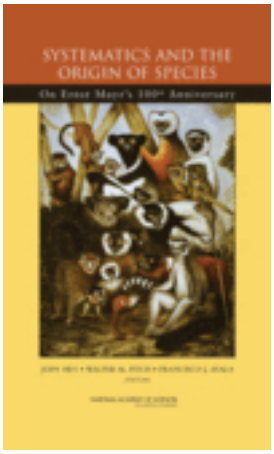


Systematics and the Origin of Species: On Ernst Mayr's 100th Anniversary

Jody Hey, Walter M. Fitch, Francisco J. Ayala, Editors, National Academy of Sciences

ISBN: , 382 pages, 6 x 9, (2005)

SYSTEMATICS AND THE ORIGIN OF SPECIES

On Ernst Mayr's 100th Anniversary

Jody Hey, Walter M. Fitch, and Francisco J. Ayala
Editors

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS • 500 Fifth St, N.W. • Washington, D.C. 20001

This volume is based on the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Systematics and the Origin of Species: On Ernst Mayr's 100th Anniversary," held December 16-18, 2004, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, California. The articles appearing in these pages were contributed by speakers at the colloquium and have not been independently reviewed. Any opinions, findings, conclusions, or recommendations expressed in this volume are those of the authors and do not necessarily reflect the views of the National Academy of Sciences.

Library of Congress Cataloging-in-Publication Data

Systematics and the origin of species : on Ernst Mayr's 100th anniversary / Jody Hey, Walter M. Fitch, and Francisco J. Ayala, editors.
p. cm.
Includes bibliographical references.
ISBN 0-309-09536-0 (hardcover) —ISBN 0-309-54760-1 (pdf) 1. Species. 2. Biology —Classification. I. Mayr, Ernst, 1904-2005. II. Hey, Jody. III. Fitch, Walter M., 1929- IV. Ayala, Francisco José, 1934-
QH380.S97 2005
576.8'6—dc22
2005017917

Additional copies of this book are available from the National Academies Press, 500 Fifth St., N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-33313 (in the Washington metropolitan area); Internet, http://www.nap.edu.

Cover design by Michele de la Menardiere; illustration by Stephen D. Nash/ Conservation International.

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

Preface

Ernst Mayr, one of the 20th century's greatest scientists and a principal author of the modern theory of evolution, passed away on February 3, 2005, at the age of 100. From December 16 to 18, 2004, before Mayr's passing, a colloquium on ''Systematics and the Origin of Species'' sponsored by the National Academy of Sciences was held in his honor. The colloquium's title was the same as that of Mayr's 1942 book, generally considered one of the foundation books of the theory of evolution. The 17 papers that follow explore current knowledge about the main topics of Mayr's book.

The modern theory of evolution embodies a complex array of biological knowledge centered around Darwin's theory of evolution by natural selection couched in genetic terms. It is not one single theory with its corroborating evidence, but a multidisciplinary body of knowledge bearing on biological evolution: an amalgam of well established theories and working hypotheses together with the observations and experiments that support accepted hypotheses (and falsify rejected ones), which jointly seek to explain the evolutionary process and its outcomes. These hypotheses, observations, and experiments originate in disciplines such as genetics, developmental biology, neurobiology, zoology, botany, paleontology, and molecular biology.

Darwin's theory of evolution (1859) argued that natural selection, the process accounting for the adaptation and diversity of organisms, emerges as a necessary conclusion from two premises: (*i*) the assumption that hereditary variations useful to organisms occur and (*ii*) the observation

that more individuals are produced than can possibly survive. A serious difficulty facing Darwin's evolutionary theory was the lack of an adequate theory of inheritance that would account for the preservation through the generations of the variations on which natural selection was supposed to act. Theories then current of ''blending inheritance'' proposed that offspring struck an average between the characteristics of their parents. As Darwin became aware, blending inheritance could not account for the conservation of variations, because differences among variant offspring would be halved each generation, rapidly reducing the original variation to the average of the preexisting characteristics.

The missing link in Darwin's argument was provided by Mendelian genetics. Mendel's paper published in 1866 formulated the fundamental principles of a theory of heredity that accounts for biological inheritance through particulate factors (now called ''genes'') inherited one from each parent that do not mix or blend but segregate in the formation of the sex cells, or gametes. Mendel's discoveries, however, remained unknown to Darwin and, indeed, did not become generally known until 1900, when they were simultaneously rediscovered by several scientists.

The synthesis of Mendelian genetics with Darwin's theory of natural selection was initially accomplished in the 1920s and 1930s through the theoretical work of several geneticists who used mathematical arguments to show, first, that continuous variation (in such characteristics as size, number of progeny, longevity, and the like) could be explained by Mendel's laws and, second, that natural selection acting cumulatively on small variations could yield major evolutionary changes in form and function. Distinguished members of this group of theoretical geneticists were R. A. Fisher (1930) and J. B. S. Haldane (1932) in Great Britain and Sewall Wright (1931) in the United States. Their work provided a theoretical framework for the integration of genetics into Darwin's theory of natural selection but had a limited impact on contemporary biologists because (*i*) it was formulated in a mathematical language that most biologists could not understand; (*ii*) it was almost exclusively theoretical, with little empirical corroboration; and (*iii*) it was limited in scope, largely omitting many issues, like speciation, that were of great importance to evolutionists.

The synthesis accomplished by the theoreticians was greatly expanded in the following decades by biologists coming from various disciplines who enlarged the initial theoretical synthesis with relevant concepts and theories and provided supporting empirical evidence. Several books are considered emblematic of this original expansion of the theory in addition to Mayr's *Systematics and the Origin of Species* (1942), notably, Theodosius Dobzhansky's *Genetics and the Origin of Species*, published in 1937, George Gaylord Simpson's *Tempo and Mode in Evolution* (1944), and G. Ledyard Stebbins' *Variation and Evolution in Plants* (1950). Three earlier

colloquia sponsored by the National Academy of Sciences were dedicated to current knowledge concerning the distinctive topics originally explored in these books (Ayala *et al.*, 2000; Ayala and Fitch, 1997; Fitch and Ayala, 1995).

One key development of the theory of evolution is the replacement of ''population thinking'' by ''typological thinking.'' Darwin had postulated that hereditary variations occur in organisms that are useful to the organisms themselves. Natural selection could only occur if such variations were pervasive. In genetics, populational thinking gave rise to a new branch of genetics that, as Dobzhansky (1937) noted, ''has as its province the processes taking place in groups of individuals—in populations—and therefore is called the genetics of populations. . . . The rules governing the genetic structure of a population are distinct from those governing the genetics of individuals.''

Mayr's *Systematics and the Origin of Species* (1942) represents a self-conscious effort to explicate the significance of population variation in the understanding of evolutionary processes and the origin of new species. ''It is true that the change from the static species concept of Linnaeus to the dynamic species concept of the modern systematist has not entirely escaped the attention of progressive students of genetics and evolution. However, the whole significance of the polytypic species, of the phenomena of geographic variation, of the differences between geographic and other forms of isolation are by no means as widely appreciated . . . as they deserve'' (1942).

Mayr would later write: ''Systematics, contrary to widespread misconceptions, . . . was not at all in a backward and static condition during the first third of the 20th century. . . . Population thinking was widely adopted, and, as a consequence, variation within and between populations was actively studied, which led to the development of the biological species concept, to the widespread adoption of polytypic species taxa, and to the study of species in space and time as adapted systems. . . . [T]he experimental geneticists, with few exceptions, were quite unaware that a populational species concept had been widely adopted by naturalists'' (1980a). ''The biological species concept emphasizes the species as a community of populations, reproductive isolation, . . . and the ecological interactions of sympatric populations that do not belong to the same species'' (1980a). Species as taxonomic entities and, most of all, as populations and units of evolution have remained Mayr's supreme subject of intellectual engagement. In his most recent book, Mayr writes: ''The species is the principal unit of evolution. A sound understanding of the biological nature of species is fundamental to writing about evolution and indeed about almost any aspect of the philosophy of biology. . . . I define biological species as 'groups of interbreeding natural populations that are repro-

ductively (genetically) isolated from other such groups.' The emphasis of this definition is . . . on genetic relationship. . . . This new interpretation of species of organisms emphasizes that biological species are something very different from the natural kinds of inanimate nature'' (2004).

Ernst Mayr was born on July 5, 1904, in Kempton, Bavaria, Germany. On July 1, 1926, he became an assistant at the University Museum in Berlin ''but left for New Guinea and the Salomon Islands in February 1928. I did not return until the end of April 1930'' (Mayr, 1980b). He came to the United States in 1931 to be curator for birds at the American Museum of Natural History in New York. In 1953, he became Alexander Agassiz Professor of Zoology at Harvard University, where from 1961 to 1970 he was director of the Museum of Comparative Zoology and retired from the faculty in 1974.

Mayr's scholarly publications span >80 years, starting with his first two scientific papers, published in 1923, and reaching to 2004. Walter J. Bock, who has written a fairly comprehensive overview of Mayr's career, divides his contributions into three major periods. ''The first period (1923 until 1953 when he left the American Museum of Natural History) was devoted mainly to avian systematics and the theory of systematics. This work formed the foundation for the second period (beginning in 1942 but becoming more dominant in the latter part of the 1940s and lasting until his formal retirement from Harvard University in 1974), which was devoted largely to evolutionary theory. His systematic and evolutionary contributions, in turn, provided the basis for the last period (beginning in the early 1970s), devoted chiefly to the history and philosophy of biology'' (Bock, 1994).

The bibliography listed by Bock includes 176 publications (1923–1994), all but a baker's dozen single-authored by Mayr. Some of Mayr's most important books, in addition to *Systematics and the Origin of Species*, are *Animal Species and Evolution* (1963), *Principles of Systematic Zoology* (1969), *Populations, Species, and Evolution* (1970), *Evolution and the Diversity of Life* (1970), *The Growth of Biological Thought* (1982), and *Toward a New Philosophy of Biology* (1988). Remarkably, Mayr has continued publishing essays, articles, and books to the present. Mayr's most recent book, *What Makes Biology Unique? Considerations on the Autonomy of a Scientific Discipline* (2004), was published in August 2004, one month after his 100th birthday. In 2001, at age 97, Mayr had published two other important books: Ernst Mayr and Jared Diamond's *The Birds of Northern Melanesia: Speciation, Ecology and Biogeography* (2001) and Ernst Mayr's *What Evolution Is* (2001). Mayr's other recent books are *This Is Biology: The Science of the Living World* (1997) and *One Long Argument: Charles Darwin and the Genesis of Modern Evolutionary Theory* (1991).

The colloquium honoring Mayr's book and its legacy featured 17

presentations, including one by E. O. Wilson, "Introductory Essay: Systematics and the Future of Biology" (Chapter 1), which appears immediately after this preface. Wilson makes the point that the tremendous growth in molecular and cellular biology will be hampered if it is not balanced by similar progress in our understanding of biological diversity. He argues forcefully for growth in systematics and biodiversity research and for the establishment and increasing use of Internet-based virtual collections so that researchers and laypersons can freely access the resources of museums worldwide.

The remaining chapters in this volume are distributed in four parts: Part I, The Origins of Species Barriers; Part II, Discerning Recent Divergence; Part III, The Nature of Species and the Meaning of ''Species''; and Part IV, Genomic Approaches and New Insights on Diversity.

REFERENCES

Ayala, F. J. & Fitch, W. M., eds. (1997) *Genetics and the Origin of Species: From Darwin to Molecular Biology 50 Years After Dobzhansky* (Natl. Acad. Sci., Washington, DC).

Ayala, F. J., Fitch, W. M. & Clegg, M. T., eds. (2000) *Variation and Evolution in Plants and Microorganisms: Toward a New Synthesis 50 Years After Stebbins* (Natl. Acad. Press, Washington, DC), pp. xi, 340.

Bock, W. J. (1994) Ernst Mayr, naturalist: His contributions to systematics and evolution. *Biol. Philos*. **9**, 267–327.

Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection* (Murray, London).

Dobzhansky, T. (1937) *Genetics and the Origin of Species* (Columbia Univ. Press, New York).

Fisher, R. A. (1930) *The Genetical Theory of Natural Selection* (Clarendon, Oxford).

Fitch, W. M. & Ayala, F. J., eds. (1995) *Tempo and Mode in Evolution: Genetics and Paleontology 50 Years After Simpson* (Natl. Acad. Press, Washington, DC), pp. viii, 325.

Haldane, J. B. S. (1932) *The Causes of Evolution* (Longmans, Green & Co., London).

Mayr, E. (1942) *Systematics and the Origin of Species* (Columbia Univ. Press, New York).

Mayr, E. (1980a) Prologue: Some thoughts on the history of the evolutionary synthesis. In *The Evolutionary Synthesis*, eds. Mayr, E. & Provine, W. (Harvard Univ. Press, Cambridge, MA), pp. 1–48.

Mayr, E. (1980b) How I became a Darwinian. In *The Evolutionary Synthesis*, eds. Mayr, E. & Provine, W. (Harvard Univ. Press, Cambridge, MA), pp. 413–423.

Mayr, E. (2004) *What Makes Biology Unique? Considerations on the Autonomy of a Scientific Discipline* (Cambridge Univ. Press, Cambridge, U.K.).

Mendel, G. (1866) Versuche über Pflanzen-Hybriden. *Verhandlungen des Naturforschenden Vereines in Brünn* **4**, 3–47.

Simpson, G. G. (1944) Tempo and Mode in Evolution (Columbia Univ. Press, New York).

Stebbins, G. L. (1950) *Variation and Evolution in Plants* (Columbia Univ. Press, New York).

Wright, S. (1931) Evolution in Mendelian populations. *Genetics* **16**, 97–159.

Contents

PART IV: GENOMIC APPROACHES AND NEW INSIGHTS ON DIVERSITY/265

1

Introductory Essay: Systematics and the Future of Biology

EDWARD O. WILSON*

Biology is a science of three dimensions. The first is the study of each species across all levels of biological organization, molecule to cell to organism to population to ecosystem. The second dimension is the diversity of all species in the biosphere. The third dimension is the history of each species in turn, comprising both its genetic evolution and the environmental change that drove the evolution. Biology, by growing in all three dimensions, is progressing toward unification and will continue to do so. A large part of the future of biology depends on interdisciplinary studies that allow easy travel across the three dimensions.

Molecular and cellular biology, the subdisciplines of maximum support and activity today, occupy the two lowest levels of biological organization. They are focused on the first dimension in a small set of model species, selected primarily for their ease of culturing and the special traits that make them amenable for different kinds of analysis. The triumph of molecular and cellular biology has been the documentation of one of the two overarching principles of biology: that all living phenomena are obedient to the laws of physics and chemistry. The other overarching principle of biology is that all living phenomena originated and evolved by natural selection. That, in turn, has been the triumph of organismic and evolutionary biology.

Viewed in the framework of the history of biology, the subdisciplines

*Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA 02138-2902.

of molecular and cellular biology are in the natural history period of their development. This perhaps surprising characterization can be clarified with the following metaphor. The cell is a system consisting of a very large number of interacting elements and processes. It can be compared to a conventional ecosystem, such as a lake or forest, in this sense: Researchers are discovering the anatomy and functions of the vast array of molecules, the equivalent of the species of plants and animals in ecosystems, that compose the cell. These scientists are the Humboldts, the Darwins, the Mayrs, and other explorer-naturalists in a new age, and of a new kind. Mercifully free of mosquito bites and blistered feet, they press forward into the unmapped regions at the lowest levels of biological organization. They are not in the business of creating fundamental principles, which they take mostly from physics and chemistry. Their spectacular success has come instead from technology invented and applied with creative genius. They render visible, by crystallography, immunology, genic substitution, and other methods, the anatomy and functions of the ultramicroscopic inhabitants of the cell, which are otherwise entirely beyond the range of the unaided human senses. They aim and can be expected in time to join with researchers in other subdisciplines of biology to develop the fundamental principles of biological organization.

There remains the rest of biology, the vast and mostly unexplored upper reaches of the first dimension (levels of organization), and the little known second dimension (diversity) and third dimension (evolution). These domains do not belong to the past of biology, nor are they antiquated and declining in any manner or to any degree whatsoever, as is sometimes misperceived. They are, in large part, the future of biology. Probably <10% of species on the planet have been discovered, and of these only a tiny fraction—<1%—have been addressed with more than a cursory anatomical description.

Keep in mind that each species is unique in its genotype, proteome, behavior, history, and environmental relationship to other species. Until we learn more about the immense array of Earth's little known and entirely unknown species, how they came into being, and what they are doing in the biosphere, the development of the rest of biology, including molecular and cellular biology, will be vastly incomplete. A further consequence of the imbalance is that the relation of humanity to the rest of the biosphere will stay largely uncharted territory. And the means to save and manage the living environment for the long term will remain very much a guessing game.

The proportionate shortfall of the disciplines can be expressed in practical terms as follows. A large part of the success of molecular and cellular biology is due to their relevance to medicine. In public perception and support, they are virtually married to medicine. Hence, molecular and

cellular biology are rich not so much because they have been successful; rather, they are successful because they have been rich. What needs to be appreciated for the future of organismic and evolutionary biology in practical terms is that where molecular and cellular biology are vital to personal health, organismic and evolutionary biology are vital to environmental health, and thence also to personal health.

The exploration of life on this little known planet needs to be made the equivalent of the Human Genome Project. It should bean all-out effort that sets the global biodiversity census as a goal with a timetable and not just a result eventually to be reached. In fact, the technology now exists to speed exploratory and monographic systematics by at least an order of magnitude. This technology includes high-resolution digital photography with computer-aided automontaged depth of focus for even the smallest and hence most highly magnified specimens. As a practicing systematist, I am certain that the first order of business must be to thoroughly photograph representatives of all species for which either type specimens or indisputably authenticated substitutes exist. The images can then be published on the Internet for immediate access on command, available to anyone, at any time, anywhere in the world.

