES

Adrian Friday

THE EVOLUTION AND BIOGEOGRAPHY OF ASSOCIATIONS
BETWEEN LYCAENID BUTTERFLIES AND ANTS

Naomi E. Pierce

THE EVOLUTION OF CLUTCH SIZE IN INVERTEBRATES

H. C. J. Godfray

NEST BUILDING AS A FACILITATING AND LIMITING
FACTOR IN THE EVOLUTION OF EUSOCIALITY IN THE
HYMENOPTERA

Michel Hansell

WHAT IS BET-HEDGING?

Jon Seger and H. Jane Brockmann

RONALD FISHER AND THE DEVELOPMENT OF EVOLU-
TORY THEORY: II. INFLUENCES OF NEW

VARIATION ON EVOLUTIONARY PROCESS

Egbert Giles Leigh, Jr

OXFORD SURVEYS IN EVOLUTIONARY BIOLOGY

Volume 6: 1989

Edited by P. H. Harvey and L. Partridge

Contents

WEISMANN AND MODERN BIOLOGY

John Maynard Smith

**LIFE HISTORY VARIATION IN PLACENTAL MAMMALS:
UNIFYING THE DATA WITH THEORY**

Paul H. Harvey, Andrew F. Read, and Daniel E. L. Promislow

**COMPARATIVE METHODS USING PHYLOGENETICALLY
INDEPENDENT CONTRASTS**

Austin Burt

CASTE AND CHANGE IN SOCIAL INSECTS

Deborah M. Gordon

INCLUSIVE FITNESS IN A NUTSHELL

David C. Queller

**LEVELS OF SELECTION AND SORTING WITH SPECIAL
REFERENCE TO THE SPECIES LEVEL**

Elisabeth S. Vrba

**THE EMERGENCE, MAINTENANCE, AND TRANSITIONS OF
THE EARLIEST EVOLUTIONARY UNITS**

Eörs Szathmáry

**POPULATION GENETICS AT THE DNA LEVEL: A REVIEW OF
THE CONTRIBUTION OF RESTRICTION ENZYME STUDIES**

Andrew J. Leigh Brown

OXFORD SURVEYS IN EVOLUTIONARY BIOLOGY

Volume 7: 1990

Edited by D. Futuyma and J. Antonovics

Contents

GENE GENEALOGIES AND THE COALESCENT PROCESS
Richard R. Hudson

PRINCIPLES OF GENEALOGICAL CONCORDANCE IN
SPECIES CONCEPTS AND BIOLOGICAL TAXONOMY
John C. Avise and R. Martin Ball, Jr

HYBRID ZONES: WINDOWS ON EVOLUTIONARY PROCESS
Richard C. Harrison

EVOLUTIONARY PROCESS AS STUDIED IN POPULATION
GENETICS: CLUES FROM PHYLOGENY
Hampton L. Carson

SEXUAL SELECTION, SENSORY SYSTEMS AND SENSORY
EXPLOITATION
Michael J. Ryan

B CHROMOSOMES, SELFISH DNA AND THEORETICAL
MODELS: WHERE NEXT
M. W. Shaw and G. M. Hewitt

CO-EVOLUTION BETWEEN TWO SYMBIOTHS: THE CASE OF
CYTOPLASMIC MALE-STERILITY IN HIGHER PLANTS
*D. Couvet, A. Atlan, E. Belhassen, C. Gliddon, P. H. Gouyon and
F. Kjellberg*

STUDYING THE EVOLUTION OF PHYSIOLOGICAL PERFORMANCE
Albert F. Bennett and Raymond B. Huey

PLANT CONSUMERS AND PLANT SECONDARY CHEMISTRY:
PAST, PRESENT AND FUTURE
May R. Berenbaum