Such virtual type collections can in most cases remove most of the time-consuming necessity of visiting museums or securing loans of often fragile specimens. Combined with online reproductions of the original published descriptions and earlier monographs, they will accelerate the identification of vast numbers of specimens now backed up in unprocessed collections around the world. They will also allow the rapid preparation of revisionary monographs, local biodiversity analyses, and the publication of made-to-order field guides.

Systematics is now in the early stages of a technology-driven revolution. A conference of biodiversity and informatics leaders held at Harvard University in 2001 agreed that by using new methods already available and including genomics, systematists have the capacity to complete or nearly complete a global biodiversity survey within a single human generation—25 years—at about the cost of the Human Genome Project. To give a better feel for the magnitude of this biological moon shot, consider that whereas very roughly 10% of Earth's species (out of, say, 15–20 million extant) have been discovered and named in the past 250 years since Linnaeus inaugurated the hierarchical and binomial system of classification, now it seems possible that the remaining 90% can be so covered in 1/10th of that time.

If such a goal seems out of reach, consider that perhaps 1million species of all kinds, mostly eukaryotes, have type specimens in good enough condition for electronic republication. The New York Botanical Garden has already processed and put online the types of 90,000 plant

species, and Harvard's Museum of Comparative Zoology is well on its way to processing ≈28,000 insect species. Thus, even in its earliest stages, at only two institutions, the republication program has already reached the hypothetical 10% level.

A new global systematics initiative is under way on three fronts. The first is the all-species program, which aims to measure the full breadth of biodiversity in all of the three recognized domains, the Bacteria, Archaea, and Eucarya, by using new and future technology. The next is the Encyclopedia of Life, expanding the all-species program by providing an indefinitely expansible page for each species and containing information either directly available or by linkage to other databases. The final and rapidly growing body of knowledge is the Tree of Life, the reconstructed phylogeny of life forms in ever finer detail, with particular reliance on genomics.

The upper levels of biological organization, from organism to ecosystem, the mapping and analysis of biodiversity, and the development of the Tree of Life all of the way from genes to species will eventually amount to most of biology. These proportionately still-neglected domains, therefore, offer intellectual stock of substantial growth potential to universities and other research-oriented organizations that invest in them now. It is still relatively easy to provide leadership at the cutting edge of biology extended beyond the molecular and cellular levels of a few species. The cost would be low, and the returns to scale incalculably great.

Part I

THE ORIGINS OF SPECIES BARRIERS

Mayr was well known for his championship of the biological species concept and for asserting a predominant role for the geographic separation of populations in the diversification process that gives rise to separate species. The genetic version of this perspective is, to a good approximation, the Dobzhansky–Muller model of divergence, in which genes that have been the site of adaptive fixations within separate populations may also be the site of negative epistatic interactions in species hybrids and cause inviability or sterility when hybridization occurs (Dobzhansky, 1936; Muller, 1940; Orr, 1995). But what are these adaptations that accumulate within separate populations and give rise to reproductive barriers? This question is addressed by Allen Orr in "The Genetic Basis of Reproductive Isolation: Insights from *Drosophila*" (Chapter 2), and he explains what we know from the growing handful of cases in which the actual genes that contribute to low hybrid fitness have been isolated. The most striking commonality to emerge from these studies is that these genes have extraordinarily rapid rates of adaptive amino acid replacement. This finding makes sense, for if we suppose that some fraction of amino acid replacements are prone to negatively epistatic interactions when placed in a hybrid background, then those genes that have the highest rates of amino acid replacement will also tend to be those that cause these types of Dobzhansky–Muller incompatibilities.

This discovery of rapidly evolving genes that contribute to reproductive barriers also necessarily focuses our attention on the kinds of phenotypes and genes that are particularly prone to evolve rapidly. Classically,

much of the discussion on rapid population divergence tended to focus on the kinds of environmental or geographic circumstances that might promote rapid evolution. However, in recent years, the attention has shifted to situations where intraspecies and intragenomic conflicts can lead to rapid evolution of genes (Rice, 1998). Such conflicts arise whenever natural selection favors alleles that are penetrant under some circumstances (such as in one sex), even though those same alleles may reduce other components of fitness that are manifest in other contexts (such as in the other sex). Genomic conflicts can lead to tit-for-tat, or arms-race, evolution between groups of genes within the same genome. In "Inter-Locus Antagonistic Coevolution as an Engine of Speciation: Assessment with Hemiclonal Analysis" (Chapter 3), William Rice *et al.* explore this issue directly by developing a model evolutionary system with *Drosophila melanogaster*. In this system, individual haploid chromosome complements (hemiclones) are drawn, using genetic tricks that are possible with *Drosophila*, from a longstanding laboratory population. Once isolated, each hemiclone can be measured, by replicating in combination with other hemiclones, for its net effect upon fitness of particular phenotypes. The approach allows a careful assessment of the selection gradient and additive genetic variance for traits that enhance fitness in males but reduce fitness in their female mates.

The findings that rapid evolution of genes, including that caused by genomic conflict, can lead to the formation of reproductive barriers notwithstanding, there remains the question of how much gene flow can be tolerated between diverging populations if speciation is to occur (Wright, 1940). Not even populations with many rapidly evolving genes can be expected to become reproductively isolated from other populations if gene flow rates are high. With this point in mind, Francisco Ayala and Mario Coluzzi, in "Chromosome Speciation: Humans, *Drosophila,* and Mosquitoes" (Chapter 4), explore models in which recombination suppressors, such as chromosomal inversions, can enhance the opportunity for adaptive divergence in the face of gene flow between parapatric populations (Coluzzi, 1982; Noor *et al.*, 2001; Rieseberg, 2001). These models are more plausible than those in which chromosomal inversions enable divergence by causing low hybrid fitness, and they are supported particularly by recent evidence from *Drosophila* and *Anopheles.*

Regardless of the rates and roles of genetic changes that contribute to divergence, there remain very large questions about the phenotypic manifestation of these changes. Mary Jane West-Eberhard, in "Developmental Plasticity and the Origin of Species Differences" (Chapter 5), addresses these questions and stresses that phenotypes may derive from genes in ways that are highly contingent upon other genes and upon environmental circumstances. Thus, single new alleles or new environmental circum-

stances alone, in the absence of genetic changes, may trigger large changes in the phenotype. Furthermore, if genetic accommodation is common, such as by the Baldwin effect (Baldwin, 1896; Simpson, 1953), then phenotypic variation that arises primarily by environmental causes may play a driving role in divergence.

REFERENCES

Baldwin, J. M. (1896) A new factor in evolution. *Am. Nat.* **30,** 441–451.

Coluzzi, M. (1982) Spatial distribution of chromosomal inversions and speciation in anopheline mosquitoes. In *Mechanisms of Speciation*, ed. Barigozzi, C. (Liss, New York), pp. 143–153.

Dobzhansky, T. (1936) Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* **21,** 113–135.

Muller, H. J. (1940) Bearings of the *Drosophila* work on systematics. In *The New Systematics*, ed. Huxley, J. (Clarendon, Oxford), pp. 185–268.

Noor, M. A., Grams, K. L., Bertucci, L. A. & Reiland, J. (2001) Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci. USA* **98,** 12084–12088.

Orr, H. A. (1995) The population-genetics of speciation—The evolution of hybrid incompatibilities. *Genetics* **139,** 1805–1813.

Rice, W. R. (1998) Intergenomic conflict, interlocus antagonistic coevolution, and the evolution of reproductive isolation. In *Endless Forms: Species and Speciation*, eds. Howard, D. J. & Berlocher, S. H. (Oxford Univ. Press, New York), pp. 261–270.

Rieseberg, L. H. (2001) Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16,** 351–358.

Simpson, G. G. (1953) The Baldwin effect. *Evolution* **7,** 110–117.

Wright, S. (1940) Breeding structure of populations in relation to speciation. *Am. Nat.* **74,** 232–248.

2

The Genetic Basis of Reproductive Isolation: Insights from *Drosophila*

H. ALLEN ORR*

Recent studies of the genetics of speciation in *Drosophila* have focused on two problems: (*i*) identifying and characterizing the genes that cause reproductive isolation, and (*ii*) determining the evolutionary forces that drove the divergence of these "speciation genes." Here, I review this work. I conclude that speciation genes correspond to ordinary loci having normal functions within species. These genes fall into several functional classes, although a role in transcriptional regulation could prove particularly common. More important, speciation genes are typically very rapidly evolving, and this divergence is often driven by positive Darwinian selection. Finally, I review recent work in *Drosophila pseudoobscura* on the possible role of meiotic drive in the evolution of the genes that cause postzygotic isolation.

Ernst Mayr made at least three contributions that are fundamental to the genetic study of speciation. The first, and surely most important, was his codification of what we mean by species and thus by speciation. In his book, *Systematics and the Origin of Species*, Mayr (1942) famously argued that species are best defined by the Biological Species Concept: species are groups of actually or potentially interbreeding natural populations that are reproductively isolated from other such groups. Reproductive isolation is thus the *sine qua non* of good species.

Mayr's second contribution, elaborated in his book, *Animal Species*

*Department of Biology, University of Rochester, Rochester, NY 14627.

and Evolution (1963), was his argument that reproductive isolation usually evolves in allopatry. This has, of course, been an enormously controversial claim, and I will not enter into it in detail. Suffice it to say that the weight of evidence, both biogeographic and comparative, now strongly suggests that Mayr was correct. Although sympatric speciation may well occur, reproductive isolation appears usually to evolve in allopatry (reviewed in Coyne and Orr, 2004, chapters 3 and 4).

Mayr's third key claim, introduced in his 1942 book but elaborated in Mayr (1954) and later in *Animal Species and Evolution*, was that genetic drift plays a critical role in speciation. According to Mayr, large populations suffer a sort of evolutionary inertia: the conservative forces of gene flow and epistasis prevent, or at least render unlikely, the evolution of novel morphologies or new coadapted gene complexes in large, geographically widespread species. Such evolution is, he suggested, far more likely in island or peripheral populations founded by a few individuals. Genetic drift in such populations may allow rapid evolution to new fitness peaks, permitting the evolution of reproductive isolation between ancestral and founder populations. Such "founder effect" models of speciation proved extraordinarily popular, representing, as Provine (1989) noted, "the favored explanation for at least island speciation since 1954." Despite this popularity, it is difficult to point to unambiguous evidence for founder effect speciation, and the idea has grown controversial. I return to this issue below.

The research program that Mayr helped to formulate, that to understand the origin of species one must understand the origin of reproductive isolation, has experienced a renaissance over the last two decades. Evolutionists have performed an impressive number of new and careful studies of the biogeography, ecology, and genetics of speciation, and we now understand a good deal about the evolution of reproductive isolation (Coyne and Orr, 2004). We can, for instance, describe the rate at which this reproductive isolation evolves (Fig. 2.1), and the ecological factors that drive speciation, at least in some cases (Schluter, 2000). We can also point to plausible examples in which the process of reinforcement may have completed speciation (Coyne and Orr, 2004; Noor, 1999; Servedio and Noor, 2003). Finally, we know a fair amount about the number and location of the genes that cause reproductive isolation (at least in its postzygotic form), and the causes of patterns like Haldane's rule (Coyne and Orr, 2004; Laurie, 1997; Orr, 1997).

Given this progress, it is natural to ask: What now are the major outstanding problems in speciation? On the genetical front (I am not competent to address any other), the answers seem clear. We must (*i*) find the genes that cause reproductive isolation and (*ii*) identify the evolutionary forces that drove their divergence.

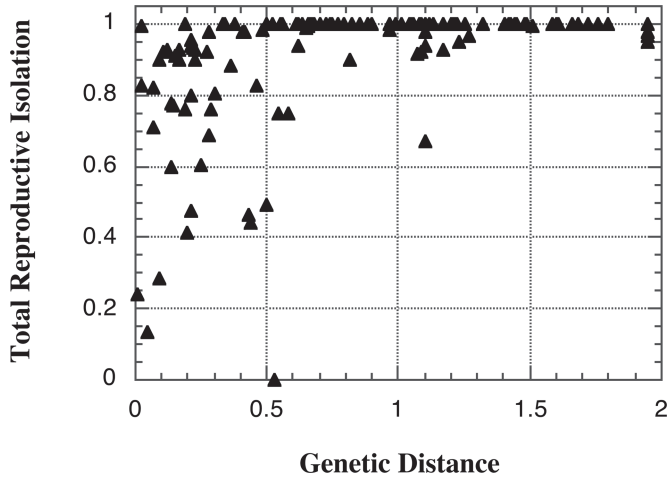


FIGURE 2.1 The rate of increase in "total" reproductive isolation with genetic distance in *Drosophila.* Postzygotic isolation is measured such that a value of 0 means no reproductive isolation (either pre- or postzygotic) and a value of 1 means complete reproductive isolation. Genetic distance is measured by Nei's genetic distance, which increases approximately linearly with time over the values shown. Data include both allopatric and sympatric species pairs. Although the measure of total reproductive isolation shown includes aspects of both pre- and postzygotic isolation, it is imperfect and generally excludes forms of reproductive isolation (e.g., ecological isolation) that are not readily measured in the laboratory. Data are based on Coyne and Orr (1989, 1997). [Reproduced with permission from Coyne and Orr, 1997 (Copyright 1997, International Journal of Organic Evolution).]

THE PROBLEM

Biologists who do not specialize in speciation are invariably surprised to learn that evolutionists have identified very few genes causing reproductive isolation. The reason for this slow progress is, however, simple. If species are taxa that are reproductively isolated, a genetics of speciation must, almost by definition, be a genetics where such a thing is not possible, between organisms that do not exchange genes. The consequence of this methodological dilemma is that evolutionists have been unable to address a large set of fundamental questions about the genes that cause reproductive isolation, so-called speciation genes. (This perhaps unfortunate term, which is now entrenched in the literature, refers to any locus that causes reproductive isolation, whether in F_1 or later-generation hybrids, and whether the gene was among the first to cause isolation or not.)

TABLE 2.1 Speciation Genes Causing Postzygotic Reproductive Isolation That Have Been Identified at the DNA Sequence Level

Gene	Taxon	Hybrid Phenotype	Gene Type	References
Xmrk-2	*Xiphophorus*	Inviability	Receptor tyrosine kinase	Wittbrodt *et al.* (1989), Schartl *et al.* (1999), Malitschek *et al.* (1995)
OdsH	*Drosophila*	Sterility	Transcription factor	Ting *et al.* (1998), Wu and Ting (2004), Sun *et al.* (2004)
Hmr	*Drosophila*	Inviability	Transcription factor	Barbash *et al.* (2003, 2004)
Nup96	*Drosophila*	Inviability	Nucleoporin	Presgraves *et al.* (2003)

These questions include: Are speciation genes "ordinary" genes that have normal functions within species? If so, do speciation genes fall into one or at least a few functional classes? Are substitutions in speciation genes concentrated in coding or regulatory sequences? And do speciation genes diverge by natural selection or genetic drift? This last question is perhaps the most important: if we can identify speciation genes at the level of DNA sequences, we should be able to bring to bear a powerful set of molecular population genetic tools (e.g., McDonald–Kreitman and Hudson–Kreitman–Aguade tests) on the Mayrian question of the relative roles of deterministic vs. stochastic forces in the origin of species.

The methodological dilemma that plagues speciation genetics is sufficiently serious that, until recently, evolutionists could point to only a single gene that causes postzygotic reproductive isolation (Table 2.1). (I focus on postzygotic isolation because it has been the subject of most genetic analysis, both in general and in my own laboratory; see, however, Metz and Palumbi, 1996; Palumbi, 1992; Vacquier, 1998; Swanson and Vacquier, 2002, for discussion of genes causing prezygotic isolation.) It had long been known that crosses between certain fish species of the genus *Xiphophorus*, e.g., between the platyfish and swordtail, result in hybrid inviability. In particular, backcross hybrids between these species often develop malignant melanomas and die (Schartl *et al.*, 1994). Half a

century of genetic study has revealed that this hybrid lethality results from a two-locus "Dobzhansky–Muller" incompatibility: one locus on the X chromosome of the platyfish is incompatible with another (apparently single) locus on an autosome of the swordtail. Melanoma results when these loci are brought together in a hybrid genome. Molecular genetic analyses allowed identification and characterization of the X-linked partner in this incompatibility: *Xmrk-2*. *Xmrk-2* encodes a novel receptor tyrosine kinase, which, while found as a duplicate gene on the platyfish X, is absent from the swordtail X (Wittbrodt *et al.,* 1989; Schartl *et al.,* 1999; Malitschek *et al.*, 1995). *Xmrk-2* appears to be misexpressed in species hybrids, causing tumor formation (Malitschek *et al.*, 1995).

Xmrk-2 was subjected to intense genetic analysis because of its role in cancer. There is little reason to believe, however, that postzygotic isolation usually involves malignancies. The recent renaissance in the genetics of speciation has therefore featured a determined effort to find additional speciation genes regardless of the form of hybrid inviability or sterility involved. One cannot, after all, draw robust conclusions about the identities or evolutionary histories of the genes causing reproductive isolation from a sample of one. Fortunately, these recent efforts, by a number of laboratories, have largely succeeded, and we now know the identities of several additional speciation genes. I briefly review this work. I then turn to recent efforts to characterize the evolutionary forces that drive the divergence of such genes. As we will see, these forces may sometimes assume a surprising form.

THE SEARCH FOR SPECIATION GENES

The general methodological dilemma facing the genetics of speciation involves a number of particularly unfortunate special cases. Perhaps worst of all, one of our most sophisticated genetic model systems, *Drosophila melanogaster*, forms only inviable or sterile hybrids when crossed to any other species. As a result, *D. melanogaster* has, until recently, proved nearly useless in the study of speciation.

Drosophila geneticists have taken two approaches to sidestep this problem. The first step involves study of non-*melanogaster* fly species, wherein at least some of the tools first developed in *D. melanogaster* may be available (e.g., germ-line transformation, or *P* element insertions). This approach allowed identification of a gene that causes hybrid male sterility. Coyne and Charlesworth (1986) mapped a locus causing sterility of *Drosophila simulans–Drosophila mauritiana* hybrid males to a small region of the *D. mauritiana* X chromosome. Subsequent work (Perez and Wu, 1995; Perez *et al.*, 1993) refined this mapping, showing that at least two hybrid sterility factors reside in the region, one of which reduces hybrid fertility

by half when introgressed alone onto a *D. simulans* genetic background. Ting *et al.* (1998) identified this locus, *Odysseus*-Homeobox gene (*OdsH*), at the molecular level. As its name indicates, *OdsH* is an X-linked homeobox gene and thus a presumed transcription factor (Ting *et al.*, 1998; Wu and Ting, 2004). Recent transformation experiments appear to confirm that *OdsH* causes partial male sterility among backcross (although not F_1) hybrids (Sun *et al.*, 2004). Sterility may reflect misexpression of *OdsH* in hybrid testes (Wu and Ting, 2004). *OdsH* evolved rapidly between *D. simulans* and *D. mauritiana*, and this evolution was almost certainly driven by positive Darwinian selection: nonsynonymous substitutions greatly outnumber synonymous ones in the lineage leading to *D. mauritiana* (Ting *et al.*, 1998).

The second approach to identifying speciation genes has taken more direct advantage of *D. melanogaster*. Although any hybrids formed between this species and its closest relatives are completely sterile, this fact does not preclude its use in the search for hybrid *inviability* genes. Several workers (Barbash *et al.*, 2000; Hutter and Ashburner, 1987; Hutter *et al.*, 1990; Orr and Irving, 2000) used tools from *D. melanogaster* (e.g., chromosomal duplications and deletions, loss-of-function mutations, and germline transformation) to map and characterize an X-linked gene, *Hybrid male rescue* (*Hmr*), that causes the inviability of certain F_1 hybrids between *D. melanogaster* and its sister species, e.g., *D. simulans*. This work culminated in the recent identification by Barbash *et al.* (2003) of *Hmr* at the sequence level: *Hmr* encodes a transcriptional regulator of or related to the MYB family. It seems likely that *Hmr*'s presumed normal function in gene regulation is disrupted in species hybrids, causing inviability (Barbash *et al.*, 2003). Very recent work reveals that *Hmr* is also rapidly evolving, and that this evolution was also driven by positive Darwinian selection (Barbash *et al.*, 2004).

My laboratory has focused on deficiency mapping to locate and identify speciation genes. The critical point is that such mapping, which involves the introduction of cytologically defined chromosomal deletions from *D. melanogaster* into species hybrids, can be performed within F_1 hybrids, whose sterility is therefore irrelevant. Daven Presgraves (2003) used a large set of deficiencies from *D. melanogaster* to screen for autosomal genes from *D. simulans* that cause hybrid inviability when present with an X chromosome from *D. melanogaster*. The essence of this approach is that it allows detection and mapping of recessive, and thus, normally masked, speciation genes: if a chromosomal region includes a recessive hybrid inviability gene(s) from *D. simulans*, control hybrids that carry an undeleted (balancer) chromosome from *D. melanogaster* should be viable, whereas experimental hybrids that carry a deleted chromosome (in the relevant region) from *D. melanogaster* should be inviable.

Presgraves (2003) singly introduced >200 deficiencies into *D. melanogaster–D. simulans* species hybrids. He found that 20 deficiencies significantly reduced hybrid fitness, and that 10 of these deficiencies were hybrid lethal, i.e., uncovered genes from *D. simulans* that cause essentially complete inviability on a partly *D. melanogaster* genetic background. Further crosses using attached-X chromosome manipulations revealed that these *D. simulans* autosomal regions interact with the *D. melanogaster* X chromosome to cause hybrid inviability (Presgraves, 2003). This work implies that *D. melanogaster* and *D. simulans*, although diverged for only ≈2.5 million years (Hey and Kliman, 1993; Li, 1997), are separated by many genes causing hybrid inviability. Indeed, it appears that ≈10% of all viability-essential genes have diverged functionally between these species to such an extent that a gene from one species kills hybrids if made homozygous (hemizygous) in the genetic background of the other species (Presgraves, 2003).

In subsequent work, Presgraves *et al.* (2003) dissected a small deficiency region known to cause the inviability of *D. melanogaster–D. simulans* hybrids. A combination of fine-scale deletion mapping and complementation testing (Fig. 2.2) revealed that the hybrid inviability effect of this third chromosomal region mapped to a single complementation group, which was ultimately shown to correspond to *Nucleoporin96* (*Nup96*). This locus interacts with another (presently unknown) locus on the *D. melanogaster* X chromosome to cause hybrid inviability. *Nup96* encodes a component of the nuclear pore complex, a large structure that perforates the nuclear membrane of all eukaryotic cells and that regulates the trafficking of proteins and RNA in and out of the nucleus. Although the ≈30 nucleoporin proteins that constitute the nuclear pore complex are typically conserved evolutionarily, *Nup96* evolved rapidly between *D. melanogaster* and *D. simulans*. Molecular population genetic analyses further reveal that this divergence was driven by positive Darwinian selection in both species lineages. Because neither species shows evidence of recent adaptive sweeps (e.g., depressed nucleotide diversity or skewed frequency spectra), this adaptive evolution likely occurred early in the split of the two species (Presgraves *et al.*, 2003).

In very recent work, my laboratory has turned to the attempt to identify and characterize speciation genes that cause hybrid sterility, not inviability. The chief reason is that it is now clear that hybrid sterility, and, especially hybrid *male* sterility, evolves faster than hybrid inviability in several groups of animals, including *Drosophila* (reviewed in Coyne and Orr, 2004). Hybrid male sterility may thus represent a more important phenotype than hybrid inviability in the earliest stages of speciation. Our current efforts focus on the so-called *4-sim* hybrid sterility system: Muller and Pontecorvo (1942) and Pontecorvo (1943) showed long ago that *D.*

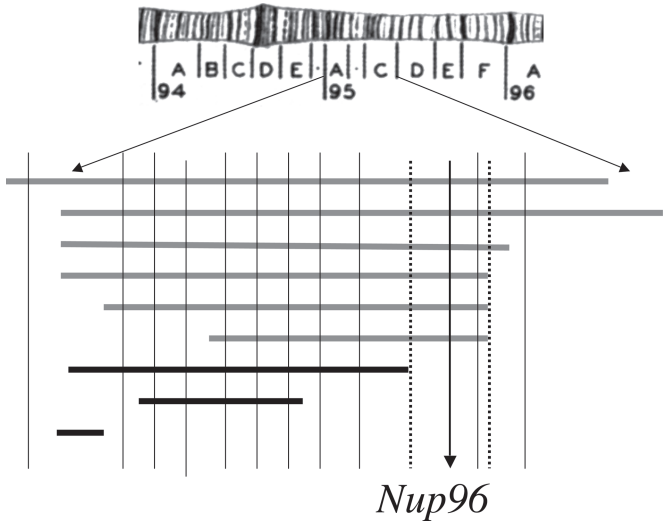


FIGURE 2.2 Deficiency and complementation mapping of a hybrid inviability gene. The region shown corresponds to cytological region 95 of the third chromosome of *D. melanogaster*. Bars shown in gray below the chromosome uncover recessive hybrid inviability in *D. melanogaster–D. simulans* hybrids and those shown in black do not. Vertical lines represent known viability-essential complementation groups in the region. Hybrid inviability maps to a single locus, *Nup96*. Data are based on Presgraves *et al.*, 2003. [Reproduced with permission from Presgraves *et al.*, 2003 (Copyright 2003).]

melanogaster males that are homozygous for the "dot" fourth chromosome from *D. simulans* are sterile; sterility results from sperm immotility. Heterozygous *4-sim* males and all genotypes of females are fertile. Deficiency mapping shows that hybrid male sterility localizes to a very small region of this very small chromosome (Orr, 1992; the fourth includes only ≈70 loci [Adams *et al.*, 2000]). It seems likely therefore that *4-sim* hybrid sterility reflects the action of a single locus that interacts with other loci elsewhere in the *D. melanogaster* genome. My laboratory has recently further narrowed the location of this putative hybrid sterility gene and has begun to test the few remaining candidate loci in the region for their possible role in hybrid sterility (J. P. Masly, personal communication).

In summary, the above genetic studies have finally allowed evolutionists to examine the factors that cause postzygotic isolation. It is clear that these factors correspond to ordinary genes, having normal functions within species: no evidence has been uncovered for the idea that speciation involves novel processes like the mass mobilization of transposable

elements in hybrids (Engels and Preston, 1979; Kidwell, 1983; Rose and Doolittle, 1983). It is also clear that speciation genes do not fall into a single functional class. Although two of the above genes (*OdsH* and *Hmr*) play a role in transcriptional regulation, and this occurrence could well prove common, others do not. *Nup96*, for example, encodes a structural protein. Similarly, it is clear that the divergence of noncoding regulatory sequences is not the invariable cause of postzygotic isolation: complementation tests with multiple mapped loss-of-function mutations show that *Nup96*'s effect on hybrid inviability is due to divergence of the Np96 protein itself (Presgraves *et al.*, 2003). Also, the genes causing postzygotic isolation are sometimes members of duplicate gene families (*Xrmk-2* and *OdsH*), but are sometimes single-copy genes (*Hmr* and *Nup96*). Although there is good evidence that the genes causing postzygotic isolation are on average partially recessive, as predicted by the dominance theory of Haldane's rule (reviewed in Coyne and Orr, 2004), it also clear that speciation genes can differ dramatically in dominance. *Nup96* and *Hmr*, for instance, act mostly recessively in hybrids, whereas the (currently unknown) gene(s) from *D. simulans* with which *Hmr* interacts to cause hybrid inviability appears quite dominant (Hutter *et al.*, 1990). Recessive speciation genes can obviously contribute only to the sterility or inviability of F_2, or backcross, not F_1, hybrids (unless, of course, the gene is X-linked). Whereas any given recessive gene may well cause less reproductive isolation than any given dominant one, recessive speciation genes appear more common than dominant; not surprisingly, then, later-generation hybrid problems often appear before F_1 problems (reviewed in Coyne and Orr, 2004, chapters 7 and 8).

Although the study of speciation genes remains in its infancy, two patterns do, so far at least, appear to characterize their evolution. First, speciation genes are rapidly evolving. Second, these genes often evolve by positive Darwinian selection (Barbash *et al.*, 2003; Presgraves *et al.*, 2003).

SPECIATION GENES: INTRAGENOMIC CONFLICT?

While the above work shows that positive selection plays a key role in the divergence of speciation genes, it says nothing about the nature of this selection. Departures from neutrality as revealed by, e.g., McDonald–Kreitman tests, are equally consistent with sexual selection, natural selection to abiotic factors, or natural selection to biotic factors. Surprisingly, recent work suggests that a particular form of the last hypothesis, one involving adaptation not to other organisms but to intragenomic events, may play a role in speciation. In particular, it now appears that meiotic drive may sometimes drive the evolution of hybrid sterility genes.

This idea was first suggested in the early 1990s by Frank (1991) and

Hurst and Pomiankowski (1991). These authors argued that mutations that cause meiotic drive may often become suppressed within species (because drive at one locus causes selection at other loci for suppression of drive), but can become reexpressed in species hybrids: if the mutations that suppress drive are less than completely dominant, they may fail to suppress segregation distortion in heterozygous hybrids. Frank (1991) and Hurst and Pomiankowski (1991) further speculated that such drive might sometimes cause the sterility of hybrids. Other variations on the meiotic drive theory of postzygotic isolation have appeared recently (Henikoff and Malik, 2002; Henikoff *et al.*, 2001; Tao and Hartl, 2003; Tao *et al.*, 2001).

Although early experimental work found no evidence of meiotic drive in species hybrids (Coyne and Orr, 1993; Johnson and Wu, 1992), more recent work has uncovered such evidence (Cazemajor *et al.*, 1997; Mercot *et al.*, 1995; Montchamp-Moreau and Joly, 1997; Tao *et al.*, 2001). In the most impressive of these efforts, Tao *et al.* (2001) showed that a <80-kb region of the third chromosome of *D. mauritiana* causes meiotic drive in an otherwise *D. simulans* genetic background. Remarkably, this same small region also causes hybrid male sterility, leading Tao *et al.* (2001) to conclude that the same gene, *too much yin* (*tmy*), likely causes both hybrid segregation distortion and hybrid sterility; *tmy* has not yet been identified at the DNA sequence level.

Recently, my laboratory has discovered strong segregation distortion in hybrids between two closely related subspecies of *D. pseudoobscura*. When females of the Bogota subspecies are crossed to males of the USA subspecies, fertile F_1 females and sterile F_1 males result. Although these males have traditionally been described as completely sterile, we have discovered that they are, in fact, weakly fertile. Surprisingly, these F_1 males produce nearly all daughters (≈95%) when crossed to any genotype of females (pure Bogota, pure USA, and hybrid F_1 sisters) (Orr and Irving, 2005). X-linked genetic markers reveal that this sex ratio distortion is not due to somatic sexual transformation (i.e., phenotypic females are genetic females) and egg-to-adult viability studies strongly suggest that it is not due to the inviability of sons (Orr and Irving, 2005). Instead, hybrid F_1 males appear to suffer severe segregation distortion. Although we do not yet know the proximate mechanism of segregation distortion (e.g., classical meiotic drive in the male germ line, immotility of Y-bearing sperm in the female reproductive tract, postfertilization failure of pronuclei fusion, etc.), we have completed a preliminary genetic analysis of this hybrid segregation distortion.

Our results reveal that the genes that normally suppress segregation distortion within the Bogota subspecies are autosomal (and perhaps also Y-linked) (Orr and Irving, 2005). More important, the genes that cause hybrid segregation distortion reside in several regions of the Bogota X

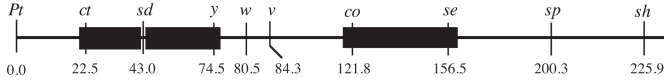


FIGURE 2.3 Regions of the *D. pseudoobscura* X chromosome implicated in hybrid sterility between the Bogota and USA subspecies (Orr and Irving, 2001). Recent genetic analysis of segregation distortion in the same hybridization implicates approximately the same chromosomal regions (Orr and Irving, 2005). The markers shown are all visible mutations. Kosambi-corrected map positions are provided. [Reproduced with permission from Orr and Irving, 2001 (Copyright 2001, Genetics Society of America).]

chromosome. Remarkably, these regions show essentially complete epistasis: little or no distortion occurs unless genes from both the left (*XL*) and right (*XR*) arms of the X derive from Bogota. These results are strikingly similar to those characterizing hybrid male sterility between the same subspecies (Orr, 1989; Orr and Irving, 2001; see Fig. 2.3). More suggestive still, the severity of hybrid sterility and the extent of offspring sex ratio distortion are strongly correlated across individual backcross hybrid males (Fig. 2.4). Future high-resolution introgression analysis

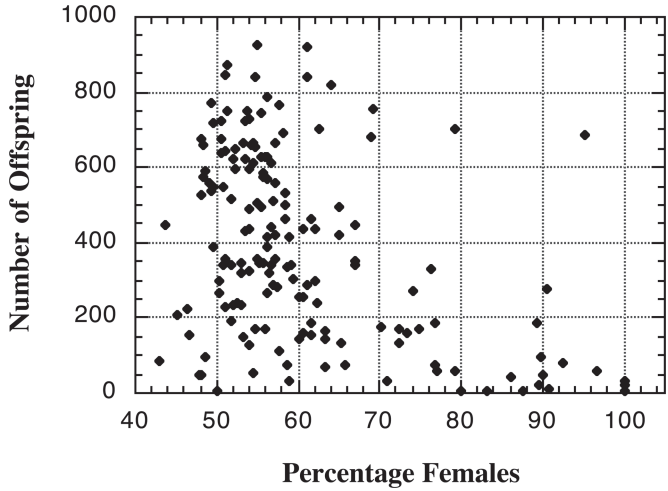


FIGURE 2.4 Correlation between *D. pseudoobscura* Bogota-USA hybrid fertility (as measured by offspring production) and hybrid segregation distortion (as measured by offspring sex ratio). Each data point reflects the offspring of a single recombinant backcross generation male. Data are based on Orr and Irving (2005). [Reproduced with permission from Orr and Irving, 2005 (Copyright 2005, Genetics Society of America).]

within a small but critical region of *XR* will let us determine whether the genes that cause hybrid sterility can be separated meiotically from those that cause hybrid segregation distortion is warranted (N. Phadnis, personal communication).

It is clear, then, that hybrid males between these young subspecies suffer both sterility and segregation distortion, and that these two forms of hybrid dysfunction have similar genetic bases. We cannot, therefore, reject the possibility that at least some of the same genes cause both phenomena, as first suggested by Frank (1991) and Hurst and Pomiankowski (1991). More generally, we cannot reject the possibility that arms races between selfish genetic factors like those that cause meiotic drive contribute to the evolution of postzygotic reproductive isolation.

CONCLUSIONS

Molecular evolutionary analyses of speciation genes show that these loci are rapidly evolving, and that this evolution is often driven by positive Darwinian selection. Although the sample of genes characterized thus far by various laboratories remains small, and concentrated in the genus *Drosophila*, I suspect that these patterns may prove general, although likely not universal. Our recent work, along with that of several other groups, also suggests that the selection underlying the evolution of speciation genes may sometimes assume a surprising form, response to intragenomic conflicts, perhaps involving meiotic drive.

In summary, it would appear that two of Mayr's three seminal contributions to the study of speciation were correct, or, at the least, extremely productive. First, the entire research program of the genetics of speciation over the last half-century arose out of Mayr's Biological Species Concept. Whatever one's views on the philosophical strengths or weaknesses of this concept, it has, as a practical matter, given rise to an extraordinarily rich research program, one that has led to a number of substantive discoveries. In a phrase, the modern genetics of speciation is a genetics of reproductive isolation. Second, there can be little doubt that this reproductive isolation typically, if not always, evolves in allopatry (Coyne and Orr, 2004; Mayr, 1963). Finally, however, recent work provides no evidence for a crucial role of genetic drift in speciation. Instead, both ecological studies (Schluter, 2000) and the genetical studies reviewed here point to an important role for positive Darwinian selection in the evolution of reproductive isolation. Although the above data do not exclude all varieties of drift-based models of speciation (founder-effect theories do, after all, feature a role for positive selection), such models seem at present unparsimonious (see also Coyne, 1994; Coyne and Orr, 2004).

It remains a testament to Mayr's vast influence, however, that all of

the questions asked above, whatever their answers, were first raised or emphasized by Mayr nearly half a century ago. It seems doubtful that a modern genetics of speciation in anything like its present form would have arisen without his fundamental contributions.

ACKNOWLEDGMENTS

I thank F. Ayala, W. Fitch, and J. Hey for organizing the symposium in which this work was presented. I also thank J. Hey and an anonymous reviewer for helpful comments on an earlier draft of this paper. This work was supported by National Institutes of Health Grant GM519k32.

REFERENCES

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., *et al.* (2000) The genome sequence of *Drosophila melanogaster. Science* **287,** 2185–2195.

Barbash, D. A., Roote, J. & Ashburner, M. (2000) The *Drosophila melanogaster Hybrid male rescue* gene causes inviability in male and female species hybrids. *Genetics* **154,** 1747–1771.

Barbash, D. A., Siino, D. F., Tarone, A. M. & Roote, J. (2003) A rapidly evolving MYB-related protein causes species isolation in *Drosophila. Proc. Natl. Acad. Sci. USA* **100,** 5302–5307.

Barbash, D. A., Awadalla, P. & Tarone, A. M. (2004) Functional divergence caused by ancient positive selection of a *Drosophila* hybrid incompatibility locus. *PLoS* **2,** 839–848.

Cazemajor, M., Landre, C. & Montchamp-Moreau, C. (1997) *Genetics* **147,** 635–642.

Coyne, J. A. (1994) Ernst Mayr on the origin of species. *Evolution (Lawrence, Kans.)* **48,** 19–30.

Coyne, J. A. & Charlesworth, B. (1986) Location of an X-linked factor causing male sterility in hybrids of *Drosophila simulans* and *D. mauritiana*. *Heredity* **57,** 243–246.

Coyne, J. A. & Orr, H. A. (1989) Patterns of speciation in *Drosophila. Evolution (Lawrence, Kans.)* **43,** 362–381.

Coyne, J. A. & Orr, H. A. (1993) Further evidence against meiotic-drive models of hybrid sterility. *Evolution (Lawrence, Kans.)* **47,** 685–687.

Coyne, J. A. & Orr, H. A. (1997) "Patterns of speciation in *Drosophila*" revisited. *Evolution (Lawrence, Kans.)* **51,** 295–303.

Coyne, J. A. & Orr, H. A. (2004) *Speciation* (Sinauer, Sunderland, MA).

Engels, W. R. & Preston, C. R. (1979) Hybrid dysgenesis in *Drosophila melanogaster*: The biology of male and female sterility. *Genetics* **92,** 161–174.

Frank, S. H. (1991) Divergence of meiotic drive-suppressors as an explanation for sex-biased hybrid sterility and inviability. *Evolution (Lawrence, Kans.)* **45,** 262–267.

Henikoff, S. & Malik, H. S. (2002) Selfish drivers. *Science* **417,** 227.

Henikoff, S., Ahmad, K. & Malik, H. S. (2001) The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science* **293,** 1098–1102.

Hey, J. & Kliman, R. M. (1993) Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. *Mol. Biol. Evol.* **10,** 804–822.

Hurst, L. D. & Pomiankowski, A. (1991) Causes of sex ratio bias may account for unisexual sterility in hybrids: A new explanation of Haldane's rule and related phenomena. *Genetics* **128,** 841–858.

Hutter, P. & Ashburner, M. (1987) Genetic rescue of inviable hybrids between *Drosophila melanogaster* and its sibling species. *Nature* **327,** 331–333.

Hutter, P., Roote, J. & Ashburner, M. (1990) A genetic basis for the inviability of hybrids between sibling species of *Drosophila. Genetics* **124,** 909–920.

Johnson, N. A. & Wu, C.-I. (1992) An empirical test of the meiotic drive models of hybrid sterility: Sex ratio data from hybrids between *Drosophila simulans* and *Drosophila sechellia. Genetics* **130,** 507–511.

Kidwell, M. G. (1983) Intraspecific hybrid sterility. In *The Genetics and Biology of Drosophila*, eds. Ashburner, M., Carson, H. L., & Thompson, J. J. N. (Academic, London), Vol. 3c, pp. 125–154.

Laurie, C. C. (1997) The weaker sex is heterogametic: 75 years of Haldane's rule. *Genetics* **147,** 937–951.

Li, W.-H. (1997) *Molecular Evolution* (Sinauer, Sunderland, MA).

Malitschek, B., Fornzler, D. & Schartl, M. (1995) Melanoma formation in *Xiphophorus*: a model system for the role of receptor tyrosine kinases in tumorigenesis. *BioEssays* **17,** 1017–1023.

Mayr, E. (1942) *Systematics and the Origin of Species* (Columbia Univ. Press, New York).

Mayr, E. (1954) Change of genetic environment and evolution. In *Evolution as a Process*, eds. Huxley, J., Hardy, A. C. & Ford, E. B. (Allen & Unwin, London), pp. 157–180.

Mayr, E. (1963) *Animal Species and Evolution* (Belknap Press of Harvard Univ. Press, Cambridge, MA).

Mercot, H., Atlan, A., Jacques, M. & Montchamp-Moreau, C. (1995) Sex-ratio distortion in *Drosophila simulans*: Co-occurrence of a meiotic drive and a suppressor of drive. *J. Evol. Biol.* **8,** 283–300.

Metz, E. C. & Palumbi, S. R. (1996) Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol. Biol. Evol.* **13,** 397–406.

Montchamp-Moreau, C. & Joly, D. (1997) The *sex-ratio* trait in *Drosophila simulans*: Genetic analysis of distortion and suppression. *Heredity* **79,** 24–30.

Muller, H. J. & Pontecorvo, G. (1942) Recessive genes causing interspecific sterility and other disharmonies between *Drosophila melanogaster* and *simulans. Genetics* **27,** 157.

Noor, M. (1999) Reinforcement and other consequences of sympatry. *Heredity* **83,** 503–508.

Orr, H. A. (1989) Complex epistasis and the genetic basis of hybrid sterility in the *Drosophila pseudoobscura* Bogota-USA hybridization. *Evolution (Lawrence, Kans.)* **43,** 180–189.

Orr, H. A. (1992) Mapping and characterization of a "speciation gene" in Drosophila. *Genet. Res.* **59,** 73–80.

Orr, H. A. (1997) Haldane's rule. *Annu. Rev. Ecol. Syst.* **28,** 195–218.

Orr, H. A. & Irving, S. (2000) Genetic analysis of the *Hybrid male rescue* locus of Drosophila. *Genetics* **155,** 225–231.

Orr, H. A. & Irving, S. (2001) Segregation distortion in hybrids between the Bogota and USA subspecies of *Drosophila pseudoobscura. Genetics* **158,** 1089–1100.

Orr, H. A. & Irving, S. (2005) Abnormal spermiogenesis is associated with the X-linked sex-ratio trait in *Drosophila simulans. Genetics* **169,** 671–682.

Palumbi, S. R. (1992) Marine speciation on a small planet. *Trends Ecol. Evol.* **7,** 114–118.

Perez, D. E. & Wu, C.-I. (1995) Further characterization of the *Odysseus* locus of hybrid sterility in *Drosophila*: One gene is not enough. *Genetics* **140,** 201–206.

Perez, D. E., Wu, C.-I., Johnson, N. A. & Wu, M.-L. (1993) Genetics of reproductive isolation in the *Drosophila simulans* clade: DNA-marker assisted mapping and characterization of a hybrid-male sterility gene, *Odysseus* (*Ods*). *Genetics* **134,** 261–275.

Pontecorvo, G. (1943) Hybrid sterility in artificially produced recombinants between *Drosophila melanogaster* and *D. simulans. Proc. R. Soc. Edinburgh Sect. B* **61,** 385–397.

Presgraves, D. C. (2003) A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila. Genetics* **163,** 955–972.

Presgraves, D. C., Balagopalan, L., Abmayr, S. M. & Orr, H. A. (2003) Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila. Nature* **423,** 715–719.

Provine, W. B. (1989) Founder effects and genetic revolutions in microevolution and speciation:a historical perspective. In *Genetics, Speciation, and the Founder Principle*, eds. Giddings, L. V., Kaneshiro, K. Y. & Anderson, W. W. (Oxford Univ. Press, New York), pp. 43–76.

Rose, M. & Doolittle, W. F. (1983) Molecular biological mechanisms of speciation. *Science* **220,** 157–162.

Schartl, A., Dimitrijevic, N. & Schartl, M. (1994) Evolutionary origin and molecular biology of the melanoma-inducing oncogene of Xiphophorus. *Pigm. Cell. Res.* **7,** 428–432.

Schartl, M., Hornung, U., Gutbrod, H., Volff, J.-N. & Wittbrodt, J. (1999) Melanoma loss-of-function mutants in *Xiphophorus* caused by *Xmr*k-oncogene deletion and gene disruption by a transposable element. *Genetics* **153,** 1385–1394.

Schluter, D. (2000) *The Ecology of Adaptive Radiation* (Oxford Univ. Press, Oxford).

Servedio, M. R. & Noor, M. A. F. (2003) The role of reinforcement in speciation: Theory and data. *Annu. Rev. Ecol. Syst.* **34,** 339–364.

Sun, S., Ting, C.-T. & Wu, C.-I. (2004) The normal function of a speciation gene, Odysseus, and its hybrid sterility effect. *Science* **305,** 81–83.

Swanson, W. J. & Vacquier, V. (2002) Rapid evolution of reproductive proteins. *Nat. Rev. Genet.* **3,** 137–144.

Tao, Y. & Hartl, D. L. (2003) Genetic dissection of hybrid incompatibilities between Drosophila simulans and Drosophila mauritiana. III. Heterogeneous accumulation of hybrid incompatibilities, degree of dominance and implications for Haldane's rule. *Evolution (Lawrence, Kans.)* **57,** 2580–2598.

Tao, Y., Hartl, D. L. & Laurie, C. C. (2001) Sex-ratio distortion associated with reproductive isolation in *Drosophila. Proc. Natl. Acad. Sci. USA* **98,** 13183–13188.

Ting, C.-T., Tsaur, S.-C., Wu, M.-L. & Wu, C.-I. (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282,** 1501–1504.

Vacquier, V. D. (1998) Evolution of gamete recognition proteins. *Science* **281,** 1995–1998.

Wittbrodt, J., Adam, D., Malitschek, B., Maueler, W., Raulf, F., Telling, A., Robertson, S. M. & Schartl, M. (1989) Novel putative receptor tyrosine kinase encoded by the melanoma-inducing *Tu* locus in *Xiphophorus. Nature* **341,** 415–421.

Wu, C.-I. & Ting, C.-T. (2004) Genes and speciation. *Nat. Rev. Genet.* **5,** 114–122.

3
Inter-Locus Antagonistic Coevolution as an Engine of Speciation: Assessment with Hemiclonal Analysis

WILLIAM R. RICE,* JODELL E. LINDER,*
URBAN FRIBERG,*† TIMOTHY A. LEW,*
EDWARD H. MORROW,*
AND ANDREW D. STEWART*

One of Ernst Mayr's legacies is the consensus that the allopatry model is the predominant mode of speciation in most sexually reproducing lineages. In this model, reproductive isolation develops as a pleiotropic byproduct of the genetic divergence that develops among physically isolated populations. Presently, there is no consensus concerning which, if any, evolutionary process is primarily responsible for driving the specific genetic divergence that leads to reproductive isolation. Here, we focus on the hypothesis that inter-locus antagonistic coevolution drives rapid genetic divergence among allopatric populations and thereby acts as an important "engine" of speciation. We assert that only data from studies of experimental evolution, rather than descriptive patterns of molecular evolution, can provide definitive evidence for this hypothesis. We describe and use an experimental approach, called hemiclonal analysis, that can be used in the laboratory model system to simultaneously screen nearly the entire genome for both standing genetic variation within a population and the net-selection gradient acting on the variation. Hemiclonal analysis has four stages: (*i*) creation of a laboratory "island population"; (*ii*) cytogenetic cloning of nearly genomewide haplotypes to construct hemiclones; (*iii*) measurement of additive genetic

*Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, CA 93106; and †Department of Ecology and Environmental Science, Umeå University, 901 87 Umeå, Sweden.

variation among hemiclones; and (*iv*) measurement of the selection gradient acting on phenotypic variation among hemiclones. We apply hemiclonal analysis to test the hypothesis that there is ongoing antagonistic coevolution between the sexes in the *D. melanogaster* laboratory model system and then discuss the relevance of this analysis to natural systems.

There is widespread agreement among evolutionary biologists that the allopatry model is responsible for generating much of the species diversity presently found within sexually reproducing lineages (Coyne and Orr, 2004; Dobzhansky, 1937; Futuyma, 1986; Gavrilets, 2004; Mayr, 1942; Muller, 1942; White, 1978). Although substantial empirical evidence supports the operational steps of the basic allopatry model (Rice and Hostert, 1993), there is no general consensus regarding the relative importance of alternative evolutionary processes that drive the specific genetic divergence that leads to reproductive isolation. Because allopatry necessitates that populations be physically separated, there can be no direct selection for reproductive isolation and therefore it must develop as a pleiotropic byproduct of the genetic differences that accrue due to the independent evolution of populations.

At a minimum, recurrent mutation and random genetic drift of neutral variation lead to genetic differentiation among allopatric populations. However, natural selection can greatly accelerate the rate of genetic divergence. Van Valen (1973) used paleontological evidence to conclude that adaptation to the physical environment is asymptotic (declining in rate with time). Rapid adaptation occurs when a population initially experiences a new physical environment, but the rate of evolution slows as the population becomes progressively more adapted to the prevailing physical conditions. The rate of adaptation diminishes with time as the lag-load (the reduction in mean population fitness due to the average trait of a population differing from its optimal value) of the population decreases. In contrast, adaptation to the biotic environment is expected to be nonasymptotic when enemies (for example, predator and prey, host and pathogen, or resource competitors) become locked in a perpetual arms race of adaptation and counteradaptation. In this case of interspecific antagonistic coevolution, the lag-load of a species does not diminish with time because adaptive progress is continually eroded due to counter-evolution by enemy species.

An analogous cycle of antagonistic coevolution can take place between genes that reside within the genome of a single species (inter-locus antagonistic coevolution). In this case, adaptive allelic replacement at one

locus increases the lag-load at a second locus by generating selection for a new optimal allele, and the resulting adaptive allelic replacement at the second locus increases the lag-load at the first locus, thereby stimulating a new round of the antagonistic cycle of adaptation and counteradaptation. The genetic conflict that drives inter-locus antagonistic coevolution is termed "intragenomic" when conflict occurs between genes that reside in a single individual [for example, genetic conflict associated with genomic imprinting (Haig, 2002), meiotic drive (Jaenike, 2001), and/or cytonuclear conflict (Werren, 1997)] and "intergenomic" when gene products from different loci mediate conflicts of interest between different individuals of the same species (Rice and Holland, 1997).

Inter-locus antagonistic coevolution (intragenomic or intergenomic) can potentially drive rapid genetic divergence among allopatric populations because the antagonistic cycle of adaptation and counteradaptation maintains a persistent lag-load at each interacting locus and thereby drives perpetual evolutionary change. In this article, we focus on intergenomic conflict because two of its forms are predicted to contribute to reproductive isolation by causing genetic divergence among allopatric populations (Parker and Partridge, 1998; Rice, 1996, 1998; Rice and Holland, 1997). Intergenomic conflict can occur both within and between the sexes. An example of intrasexual conflict occurs between gene loci that mediate the male "offense" and "defense" phenotypes in the context of male–male competition to fertilize eggs. A male that mates a virgin female is selected for a defense phenotype, that is, to prevent the female from mating with other males and to prevent his stored sperm from being displaced by a secondary male if the female remates. A male encountering a previously mated female is selected for an offense phenotype, that is, to induce the female to mate with him even if she has ample stored sperm from another male, and then to replace the stored sperm from the previous male with his own. To the extent that the male offense and defense phenotypes are controlled, at least in part, by gene products derived from different gene loci, these interacting genes can antagonistically coevolve in a perpetual cycle of adaptation and counteradaptation.

Persistent male courtship, insemination, and seminal fluid proteins are known to be harmful to females in many species because they lower the female's survival, fecundity, and short-term fertility (Arnqvist, 1989; Burpee and Sakaluk, 1993; Chapman *et al.*, 1993, 1995; Cohet and David, 1976; Crudgington and Siva-Jothy, 2000; Dean, 1981; Fowler and Partridge, 1989; Friberg and Arnqvist, 2003; Kasule, 1986; McKinney *et al.*, 1983; Moore *et al.*, 2001; Partridge and Fowler, 1990; Pitnick and Garcia-Gonzalez, 2002; Prout and Clark, 2000). Available evidence indicates that the harm to females is an incidental byproduct of adaptations that increase male fertilization success in promiscuous species (Hosken *et al.*, 2001; Morrow *et al.*, 2003). Genes

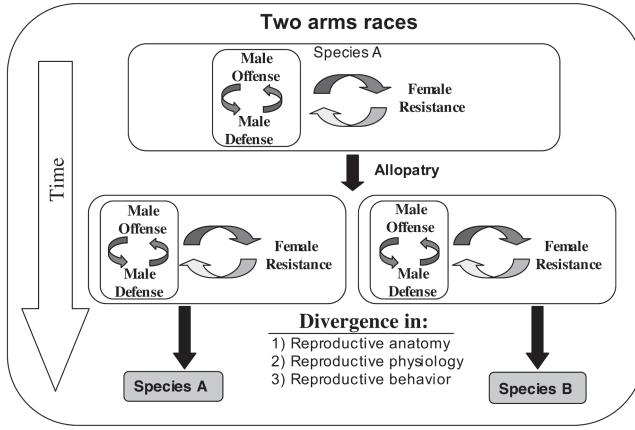


FIGURE 3.1 Inter-locus antagonistic coevolution within and between the sexes can promote the rapid evolution of reproductive traits that are expected to contribute to reproductive isolation (prezygotic mating isolation and postzygotic hybrid infertility) among allopatric populations.

expressed in females are expected to evolve to reduce the harmful effects of interacting with males. If the counteradaptations by female-expressed genes reduce the efficacy of male–male competition to fertilize eggs, then an inter-locus intersexual arms race is expected to ensue between genes that promote female resistance to male-induced harm and the genes that code for male offense and defense (Parker and Partridge, 1998; Rice, 1996, 1998; Rice and Holland, 1997).

The arms races between male offense and defense, and both of these processes with female resistance, would be expected to occur independently in allopatric populations and drive rapid genetic divergence among them (Fig. 3.1). Because, in principle, these arms races can be expected to cause changes in phenotypes influencing reproductive traits (mating behavior and reproductive physiology and anatomy), the genetic divergence produce by inter-locus antagonistic coevolution is expected to cause reproductive isolation by means of pleiotropy in the context of prezygotic reproductive isolation and hybrid infertility, and be an important engine of speciation (Fig. 3.1).

There are two major lines of evidence supporting the hypothesis that inter-locus arms races within and between the sexes drive rapid genetic divergence among allopatric populations: studies of molecular evolution (Swanson and Vacquier, 2002) and studies of experimental evolution (Holland and Rice, 1999; Hosken *et al.*, 2001; Martin and Hosken, 2003; Rice, 1996; Wigby and Chapman, 2004). Molecular studies have demonstrated

that reproductive proteins evolve more rapidly than most other types of proteins (Swanson and Vacquier, 2002). For example, in abalone (genus *Haliotis*) the gene *lysin* codes for a protein that dissolves a tunnel through the glycoprotein coat surrounding the egg, and the gene *verl* codes for the glycoprotein. From the sperm's perspective, a *lysin* gene product that more rapidly dissolves a tunnel through the glycoprotein coat is favored because it helps the sperm win in sperm competition. But, from the egg's perspective, slower penetration of the sperm through the glycoprotein coat is expected to be favored because it provides more time for a block to polyspermy (polyspermy is fatal to the egg) when many sperm compete to fertilize the same egg. This opposing selection on sperm penetration rate sets up a potential arms race between the *verl* and *lysin* genes (Frank, 2000; Rice, 1998; Rice and Holland, 1997; Swanson and Vacquier, 2002; Vacquier *et al.*, 1997). In support of this arms race, studies of molecular evolution have demonstrated that both *verl* and *lysin* are evolving rapidly due to positive Darwinian selection (Swanson and Vacquier, 2002). As the arms race progresses independently among allopatric populations, the capacity for fertilization to occur between sperm and eggs derived from different populations would be expected to diminish due to coevolution between *lysin* and *verl* following different evolutionary trajectories in separated populations.

However, the pattern observed in the studies of molecular evolution has an alternative explanation. The *verl* gene may be antagonistically evolving due to an interspecific arms race with one or more pathogens that gain entry to the egg by transgressing the glycoprotein coat (Rice, 1998; Rice and Holland, 1997; Vacquier *et al.*, 1997). In this case, the evolution at the *lysin* locus does not create a lag-load at the *verl* locus, but instead it evolves to track the evolution in *verl* that occurs in response to an interspecific arms race between host and pathogen. This alternative explanation for the same molecular data illustrates the problem with using descriptive studies of molecular evolution to test the hypothesis that inter-locus arms races are driving genetic divergence among populations. Because the data are correlative and do not directly measure selection, descriptive studies of molecular evolution can provide supporting evidence for inter-locus arms races but cannot provide definitive evidence.

Studies of experimental evolution in the laboratory are capable of measuring simultaneously standing genetic variance, selection on this variation, and response to the selection at a level of detail that cannot be achieved in natural populations. As a consequence, these studies can be used to provide a direct assessment of the potential for inter-locus antagonistic coevolution within and between the sexes. In the following section, we describe an experimental approach (hemiclonal analysis) to screen nearly the complete genome of *Drosophila melanogaster* for both genetic

variation (expressed under outbred conditions) and the net-fitness selection gradient on this variation. Because the flies are assayed in the outbred state and in an environment to which they have adapted at large size for hundreds of generations, we were able to obtain direct experimental evidence for the inter-locus arms race between the sexes that can drive the genetic divergence that leads to reproductive isolation. The technique that we describe can be applied only to laboratory populations, but we assert that laboratory populations can be constructed in such a way that the study of their evolution provides an essential complement to studies of natural populations.

HEMICLONAL ANALYSIS

The ability to predict the direction of evolutionary change requires that one establish that there is (*i*) standing additive genetic variation for a trait, and (*ii*) a non-zero net-fitness selection gradient on the trait in the prescribed direction. Because of the recent, rapid advance of DNA-based molecular tools, our capacity to measure certain types of standing genetic variation in natural populations is no longer a limiting factor, for example, in genes with highly conserved regions that permit PCR primers to be readily constructed. However, measuring heritable standing genetic variation for fitness-related traits, and especially net-fitness itself, in natural populations remains a daunting task, although these traits can be approximated in special circumstances (see, for example, Kruuk *et al.*, 2000). Although it is possible in the laboratory to accurately measure genetic variation and fitness components associated with many traits, the relevance of these measurements to net-fitness and heritability in nature is dubious. As a consequence, although we can presently measure the standing genetic variation for certain types of traits (such as microsatellites) of natural populations with a high degree of accuracy, levels of heritable phenotypic variation, and the selection gradients on this variation, are unknown for most traits (Fig. 3.2). To circumvent this problem we have focused on the analysis of a locally adapted laboratory population of *D. melanogaster*. In this context, by cloning nearly entire genomes and amplifying them to large numbers, we can measure accurately both standing genetic variation for a trait and net-selection acting on this variation. Our purpose in this laboratory analysis is not to duplicate the natural history of wild populations, but to study the evolution of laboratory-adapted populations in their own right to deduce basic principles of evolution. Many laboratories throughout the world study laboratory populations; however, we describe here a particular form of evolutionary investigation that we call "hemiclonal analysis" that specifically applies

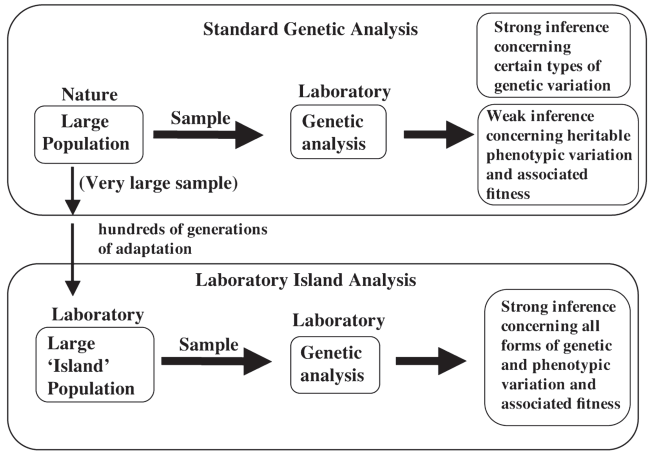


FIGURE 3.2 Genetic analysis of natural and laboratory-adapted populations. Laboratory analysis of samples of organisms taken from wild populations permits strong inference concerning levels of certain types of genetic variation (for example, microsatellites), but only weak inference concerning levels of heritable variation for phenotypic traits, and selection on this variation. When laboratory island populations are analyzed, strong inference is possible for all forms of genetic and phenotypic variation, and its adaptive significance, because the organisms are measured in the same environment to which they have adapted for hundreds of generations.

to the *D. melanogaster* model system. Our approach has four stages that we describe and discuss in the sections below.

STAGE 1: CREATE A LABORATORY ISLAND POPULATION

The study of island populations has played an important role in the study of evolution, beginning with the pioneering studies of Charles Darwin and Alfred Wallace. It is our view that one of the major reasons that island populations have been particularly informative is that they are much simpler then continental populations (in the context of both biotic and abiotic factors) and therefore easier to understand. Laboratory populations represent island populations that conveniently reside within the laboratory where joint measurements of fitness and genetic variation are more feasible. They are far simpler than natural island populations, but this simplicity provides tractability in the context of evolutionary analysis. By comparison, ecologists have gained important insights into the extinction and colonization process by studying very small island populations of vertebrates in nature, despite the fact that these "islands" are

little more than large protruding rocks, 1–16 m on a side (see, for example, Schoener *et al.*, 2004).

The utility of using laboratory populations to study evolution depends upon how much their evolution is regulated by the same principles that control the evolution of natural populations. Most past laboratory studies of *D. melanogaster* have used highly inbred stocks such as Oregon-R, Laussan-S, and Canton-S, or genetic samples that have been recently derived from nature and therefore have not adapted to the laboratory environment. In the latter case, the flies are tested in a novel environment so measures such as heritability and selection on the standing, heritable phenotypic variance are difficult to interpret. Inbred laboratory stocks lack these problems, but they have been bottlenecked many times, their extreme and uncontrolled crowding interferes with many forms of behavioral interactions that have historically been important in the species, and their laboratory culture varies among laboratories and stock centers. As a consequence, these inbred laboratory populations have evolved under crowded, uncontrolled conditions that preclude the importance of the rich repertoire of behavioral interactions within and between the sexes, and they have undoubtedly fixed for large numbers of deleterious alleles. To avoid these undesirable aspects of standard laboratory stocks, our laboratory, like others (e.g., the laboratories of Brian Charlesworth, Linda Partridge, and Michael Rose), has started a new laboratory population. The large outbred laboratory population that we study (LH_M) was founded by Larry Harshman (now at the University of Nebraska) from 400 inseminated females collected in an orchard near Modesto, California in 1991. Since then, it has been maintained at a large effective population size (>1,800 breeding adults). In 1995 the population density of juveniles and adults was reduced so that juvenile density was consistently maintained at between 150 and 200 individuals per vial, and adult density was reduced further to only 16 pairs per vial (placed on its side to allow more horizontal space for the flies to spread out). The low adult density increased the potential for behavioral interactions to contribute the adult fitness of both sexes.

The population of flies pass through three sets of 56 10-dram vials during their 2-week generation cycle. On day 1, the eggs that were laid at the end of the previous generation are randomly reduced to 150–200 per vial to prevent the extreme crowding that occurs in most mass-transfer laboratory cultures. The flies remain in these "juvenile competition" vials for 12 days during which larval competition, pupation, and the early adult stages occur. On day 12, the flies are mixed among vials and, after being randomly culled to 16 pairs per vial, are transferred to new "adult competition" vials where they reside for 48 h. Live yeast (10 mg) is applied to the top of the 10 ml of killed yeast medium in each vial. There is

a steep linear relationship between the amount of live yeast applied and average female fecundity, indicating that live yeast is the major factor limiting female fecundity (Linder and Rice, 2005). In the adult competition vials, females compete intensely for the limited supply of live yeast, and males compete to inseminate females and fertilize their eggs. Eighteen hours before the end of the 2-week generation cycle, the flies are transferred to "oviposition vials" (with no live yeast added), and the eggs laid at this time are used to begin the next generation. As a consequence, egg production in the oviposition vials represents the lifetime offspring production of both sexes. Put another way, the flies are selected to be "big-bang" reproducers, analogous to semelparous salmon. During the 2-week generation cycle, adults live for at most 6 days, and there is virtually no adult mortality during this time; however, larval mortality does occur in the juvenile competition vials at a rate of ≈10% (Chippindale *et al.*, 2001). In the following sections, all measurements of lifetime fitness and phenotypic traits are taken under conditions that closely match those of the routine culture of the LH_M base population.

For the purpose of assaying the flies for lifetime fitness and other traits such as sperm displacement, we also have backcrossed genetic markers [*brown eye* (*bw*), *brown-eye-dominant* (*bwD*), and *nubbin wings* (*nub2*)] and a compound-X [*C(1)DX y f*] into the LH_M base population. Each marker or chromosome has been backcrossed a minimum of 10 times through the LH_M base population. Only the brown-eyed (LH_M-*bw*) and compound-X(LH_M-DX) replicas of the LH_M base population were used in the assays described below.

At the time of this writing, the LH_M population has adapted to the laboratory, at continuous large effective size, for 350 generations. With such a long period to adapt to the laboratory environment, most polygenic quantitative traits should have had sufficient time to at least approach their new optima. The capacity for rapid evolution to a new environment is manifest in newly derived *Drosophila* populations (for example, see Frankham and Loebel, 1992) and is also illustrated by the transplantation experiments of Reznick *et al.* (1997) in which natural populations of guppies were shown to rapidly adapt to new environments under field conditions. Because the LH_M population has such a long history of adaptation to a prescribed laboratory environment, we are able to measure both its genetic and fitness variation in the environment to which it is adapted, and thereby assess antagonistic coevolution within and between the sexes. Because the model organism is *D. melanogaster*, we can apply the powerful set of genetic tools available only in this species to take full advantage of the laboratory island population.

STAGE 2: CYTOGENETIC CLONING AND CONSTRUCTION OF HEMICLONES

There are many quantitative genetic techniques available to test for additive genetic variation for a trait and to measure net-selection on this variation. One of the most powerful techniques available with the *D. melanogaster* model system is the isolation and amplification of individual chromosomes through the use of balancers (which suppress recombination between homologous chromosomes). With this technique, many different chromosomes can be randomly sampled, and genetic variation can then be measured among different isochromosomal lines. Because the individual chromosomes can be amplified to many copies, high statistical power can be achieved.

A limitation with the use of balancers to construct isochromosome lines is that balancers are effective in suppressing recombination only when used on single chromosomes. Therefore the entire genome cannot be screened simultaneously with this technique. To eliminate this problem, we (in collaboration with my former postdoctoral associate, A. Chippindale, and former graduate student J. Gibson) have devised a method that relies on the lack of homologous recombination in male *Drosophila*, rather than balancers, to clonally amplify chromosomes, as originally described in Rice (1996). The advantage of this technique is that a haplotype spanning nearly the entire genome (99.5%, including the X and both major autosomes, but excluding the dot fourth chromosome) can be clonally amplified simultaneously. As a consequence, nearly the entire genome of *D. melanogaster* can be screened for genetic variation (and selection, see below) simultaneously.

The protocol for cytogenetic cloning has been described in detail (Chippindale *et al.*, 2001), and it is schematically outlined in Fig. 3.3. A random male is drawn from the base population and mated to a "clone-generator" female. These females carry a random Y chromosome from the LH_M base population, an attached X (both X chromosomes in females cosegregate as a single linkage unit), and a translocation of the two major autosomes (which, in heterozygotes, causes the two autosomal chromosomes to cosegregate as a single linkage unit, among the living offspring). A single son from this cross (carrying a random genomic haplotype from the base population, which includes the X and the two major autosomes) is randomly selected and mated to many clone-generator females. The sons from these crosses that are retained (half of all sons are retained; the discarded half are homozygous for the translocation and are identifiable by the recessive markers on the translocation) carry the same genomic haplotype as their father and are next crossed to many clone-generator

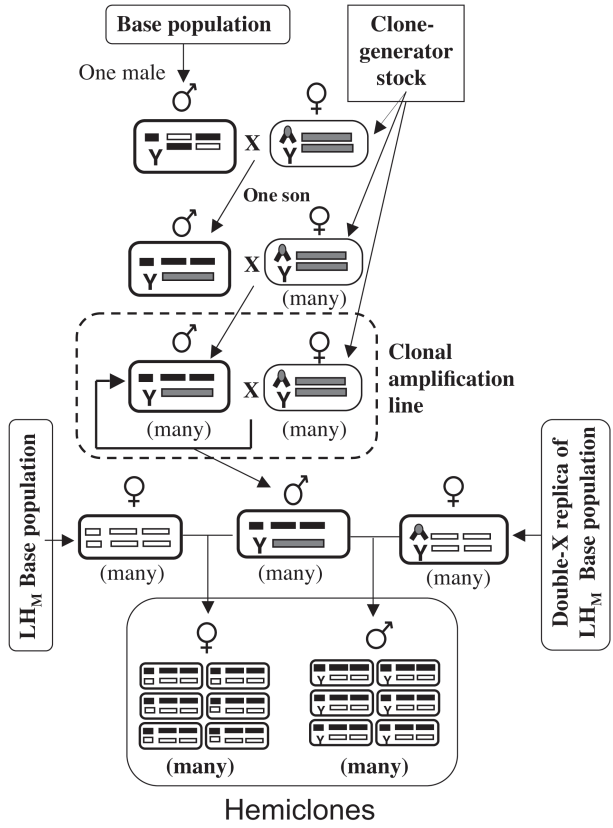


FIGURE 3.3 A schematic of the cytogenetic cloning protocol and the construction of male and female hemiclones. Ellipses depict individuals, and rectangles within ellipses depict chromosomes (X/Y at left position, autosome-2 at middle position, and autosome-3 at right position). The compound X chromosome (chevron) is *C(1)DX y f,* and the autosomal translocation (rectangle spanning the two autosomal chromosome positions) is *T(2;3) rdgC st pp bwD.*

females to produce a clonal amplification line that perpetuates the genomic haplotype (Fig. 3.3).

Males from a clonal amplification line are next crossed to one of two types of females: wild-type females from the LH_M base population or females from an attached-X replica of the LH_M base population that is continuously backcrossed to the LH_M population (Fig. 3.3). Half of the males or females from these crosses are completely wild type (the other half express a dominant genetic marker, *bwD*, and are discarded) and constitute a hemiclone. Members of a hemiclone share in common one

nearly complete genomic haplotype, each expressed in a different random genetic background. A hemiclone is equivalent to the offspring that would be produced by randomly picking a group of eggs from the base population and then fertilizing each egg with a cloned copy of the same sperm.

STAGE 3: MEASURING GENETIC VARIATION

To measure genetic variation for an arbitrary trait in the base population, multiple genomic haplotypes are independently sampled, cytogenetically cloned, and used to construct clonal amplification lines (Fig. 3.3). Next, hemiclones are constructed independently two or more times from each clonal amplification line, and the phenotypic value of each individual in each hemiclone is measured. Finally, random-effects analysis of variance is used to partition phenotypic variation among and between hemiclones to estimate additive genetic variance among hemiclones (Chippindale *et al.*, 2001). Individuals within a hemiclone share half of their genetic variation in common, so that two times the additive genetic variation among hemiclones divided by the total phenotypic variation approximates the heritability of the trait in the base population.

The additive genetic variation among hemiclones contains no nonadditive dominance variation, nor epistatic variation between alleles that reside in the genomic haplotype of a hemiclone and those in its genetic background. It does, however, potentially contain nonadditive epistatic variation between nonallelic genes that reside in the same genomic haplotype. Epistasis can occur between nonallelic genes that reside in genomic haplotypes inherited from (*i*) the father, (*ii*) the mother, or (*iii*) a mixture of these two. Only epistasis between genes that both reside in the paternal haplotype are included in the measure of additive genetic variation among hemiclones (a quarter of the four possible pair-wise types). The inclusion of some epistatic variation in the estimate of additive genetic variation is not unique to hemiclonal analysis. In fact, because of the lack of recombination in male *Drosophila*, it is a confound that is shared in common with most forms of quantitative genetic analysis with *Drosophila*. For example, a paternal half-sib design to estimate additive genetic variance includes epistatic variance among alleles that reside on the same chromosome because lack of recombination in male *Drosophila* keeps these alleles together during meiosis. Similarly, the well-known North Carolina II breeding design (Comstock and Robinson, 1952) also confounds additive and epistatic variation, when applied to *D. melanogaster*, because the protocol uses balancer chromosomes that cause whole chromosomes to segregate like single giant supergenes (e.g., see Hughes, 1997).

In effect the hemiclonal analysis technique of estimating additive ge-

netic variation among genomic haplotypes treats the genome as if it were a single highly pleiotropic locus in males (but not in females). In the more conventional procedures of paternal half-sib analysis of variance and offspring-sire regression in *D. melanogaster*, the male genome segregates as if it were three gene loci (corresponding to the X and the two major autosomes, and ignoring the 0.5% of the genome found on the dot fourth chromosome and the small number if genes residing on the Y sex chromosome). Recombination occurs in females in paternal half-sib analysis of variance and offspring-sire regression designs (as it does in hemiclonal analysis), but these recombined chromosomes do not contribute to the covariance used to estimate heritability. The lack of recombination in males causes each male chromosome to be transmitted intact from father to offspring; hence, each pair of homologous chromosomes in males behaves like a pair of alleles residing at a single highly pleiotropic locus. The only way that we see to disentangle epistatic variation from estimates of additive genetic variance in *D. melanogaster* that are free from confounding maternal effects would be to carry out analysis of more distant paternal relatives. Because our measures of additive genetic variation among hemiclones include a limited amount (25%) of the potential epistatic variation, they represent an upper bound for the level of additive genetic variation among diploid individuals.

STAGE 4: MEASURING THE NET SELECTION GRADIENT ON STANDING PHENOTYPIC VARIATION

To measure the net selection gradient on phenotypic variation, we needed to first measure the average net-fitness associated with each hemiclone. To estimate this value, a prescribed number of eggs from a hemiclone are placed in a vial with genetically marked (LH_M-*bw*) competitors (we have specifically used a 1:2 ratio of hemiclonal eggs to competitor eggs in previous assays; see Chippindale *et al.*, 2001). Next, the vials of eggs are put through the same culturing protocol as the LH_M base population, and then the numbers of offspring produced by the hemiclonal individuals are measured. In this way, each fitness component is weighted by its relative contribution to total fitness, and a measure of net-fitness is obtained. This net-fitness measure of each hemiclone is averaged across many different genetic backgrounds and thus appropriately weighted for the influence of different genetic backgrounds. Lastly, once the average net-fitness of a random set of hemiclones is measured, a bivariate plot of average net-fitness vs. average trait value is constructed, and the slope (selection gradient) is estimated with reduced major axis regression.

APPLICATION OF HEMICLONAL ANALYSIS TO INTER-LOCUS ANTAGONISTIC COEVOLUTION

To use hemiclonal analysis to assess the potential for inter-locus sexual conflict to drive genetic divergence among allopatric populations, we review results from two recent studies in our laboratory. In the first study, a random set of 35 genomic haplotypes was sampled from the LH_M base population and used to construct five replicated sets of the 35 female hemiclones (Linder and Rice, 2005). The number of females in each replicate was 20 per hemiclone. The females were collected as virgins and then mass mated by brief (90 min) exposure to a 3:1 ratio of males to females from the LM_M base population. Once mated, the females were randomly divided into two experimental treatments during the 2-day adult competition phase of their 2-week generation cycle. The treatments were (*i*) a "male-protected" environment in which the 10 hemiclonal females competed with 6 unrelated females for the resource that limited their lifetime fecundity (10 mg of live yeast, applied to the surface of the killed-yeast medium) in the absence of persistent courtship from males (no males present), and (*ii*) a "male-exposed" environment that was identical to the former treatment except that females competed in the presence of males at a 1:1: sex ratio of males to female, i.e., they competed under the social environment experienced during the normal propagation of the LH_M base population. The environmental conditions under which the flies were assayed (for example, timing of events, food levels, and densities and ages of flies) closely matched those to which the LH_M base population had adapted for >300 generations. Finally, the lifetime fecundity of the females was compared between the two treatments by measuring egg production during the last 18 h of their 2-week generation cycle (which is equivalent to egg production in the oviposition vials during the normal propagation of the LH_M base population). Any reduction in fecundity in the male-exposed compared with the male-protected treatments estimated the total cost of interacting with males, that is, the reduction in fecundity owing to resources spent by females when they decamped, kicked, and otherwise responded to persistent interactions with males. Control experiments demonstrated that males did not compete with females for their limiting resource (live yeast) and that differences in fly density did not contribute to the difference in fecundity between treatments (Linder and Rice, 2005).

All 35 hemiclones experienced a decline in lifetime fecundity due to their interactions with males (Fig. 3.4, modified from Linder and Rice, 2005). On average, lifetime fecundity was reduced by 15.4% in the male-exposed treatment relative to the male-protected treatment. However, some hemiclones were harmed more than others (Fig. 3.4), and this het-

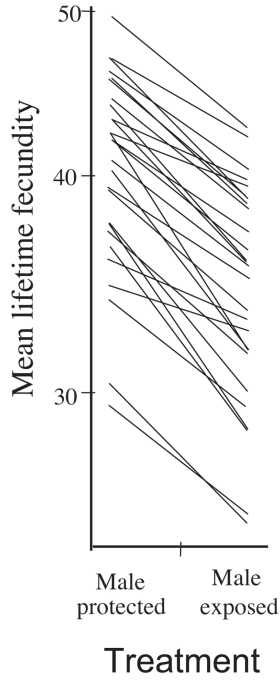


FIGURE 3.4 Aninteraction plot of the average lifetime fecundity of 35 hemiclones when reared in adult competition vials with and without males (modified from Linder and Rice, 2005).

erogeneity in the degree to which lifetime fecundity was reduced by interactions with males measured variation for female resistance to male-induced harm. Analysis of variance was used to test for and estimate heritable variation among hemiclones for the degree of resistance to males, and highly significant genetic variation was observed ($P < 0.0001$; Linder and Rice, 2005). Heritable variation among hemiclones was estimated to contribute only 2.4% of the total phenotypic variation among individuals, indicating that standing heritable variation was low, as would be expected for a polygenic trait subject to strong directional selection. However, 17% of the total genetic variation among hemiclones for lifetime fecundity was due to variation in female resistance to male-induced harm, indicating that this trait contributed substantially to total genetic fitness variation among females.

Because males harm females through their seminal fluid (Chapman *et al.*, 1995), the rate of secondary mating with different males (i.e., remating rate) was a candidate phenotype contributing to female resistance. We

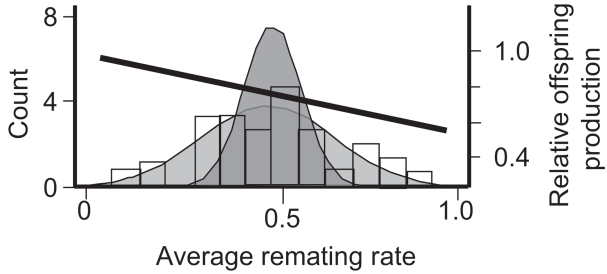


FIGURE 3.5 The distribution of average remating rate of the 35 hemiclones when expressed in females, and the selection gradient on this variation. The lightly stippled curve is a normal distribution fit to the data, and it depicts phenotypic variation among hemiclones. The darker curve depicts genetic variation: it is a normal distribution centered at the sample mean but with variance set equal to the estimated genetic variation among hemiclones (modified from Linder and Rice, 2005).

measured this trait as the percentage of females that remated at least once during their 2 days in the adult competition vials. We found significant heritable variation among hemiclones for remating rate (heritability = 18.2%, additive coefficient of variation = 14.4%, $P < 0.0001$, Fig. 3.5, modified from Linder and Rice, 2005). Female hemiclones that remated at higher average rate experienced a higher proportional reduction in their lifetime fecundity ($P < 0.0001$), indicating that reluctance to remate contributed substantially to the female resistance phenotype (Linder and Rice, 2005). We also found a significant negative correlation between our measure of female remating rate and female lifetime fecundity ($P = 0.0269$; Fig. 3.5), indicating that there was a negative selection gradient on female propensity to remate. Because (*i*) there was virtually no adult mortality during the 2-week generation cycle of the LH_M base population (Chippindale *et al.*, 2001), and (*ii*) there is no measurable correlation between adult and juvenile fitness (Chippindale *et al.*, 2001), this negative correlation indicated a negative net-selection gradient on female remating rate. At this time, we have not completed an independent assay of lifetime fitness for the 35 hemiclones, so we were unable to directly test for a net-fitness selection gradient on female remating rate. However, an independent assay of a separate, smaller set of 16 hemiclones for which we have completed both a total fitness assay (Rice and Chippindale, 2001) and an assay for remating rate (T.A.L., E.H.M., and W.R.R., unpublished results) corroborated a negative net selection gradient on remating rate.

To test for inter-locus antagonistic coevolution between the sexes, the same 35 hemiclones were expressed in males and assayed for remating

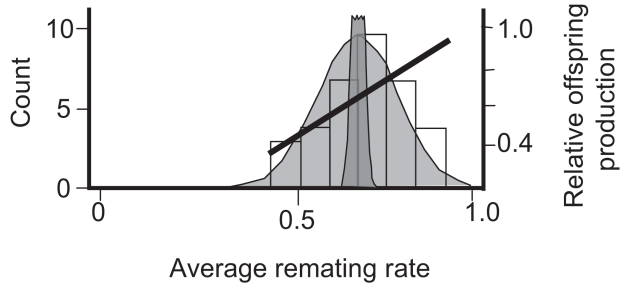


FIGURE 3.6 The distribution of average remating rate of the 35 hemiclones when expressed in males, and the selection gradient on this variation. Normal curves are defined as in Fig. 3.5.

rate. The logic underlying this second assay was to determine whether there was additive genetic variation for remating rate in males, whether this variation was at least partially nonoverlapping with that controlling remating in females, and whether the net-selection gradient on remating rate in males was positive.

The protocol for measuring remating rate in males followed that of the male-exposed treatment of the female resistance assay except that the hemiclones were expressed as males and the females expressed random genotypes drawn from the base population (LH_M-*bw*). We found significant additive variation among male hemiclones for remating rate (proportion of nonvirgin females that remated at least once with the hemiclonal males during their 2 days in the adult competition vials (heritability = 1.4%, additive coefficient of variation = 8.7%, $P < 0.05$; Fig. 3.6 and Friberg *et al.*, 2005). We also found a significant positive correlation between male remating rate and male adult fitness measured during the assay ($P = 0.0081$, Fig. 3.6). Because we found (*i*) no negative correlations between male remating rate and any other adult male fitness components, and (*ii*) past work in our laboratory found no measurable correlation between adult and juvenile fitness (Chippindale *et al.*, 2001) the positive correlation between male adult offspring production and male remating rate indicated a positive net-selection gradient on this character in males. Again, because we have not yet completed an independent assay of net fitness variation among the 35 male hemiclones, we were unable to directly test for a positive net-selection gradient on male remating rate. However, a smaller assay of 17 different hemiclones was available from a related study that tested for both remate rate (Friberg *et al.*, 2005) and lifetime fitness (Rice and Chippindale, 2001). Here, a positive correlation was found between male remating rate and male lifetime

fitness ($P = 0.029$), corroborating the significant positive net-selection gradient on this trait deduced from the measures of male adult lifetime offspring production from the larger sample of 35 hemiclones.

In sum, we found significant additive genetic variation among hemiclones for remating rate in both males and females, but the net-selection gradient on this trait was positive in males and negative in females. To look for independent genetic variation for remating rate in males and females, we constructed a bivariate plot of remating rate of the same hemiclones when expressed in males vs. females. No significant correlation was found ($r = –0.175$, $P = 0.316$), indicating that remating rate in the two sexes is controlled by different genetic variation. Because there is independent genetic variation for remating rate in the two sexes, and because it is selected in opposite directions in each sex, we conclude that this trait presently is evolving in opposite directions in the two sexes and therefore that sexually antagonistic coevolution for mating rate is currently in evidence in this laboratory island population.

INTERPRETATION OF RESULTS FROM HEMICLONAL ANALYSIS

In the above section, we used hemiclonal analysis to provide evidence that (*i*) females have genetic variation for resistance to male-induced harm, (*ii*) resistance contributes substantially to total genetic variation for net fitness, (*iii*) propensity to remating strongly influences the degree of female resistance, and (*iv*) there is unique genetic variation for remating rate in males and females that is selected and evolving in opposite directions in the two sexes. These data provide support for the hypothesis that perpetual inter-locus, intersexual arms races contribute to rapid genetic divergence among allopatric populations, and owing to the phenotypes that coevolve (reproductive behavior, physiology, and anatomy) are likely to be contributing to the specific genetic divergence that leads to reproductive isolation and speciation.

The data that we described, however, came from a laboratory island population rather than directly from nature. Some might argue that such populations are too artificial and hence tell us nothing about evolution in nature. We disagree. We cannot statistically extrapolate from our laboratory island population to natural populations of *D. melanogaster* because our laboratory population is not a random sample from the natural environment. We can, however, use laboratory island populations to make inferences about the fundamental principles of evolution and then use logic to extrapolate to the process of evolution in nature. Just as Darwin (1859) used his study of island tortoise populations to deduce general evolutionary principles (rather than extrapolate to specific continental

populations of tortoises), we used a laboratory island population to assess the evolutionary principles that underlie inter-locus antagonistic coevolution between the sexes. Our finding, that after hundreds of generations of coevolution we can detect an ongoing arms race between the sexes, supports the conclusion that perpetual arms races occur in nature and contribute substantially to the genetic divergence that leads to reproductive isolation and speciation.

A study such as ours could not feasibly be carried out in nature and therefore is possible only in the context of laboratory island populations. For example, the fact that we were able to detect heritabilities among hemiclones of only 2.4% for female resistance illustrates the substantial statistical power of this approach. The genetic measurements that we obtained for standing genetic variance and heritability took advantage of a broad array of genetic tools that are available only in laboratory populations of *D. melanogaster*. So, in general, we see two options: (*i*) study only natural populations and wait for technology to advance to the point that experiments such as ours are possible *in situ*, or (*ii*) study laboratory island populations where these experiments are possible today. We see a clear advantage to the second option.

We believe that the process of biotic evolution has basic underlying principles that apply to manmade microcosms just as they do to natural ecosystems. If we want to estimate the current or historical trajectory of a natural population, then we need to study that population *in situ*. But, if we want to understand the evolutionary principles that underlie evolution in nature, rather than specific evolutionary histories, then we can study them just as effectively, and in general more so because of reduced technical constraint, in the context of laboratory island populations. There is the danger that newly constructed laboratory populations will display misleading transients, and, as a result, caution is needed when interpreting results from laboratory populations that have not coevolved over a protracted number of generations (Sgrò and Partridge, 2000). Nonetheless, laboratory island populations make possible evolutionary analysis that cannot be achieved in nature and thereby provide an essential complement to direct studies of populations in nature.

CONCLUSIONS

The allopatric model of speciation, as originally articulated by Dobzhansy (1937) and Mayr (1942), requires genetic divergence among physically isolated populations. Although sequence data are available for only a small number of genes that cause reproductive isolation, the available data indicate that these genes evolve rapidly under positive Darwinian selection [see article by H. A. Orr, "The Genetic Basis of Reproductive

Isolation: Insights from *Drosophila"* (Chapter 2)]. A fundamental question that remains is the identification of the selective process that drives the rapid divergence of the genes that lead to speciation. In this article, we show that experimental evolution, and more specifically hemiclonal analysis, provides support for the hypothesis that inter-locus antagonistic coevolution promotes rapid genetic divergence among allopatric populations.

ACKNOWLEDGMENT

This work was supported by National Science Foundation Grants DEB99-96164, DEB-0128780, and DEB-0410112 (to W.R.R.).

REFERENCES

Arnqvist, G. (1989) Multiple mating in a water strider—Mutual benefits or intersexual conflict. *Anim. Behav.* **38,** 749–756.

Burpee, D. M. & Sakaluk, S. K. (1993) Repeated matings offset costs of reproduction in female crickets. *Evol. Ecol.* **7,** 240–250.

Chapman, T., Hutchings, J. & Partridge, L. (1993) No reduction in the cost of mating for *Drosophila-melanogaster* females mating with spermless males. *Proc. R. Soc. London Ser. B* **253,** 211–217.

Chapman, T., Liddle, L. F., Kalb, J. M., Wolfner, M. F. & Partridge L. (1995) Cost of mating in *Drosophila-melanogaster* females is mediated by male accessory-gland products. *Nature* **373,** 241–244.

Chippindale, A. K., Gibson, J. R. & Rice, W. R. (2001) Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in *Drosophila. Proc. Natl. Acad. Sci. USA* **98,** 1671–1675.

Cohet, Y. A. & David, J. R. (1976) Deleterious effects of copulation in *Drosophila* females as a function of growth temperature of both sexes. *Experientia* **32,** 696–697.

Comstock, R. E. & Robinson, H. F. (1952) In *Heterosis*, ed. Gowen, J. W. (Iowa State College Press, Ames, IA), pp. 494–516.

Coyne, J. A. & Orr, H. A. (2004) *Speciation (*Sinauer, Sunderland, MA).

Crudgington, H. S. & Siva-Jothy, M. T. (2000) Genital damage, kicking and early death—The battle of the sexes takes a sinister turn in the bean weevil. *Nature* **407,** 855–856.

Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (John Murray, London).

Dean, J. M. (1981) The relationship between lifespan and reproduction in the grasshopper *Melanoplus. Oecologia* **49,** 385–388.

Dobzhansky, T. (1937) *Genetics and the Origin of Species* (Columbia Univ. Press, New York).

Fowler, K. & Partridge, L. (1989) A cost of mating in female fruit-flies. *Nature* **338,** 760–761.

Frank, S. A. (2000) Sperm competition and female avoidance of polyspermy mediated by sperm-egg biochemistry. *Evol. Ecol. Res.* **2,** 613–625.

Frankham, R. & Loebel, D. A. (1992) Modeling problems in conservation genetics using captive drosophila populations—Rapid genetic adaptation to captivity. *Zool. Biol.* **11,** 333–342.

Friberg, U. & Arnqvist, G. (2003) Fitness effects of female mate choice: preferred males are detrimental for *Drosophila melanogaster* females. *J. Evol. Biol.* **16,** 797–811.

Friberg, U., Lew, T. A., Byrne, P. G. & Rice, W. R. (2005) Assessing the potential for an ongoing arms race within and between the sexes: Selection and heritable variation. *Evolution*, in press.

Futuyma, D. J. (1986) *Evolutionary Biology* (Sinauer, Sunderland, MA).

Gavrilets, S. (2004) *Fitness Landscapes and the Origin of Species* (Princeton Univ. Press, Princeton, NJ).

Haig, D. (2002) *Genomic Imprinting and Kinship* (Rutgers Univ. Press, New Brunswick, NJ).

Holland, B. & Rice, W. R. (1999) Experimental removal of sexual selection reverses intersexual antagonistic coevolution and removes a reproductive load. *Proc. Natl. Acad. Sci. USA* **96,** 5083–5088.

Hosken D. J., Garner T. W. J. & Ward, P. I. (2001) Sexual conflict selects for male and female reproductive characters. *Curr. Biol.* **11,** 489–493.

Hughes, K. A. (1997) Quantitative genetics of sperm precedence in *Drosophila melanogaster. Genetics* **145,** 139–151.

Jaenike, J. (2001) Sex chromosome meiotic drive. *Annu. Rev. Ecol. Syst.* **32,** 25–49.

Kasule, F. K. (1986) Repetitive mating and female fitness in *Dysdercus-cardinalis* (Hemiptera, Pyrrhocoridae). *Zool. J. Linn. Soc.* **88,** 191–199.

Kruuk, L. E. B., Clutton-Brock, T. H., Slate, J., Pemberton, J. M., Brotherstone, S. & Guinness, F. E. (2000) Heritability of fitness in a wild mammal population. *Proc. Natl. Acad. Sci. USA* **97,** 698–703.

Linder, J. E. & Rice, W. R. (2005) Natural selection and genetic variation for female resistance to harm from males. *J. Evol. Biol.* 18, 568–575.

Martin, O. Y. & Hoskin, D. J. (2003) The evolution of reproductive isolation through sexual conflict. *Nature* **423,** 979–982.

Mayr, E. (1942) *Systematics and the Origin of Species* (Columbia Univ. Press, New York).

McKinney, F., Derrickson, S. R. & Mineau, P. (1983) Forced copulation in waterfowl. *Behavior* **86,** 250–294.

Moore, A. J., Gowaty, P. A., Wallin, W. G. & Moore, P. J. (2001) Sexual conflict and the evolution of female mate choice and male social dominance. *Proc. R. Soc. London Ser. B* **268,** 517–523.

Morrow, E. H., Arnqvist, G. & Pitnick, S. (2003) Adaptation versus pleiotropy: Why do males harm their mates? *Behav. Ecol.* **14,** 802–806.

Muller, H. J. (1942) Isolating mechanisms, evolution and temperature. *Biol. Symp.* **6,** 71–125.

Parker, G. A. & Partridge, L. (1998) Sexual conflict and speciation. *Philos. Trans. R. Soc. London B* **353,** 261–274.

Partridge, L. & Fowler, K. (1990) Nonmating costs of exposure to males in female *Drosophila-melanogaster. J. Insect Physiol.* **36,** 419–425.

Pitnick, S. & Garcia-Gonzalez, F. (2002) Harm to females increases with male body size in *Drosophila melanogaster. Proc. R. Soc. London Ser. B Biol.* **269,** 1821–1828.

Prout, T. & Clark, A. G. (2000) Seminal fluid causes temporarily reduced egg hatch in previously mated females. *Proc. R. Soc. London Ser. B* **267,** 201–203.

Reznick, D. N., Shaw, F. H., Rodd, F. H. & Shaw, R. G. (1997) Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). *Science* **275,** 1934–1937.

Rice, W. R. (1996) Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature* **381,** 232–334.

Rice, W. R. (1998) In *Endless Forms: Species and Speciation*, eds., Howard, D. J. & Berlocher, S. H. (Oxford Univ. Press, New York), pp. 261–270.

Rice, W. R. & Chippindale, A. K. (2001) Sexual recombination and the power of natural selection. *Science* **294,** 555–559.

Rice, W. R. & Holland, B. (1997) The enemies within: intergenomic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. *Behav. Ecol. Sociobiol.* **41,** 1–7.

Rice, W. R. & Hostert, E. E. (1993) Laboratory experiments on speciation—What have we learned in 40 years? *Evolution* **47,** 1637–1653.

Schoener, T. W., Spiller, D. A. & Losos, J. B. (2004) Variable ecological effects of hurricanes: The importance of seasonal timing for survival of lizards on Bahamian islands. *Proc. Natl. Acad. Sci. USA* **101,** 177–181.

Sgrò, C. M. & Partridge, L. (2000) Evolutionary responses of the life history of wild-caught *Drosophila melanogaster* to two standard methods of laboratory culture. *Am. Nat.* **156,** 341–353.

Swanson W. J. & Vacquier, V. D. (2002) The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* **3,** 137–144.

Vacquier, V. D., Swanson, W. J. & Lee, Y. H. (1997) Positive Darwinian selection on two homologous fertilization proteins: What is the selective pressure driving their divergence? *J. Mol. Evol.* **44,** S15–22.

Van Valen, L. (1973) A new evolutionary law. *Evol. Theory* **1,** 1–30.

Werren, J. H. (1997) Biology of *Wolbachia*. *Annu. Rev. Entomol.* **42,** 587–609.

White, M. J. D. (1978) *Modes of Speciation* (Freeman, San Francisco).

Wigby, S. & Chapman, T. (2004) Female resistance to male harm evolves in response to manipulation of sexual conflict. *Evolution* **58,** 1028–1037.

4

Chromosome Speciation: Humans, *Drosophila,* and Mosquitoes

FRANCISCO J. AYALA* AND MARIO COLUZZI†

Chromosome rearrangements (such as inversions, fusions, and fissions) may play significant roles in the speciation between parapatric (contiguous) or partly sympatric (geographically overlapping) populations. According to the "hybrid-dysfunction" model, speciation occurs because hybrids with heterozygous chromosome rearrangements produce dysfunctional gametes and thus have low reproductive fitness. Natural selection will, therefore, promote mutations that reduce the probability of intercrossing between populations carrying different rearrangements and thus promote their reproductive isolation. This model encounters a disabling difficulty: namely, how to account for the spread in a population of a chromosome rearrangement after it first arises as a mutation in a single individual. The "suppressed-recombination" model of speciation points out that chromosome rearrangements act as a genetic filter between populations. Mutations associated with the rearranged chromosomes cannot flow from one to another population, whereas genetic exchange will freely occur between colinear chromosomes. Mutations adaptive to local conditions will, therefore, accumulate differentially in the protected chromosome regions so that parapatric or partially sympatric populations will genetically differentiate, eventually evolving into

*Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697; and †Istituto Pasteur-Fondazione Cenci Bolognetti and Dipartimento di Scienze di Sanità Pubblica, Università di Roma La Sapienza, 00185 Rome, Italy.

different species. The speciation model of suppressed recombination has recently been tested by gene and DNA sequence comparisons between humans and chimpanzees, between *Drosophila* species, and between species related to *Anopheles gambiae*, the vector of malignant malaria in Africa.

The process of evolution is continuous through time but yields in space discontinuous groups of organisms. The continuity of the process links the myriad living organisms with the last universal common ancestor, from which all living organisms descend. Organisms evolve differences because of the haphazard mutation process, adaptation to different environmental circumstances, interaction with other organisms, constraints imposed by the organisms' past evolutionary history, and the like. The discontinuities are encompassed in the Linnean system of classification, which is hierarchical, with gradually more inclusive categories: species, genus, family, order, and so on.

"Species" is a category of classification, the most basic, within which are placed groups of organisms designated by specific names such as *Homo sapiens, Mus musculus, Drosophila melanogaster,* or *Quercus borealis.* But species have a biological reality that is lacking in more inclusive groups of organisms. In sexually reproducing organisms, individual members of a species are able to interbreed and thus share in a common gene pool. Collectively, there is variation among the members of a species, but there is also continuity in space and time. Species are evolutionary units. Because of these properties, some philosophers have affirmed that species, but not more inclusive groups of organisms, are metaphysical individuals. According to Hull (1977), "species fit as naturally into the idealized category of spatiotemporally localized individuals as do particular organisms" (Ghiselin, 1974).

Dobzhansky (1935a,b) pointed out in 1935 this double biological reality of the concept of species: (*i*) as a category of classification, just like genus, family, and other categories, a logical construct pragmatically necessary for organizing the enormous diversity of the living world and (*ii*) as a category with "an attribute peculiar only to itself" (1935b), because a species is a natural entity, a collectivity that has biological continuity defined, in sexually reproducing organisms, by the capacity to interbreed among individuals of the same species and their incapacity to interbreed with individuals of other species. The biological species concept, as it came to be known, defines species precisely by these two attributes: ability to interbreed within the species and reproductive isolation from other species. The evolutionary process of speciation, by which one species splits into two, is equivalent to the evolutionary emergence of reproductive isolation.

Dobzhansky (1935a) described a species as "a group of individuals fully fertile *inter se*, but barred from interbreeding with other similar groups by its physiological properties." But Dobzhansky was aware that complex concepts such as species cannot satisfactorily be defined by a particular set of words encompassed in a single sentence. He did not place the definition cited above in quotes and provided, in the same and other writings, additional definitions that pointed toward other species characteristics such as their being temporary instantiations of the evolutionary process: "Considered dynamically, the species represents that stage of evolutionary divergence at which the once actually or potentially interbreeding array of forms becomes segregated into two or more separate arrays which are physiologically incapable of interbreeding" (1935a). This definition is quoted by Dobzhansky (1937b) in the first edition of his classic *Genetics and the Origin of Species*, in a chapter significantly titled "Species as Natural Units." He adds: "Species is a stage in a process, not a static unit" (1937b). As he had earlier pointed out: "The fundamental importance of this stage is due to the fact that it is only the development of the isolating mechanisms that makes possible the coexistence in the same geographic area of different discrete groups of organisms. . . . This, in turn, opens the possibility for the organisms dwelling together to become adapted to different places in the general economy of nature" (1935a). Species are natural units that evolve and adapt autonomously. Dobzhansky noted that "a stage must exist in the process of evolutionary divergence at which an originally panmictic population becomes split into two or more populations that interbreed with each other no longer. . . . The emphasis should be placed, however, not on the absence of actual interbreeding between the different form complexes, but rather on the presence of physiological mechanisms making interbreeding difficult or impossible" (1935a). In *Genetics and the Origin of Species* (1937b), a full chapter is dedicated to "isolating mechanisms," a label that he had first proposed in Dobzhansky (1937a).

For nearly four decades, Dobzhansky continued to use similarly worded definitions of species, with the phrases "reproductive community" or "Mendelian population" sometimes replacing or added to his earlier "group of organisms" or "array of forms." In his last major work, *Genetics of the Evolutionary Process* (1970), Dobzhansky writes: "the process of species formation, in contrast to race formation, involves the development of reproductive isolating mechanisms. An ancestral species is transformed into two or more derived species when an array of interbreeding Mendelian populations becomes segregated into two or more reproductively isolated arrays. Species are, accordingly, systems of populations; the gene exchange between these systems is limited or prevented in nature by a reproductive isolating mechanism or perhaps by a combi-

nation of several such mechanisms. In short, a species is the most inclusive Mendelian population." Moreover, he again asserts: "Species is not only a category of classification, but also a form of supraindividual biological integration" (1970).

In *Systematics and the Origin of Species*, Mayr (1942) commended Dobzhansky for identifying interbreeding and reproductive isolation as the distinguishing features of the species concept and proposed a short definition: "Species are groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups." He has used, identically or with some word differences, this definition in later writings. Indeed, Mayr is generally perceived as the leading exponent of the biological species concept and the most successful investigator of the application of this concept to a great variety of species and species groups throughout the animal world, as several papers in this collection bear witness.

Mayr repeatedly wrote that species are real and not merely human constructs that are convenient for organizing biological diversity, as some taxonomists, as well as nominalist philosophers, would claim. He supported the claim by Ghiselin (1974), Hull (1977), and others that species are metaphysical individuals, once this language was introduced in the evolutionary literature (Mayr, 1976, 1987). The integration of its gene pool provides the necessary cohesion for any particular species taxon to be considered an "individual." The integration of gene pools, in turn, derives from the two dimensions incorporated in the definition of species, the ability of its members to interbreed, and their reproductive isolation from other species.

MODELS OF CHROMOSOMAL SPECIATION

Changes in chromosome number or structure may contribute to speciation. Polyploidy, the multiplication of the chromosome complement, may yield a new species in a single generation, reproductively isolated from its ancestral species. For example, a tetraploid plant crossed with a diploid ancestor produces sterile hybrid progeny. Polyploidy is more common among angiosperms than among gymnosperms. Nearly 50% of all existing angiosperm species are estimated to have arisen by ancient polyploidy, more of them by allopolyploidy (doubling of the chromosome complement in a hybrid between two previously existing species) than by autopolyploidy (multiplication of the chromosome complement of a single species). Polyploidy is also common among ferns. Some important cultivated plants are polyploids, such as wheat, oat, tobacco, potato, banana, strawberry, sugar cane, and coffee. Polyploidy is less common in animals; polyploidy species occur among hermaphrodites, such as earth-

worms and planarians, or in species with parthenogenetic females, such as some shrimps, sow bugs, moths, and, more commonly, beetles, as well as some fish and salamanders.

Chromosome rearrangements, such as Robertsonian fusions and fissions, translocations, and inversions, may play a role in speciation. There are a number of models proposing that chromosomal rearrangements accelerate genic diversification between populations and, therefore, facilitate speciation. We will consider two classes of models, which we will call the "hybrid-dysfunction" and "suppressed-recombination" models of speciation.

Hybrid-dysfunction models claim that recombination between rearranged chromosomes generates gametes with some chromosomal segments deleted and others duplicated, thereby creating a partial reproductive barrier because the heterokaryotypic hybrid exhibits reduced reproductive fitness, also called "underdominance." Under these conditions, natural selection will, in both populations, favor mutations that reduce the probability of intercrossing and will eventually lead to complete reproductive isolation. The great multiplication of species of flightless Australian grasshoppers of the subfamily Morabinae can largely be attributed, according to White (1968, 1978), to underdominance in hybrids between populations with different chromosome rearrangements. A chromosomal rearrangement may first become established in a small local colony, either at the periphery of the distribution area of the ancestral species or inside it, by random drift. The colony may expand within a certain area and there displace the ancestral form if its members display high fitness in that area. The low fitness of the hybrids will keep the two populations separate and facilitate the evolution of prezygotic isolating mechanisms, which will inhibit the formation of hybrids. White (1968, 1978) refers to this model of speciation as "stasipatric," because it largely occurs *in situ*, rather than allopatrically, yet it differs from models of sympatric speciation that do not attribute a major role for chromosome rearrangements.

The hybrid dysfunction model of speciation encounters the following disabling difficulty (Hey, 2003; Machado *et al.*, 2002; Navarro and Barton, 2003a,b; Noor *et al.*, 2001b; Rieseberg, 2001; Spirito, 2000). A chromosome rearrangement will first appear in the population as a mutation in a single individual. This individual will be able to mate only with individuals without the mutation. If hybrids have reduced fitness, the chromosome mutation will be selected against and eliminated from the population. The hybrid dysfunction model is unlikely to have much general validity, precisely because it seems so unlikely that a chromosome rearrangement that reduces the fitness of heterozygotes will be at all established within its ancestral population, although this may occasionally occur by random

drift, particularly if hybrid underdominance is only slight (in which case natural selection for reproductive isolation will be weak as well).

A speciation model of suppressed recombination was proposed by Coluzzi (1982) in his account of multiple speciation events within the species complex related to *Anopheles gambiae*, the main vectors in Africa for the transmission of malignant malaria, caused by the protozoan *Plasmodium falciparum*. The World Health Organization estimates that there are annually 300–500 million cases of malaria and >1 million deaths, mostly in sub-Saharan Africa, and the most of those who die are children. Seven species have been identifiedwithin the *A. gambiae* complex and have arisen, according to Coluzzi (1982; Coluzzi *et al.*, 2002), within the last 5,000 years.

Suppressed-recombination models of speciation have recently been proposed by Rieseberg (2001) to account for speciation in wild sunflowers (Riesberg *et al.*, 1995); by Noor *et al.* (2001a,b) and Machado *et al.* (2002), using evidence from the closely related *Drosophila persimilis*, *Drosophila pseudoobscura*, and *D. p. bogotana*; and by Navarro and Barton (2003a,b), who have mathematically modeled the process and supported the model's predictions by comparing genomic DNA sequences between humans and chimpanzees.

We will successively examine the human–chimpanzee and *Drosophila* evidence advanced in support of the model and then return to speciation in the *A. gambiae* complex.

HUMAN SPECIATION

Genomic studies have confirmed that a substantial number of chromosomal rearrangements have occurred between humans and chimpanzees. In particular, nine chromosomes (nos. 1, 4, 5, 9, 12, 15, 16, 17, and 18) exhibit pericentric inversions between humans and chimpanzees, and human chromosome 2 represents a fusion of two acrocentric chromosomes present in chimpanzees (chromosomes 12 and 13) and other great apes (chromosomes 11 and 12 in gorillas and orangutans) (de Boer and Seuánez, 1982; Yunis and Prakash, 1982). If these chromosomal rearrangements occurred early in the divergence between ancestral populations of chimps and humans, they "would facilitate genic divergence during the time when the diverging populations are in parapatry, i.e., have limited gene flow" (Navarro and Barton, 2003b). The hypothesis proposes that alleles favored in one or the other population will be trapped at the chromosomal barrier and thus would cause the two populations to diverge genically as they adapt to their distinct prevailing environmental conditions. Accumulation of incompatibilities would gradually result in reproductive isolation and speciation. In regions not protected by chromosome

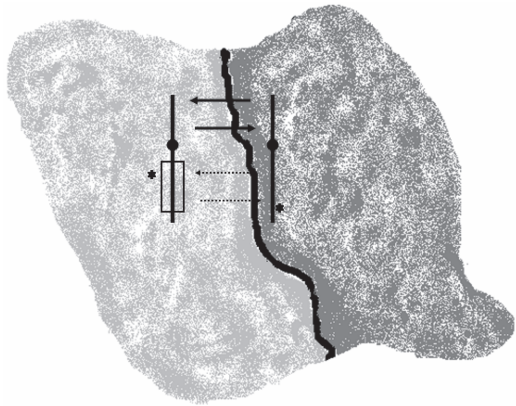


FIGURE 4.1 Two populations share a common boundary where hybridization occurs. Shown are two metacentric chromosomes that differ by an inversion (box) and incompatible alleles at two loci (*). Gene flow can readily occur along regions not linked to the inverted region (solid arrows) but is severely inhibited in regions linked to the inversion (dotted arrows). Natural selection favors the evolution of reproductive isolation between the populations by accumulation of incompatible alleles along the chromosome regions protected from recombination by the inversions. Figure was modified from Hey (2003).

rearrangements, allelic differences that might arise would not readily accumulate, because genetic recombination would tend to diffuse them between the populations (Fig. 4.1). This hypothesis can be tested, according to Navarro and Barton (2003b), by comparing genic differentiation between humans and chimps for different chromosome regions.

According to evolutionary theory, the rate of nonsynonymous nucleotide substitution per nonsynonymous site (K_A) is generally expected to be much lower than the rate of synonymous substitution per synonymous site (K_S), because random amino acid changes are usually deleterious, whereas synonymous changes are likely to be neutral or nearly so (Kimura, 1983). Thus, the expectation is $K_A << K_S$, except when positive selection is involved favoring particular amino acid replacements, in which case K_A will increase and may even become larger than K_S. K_A/K_S ratios close to or >1 indicate positive selection.

Navarro and Barton (2003b) have investigated nucleotide sequences that exhibit nucleotide differences between chimps and humans in 115 genes, about evenly distributed between rearranged chromosomes (59 genes) and colinear chromosomes (56 genes) (Table 4.1). Of the 26 genes with K_A/K_S ratios ≤1, 20 (76.9%) are located on rearranged chromosomes,

TABLE 4.1 Rate of Gene Evolution for Rearranged and Colinear Chromosomes in Humans Versus Chimpanzees

K_A/K_S	Rearranged Chromosomes	Colinear Chromosomes	Total
Number of genes (all)	59	56	115
>1	20	6	26
<1	39	50	89
Number of genes ($K_S > 0$)	54	54	108
>1	15	4	19
<1	39	50	89
K_A/K_S	0.84	0.37	0.61

Data are from Navarro and Barton (2003b). K_A and K_S are rates of evolution for nonsynonymous and synonymous nucleotides, respectively.

and only 6 are in colinear chromosomes. Of 89 chromosomes with K_A/K_S ratios >1, 39 (43.8%) are on rearranged chromosomes, and 50 are on colinear chromosomes. The average K_A/K_S ratio for rearranged chromosomes (0.84) is more than double the value for colinear chromosomes (0.37). (These ratios do not include seven genes for which $K_S = 0$, which would have given a ratio of infinity: five genes in rearranged chromosomes and two in colinear chromosomes.)

The K_A/K_S ratios and related results are consistent with the hypothesis tested, the suppressed-recombination model of speciation, although Navarro and Barton (2003b) also discussed other factors that may have been involved. The 2-fold difference between the two ratios is surprising, because it is so large. As these researchers note, and as also interpreted by Rieseberg and Livingstone (2003), a 2-fold difference could be explained, under the hypothesis, only if the chromosomal rearrangements have been barriers in parapatry and have exchanged genes through hybridization no less than half the time of divergence between the human and chimpanzee lineages. The human and chimpanzee lineages diverged 7–8 million years ago. *Sahelanthropus tchadensis* and *Orrorin tugenensis* are the oldest known taxa in the human lineage. They lived 6–7 million years ago and were prevailingly bipedal, at least when they were on the ground. *Ardipithecus ramidus*, *Australopithecus anamensis*, *Australopithecus afarensis*, and *Kenyanthropus platyops* count among the species of the human lineage that lived between 3.5 million and 5.5 million years ago, although not all of them may have been direct ancestors of *Homo sapiens*. These taxa were prevailingly or exclusively bipedal. For example, the pelvis of Lucy, a well-known specimen of *A. afarensis*, dated ≈3.5 million years ago, is very similar in configuration and proportions to a modern human pelvis and,

in any case, drastically different from that of a chimpanzee or any other modern ape. Given these and other anatomical incompatibilities, the inference that human and chimp ancestors may have hybridized for 3–4 million years after the first divergence of their lineages seems, *prima facie*, unlikely.

The inferences made by Navarro and Barton (2003b) encounter other theoretical difficulties. The model of suppressed recombination predicts that once incompatible alleles start to accumulate, the process will accelerate so that an increased fraction of allele substitutions in one of the species would be incompatible with the other species and speciation would rapidly occur (Hey, 2003; Orr and Turelli, 2001). Therefore, the period during which genetic differences accumulate preferentially in the rearranged chromosomes should not last very long. Moreover, millions of years have elapsed of separate evolution between the human and chimpanzee lineages, which should have largely erased the signal predicted by the model: namely, the expected greater genic differentiation between the rearranged chromosomes than between the colinear chromosomes, because this differentiation would have occurred so long ago, and other processes would have largely contributed to the current genetic differentiation between the two lineages. One additional difficulty is the implied assumption that the chromosomal rearrangements that differentiate humans from chimps and other apes happened all in the human lineage, early and within a short time.

Several ensuing empirical investigations have brought into question the inferences of Navarro and Barton (2003b). Lu *et al.* (2003) have estimated K_A, K_S, and K_A/K_S in 85 genes, about equally distributed between rearranged chromosomes (42 genes) and colinear chromosomes (43 genes). Between humans and chimpanzees, the K_A/K_S ratio is 1.41 times greater (0.820 versus 0.581) for rearranged chromosomes than for colinear chromosomes (Table 4.2), consistent with the result of Navarro and Barton (2003b), although the difference is smaller than the >2-fold (2.2) increase observed by these researchers. Further, as a control, they have compared the same 85 genes between humans and other primate species, namely, 21 genes from either orangutan or gibbon, plus 64 genes from Old World monkeys. Remarkably, the K_A/K_S ratios are also 1.41 times greater for rearranged chromosomes than for colinear chromosomes in humans (0.623 versus 0.443) (Table 4.2). This result is consistent with the hypothesis that positive selection may have more intensely promoted genic evolution in the rearranged chromosomes but not with Navarro and Barton's (2003b) hypothesis that the high K_A/K_S ratios are associated with speciation between humans and chimpanzees according to the suppressed-recombination model. Lu *et al.* (2003) suggest that the higher K_A/K_S ratios observed in rearranged chromosomes may simply reflect a bias in the distribution of rapidly evolving

TABLE 4.2 Rate of Gene Evolution for Rearranged (R) and Collinear (C) Chromosomes in Humans Versus Chimpanzees or Other Primates

Rate of Evolution	Rearranged	Colinear	R/C
Human: chimpanzee (all genes, 109)			
K_A/K_S	0.780 (54)	0.483 (55)	1.61
K_A	0.007 (54)	0.005 (55)	1.40
K_S	0.016 (54)	0.017 (55)	0.94
Human: chimpanzee (overlapping genes, 85)			
K_A/K_S	0.820 (42)	0.581 (43)	1.41
K_A	0.007 (42)	0.006 (43)	1.16
K_S	0.016 (42)	0.014 (43)	1.14
Human: outgroup primates (overlapping genes, 85)			
K_A/K_S	0.623 (42)	0.443 (43)	1.41
K_A	0.031 (42)	0.024 (43)	1.30
K_S	0.061 (42)	0.059 (43)	1.03

Data are from Lu *et al.*, 2003. K_A and K_S are as defined in Table 1. Number of genes is given in parentheses.

genes among chromosomes. Genes in the rearranged chromosomes may happen to be evolving faster than genes in colinear chromosomes. Faster evolution is known to be the case for certain genes; for example, genes encoding glycophorins and protamines, which happen to be located on rearranged chromosomes, are known to be rapidly evolving in all higher primates (Lu *et al.*, 2003).

Zhang *et al.* (2004) have examined nucleotide differentiation between human and chimpanzee DNA sequences amounting to 4,108,833 nucleotides: 1,831,676 nucleotides distributed among seven rearranged chromosomes, and 2,277,157 nucleotides distributed among six colinear chromosomes. The average nucleotide divergence per chromosome is 1.20 ± 0.14% for the rearranged chromosomes and larger (1.40 ± 0.08%) for the colinear chromosomes (Table 4.3). This finding is inconsistent with Navarro and Barton's (2003b) hypothesis; indeed, the rate of nucleotide evolution is faster in the colinear chromosomes than in the rearranged

TABLE 4.3 Nucleotide Sequence Divergence Between Humans and Chimpanzees

Chromosomes	Number of Nucleotides	Percent Divergence
Rearranged chromosomes	1,831,676	1.20 ± 0.14
Colinear chromosomes	2,277,157	1.40 ± 0.08
Total	4,108,833	1.29 ± 0.09

Data are from Zhang *et al.* (2004).

TABLE 4.4 Nonsynonymous (*N*) and Synonymous (*S*) Nucleotide Divergence Between Humans and Chimpanzees

Chromosomes	Number of Genes	Number of Nucleotides	*N*	*S*	*N*/*S*	d_N/d_S
Rearranged	568	202,455	288	596	0.483	0.236 ± 0.028
Colinear	512	184,728	288	532	0.541	0.239 ± 0.019
Total	1,080	387,183	576	1,128	0.510	0.237 ± 0.016

Data are from Zhang *et al.* (2004). d_N and d_S are the rates of nonsynonymous and synonymous substitution, respectively.

chromosomes. The nucleotide sequences analyzed may have largely involved noncoding DNA, thus evolving mostly neutrally. But the suppressed recombination model predicts that the rate of molecular evolution, neutral or not, should be greater in rearranged chromosomes than in colinear chromosomes.

A comparison of the number of nonsynonymous (*n*) versus synonymous substitutions between humans and chimpanzees by using cDNA sequences and thus involving protein coding genes has also been performed by Zhang *et al.* (2004). The total number of nucleotides examined is approximately the same for the 10 rearranged chromosomes (202,455 nucleotides) as for the 13 colinear chromosomes (184,728 nucleotides), and so is the number of genes in each class (568 versus 512 genes in rearranged versus colinear chromosomes). The ratio of nonsynonymous to synonymous substitutions is 0.483 in rearranged chromosomes, and the ratio is 0.541 in colinear chromosomes (Table 4.4). This result is contrary to the one observed by Navarro and Barton (2003b) and is inconsistent with their hypothesis of an accelerated rate of evolution in the rearranged chromosomes. An additional analysis involving 304 gene sequences of the ratio of nonsynonymous to synonymous rates of evolution yields a higher proportion of genes under positive selection in colinear chromosomes (29.5% of genes with ratios >1) than in rearranged chromosomes (15.5% of genes with ratios >1) (Table 4.5).

Cáceres *et al.* (2003) have importantly discovered that a large majority (≈90%) of 169 genes with a great variety of functions exhibit higher expression levels in the cerebral cortex of humans than in the cerebral cortex of chimpanzees, which suggests elevated levels of neuronal activity in the human brain. Marquès-Bonet *et al.* (2004) have distributed the genes analyzed by Cáceres *et al.* (2003) into those present in colinear and rearranged chromosomes, observing that the average proportional increase of expression in humans is 1.463 ± 0.014 (mean ± SE) in colinear chromosomes but 1.543 ± 0.019 in rearranged chromosomes, a difference that is statisti-

TABLE 4.5 Number of Genes Under Positive Selection ($d_N/d_S > 1$) in Rearranged (R) Versus Colinear (C) Human Chromosomes

d_N/d_S	Rearranged	Colinear	R/C
Total genes	148	156	304
>1	23 (15.5)	46 (29.5)	0.500
<1	125 (84.5)	110 (71.5)	1.136

Data are from Zhang *et al.* (2004). d_N and d_S are as defined in Table 4.4. Percentages are given in parentheses.

cally significant by a permutation test. They interpret these results as supporting their theory of chromosomal speciation by suppressed recombination. However, Zhang *et al.* (2004) have analyzed the same data and similarly observed increased expression in rearranged chromosomes (80 genes) relative to colinear chromosomes (72 genes), but neither the difference nor the average of the differences between individual chromosomes is statistically significant. Table 4.6 shows that genes in rearranged chromosomes 4, 5, and 9, as well as in colinear chromosome 22, have human expression levels >1.6 times higher than in chimpanzees, but in none of these chromosomes is the expression level significantly higher than the average for all genes of 1.503 (1.463 and 1.543 for colinear and rearranged chromosomes, respectively). The importance of Cáceres *et al.* (2003) is the higher expression levels in the human cerebral cortex, but gene expression increases may have happened at various times. The evidence shown in Table 4.6 does not warrant the conclusion that gene expression increases in the human cortex were distinctly promoted in rearranged chromosomes at the time when the human lineage first separated from the chimpanzee lineage some 7–8 million years ago. Brain size in the human lineage notably increased starting with *Homo habilis*, ≈2 million years ago. It may very well have been the case that gene expression increases in the cerebral cortex were, at least in good part, associated with the increase, by about a factor of 3, that happened in the human lineage during the last 2 million years.

Any increased genic differentiation that may have preferentially occurred in rearranged chromosomes during the original divergence of humans and chimps would likely be undetectable after millions of years of further divergent evolution in the two lineages. Among the apes, there are favorable populations for testing the suppressed-recombination model of speciation, such as the eight Bornean subspecies or races of orangutans that are parapatric in their distribution, with borderlines formed by rivers [the eight subspecies have been given formal taxonomic names (von

TABLE 4.6 Average Proportional Increase in Gene Expression in the Cerebral Cortex of Humans Compared with Chimpanzees

Chromosomes	Colinear	Rearranged
1		1.55 ± 0.04
2		1.45 ± 0.03
3	1.46 ± 0.03	
4		1.62 ± 0.08
5		1.67 ± 0.08
6	1.43 ± 0.03	
7	1.51 ± 0.05	
8	1.41 ± 0.04	
9		1.66 ± 0.10
10	1.38 ± 0.03	
11	1.42 ± 0.03	
12		1.43 ± 0.04
13	1.45 ± 0.05	
14	1.50 ± 0.06	
15		1.50 ± 0.07
16		1.56 ± 0.07
17		1.53 ± 0.05
18		1.54 ± 0.09
19	1.53 ± 0.05	
20	1.41 ± 0.04	
21	1.48 ± 0.07	
22	1.63 ± 0.10	
Total	1.463 ± 0.014	1.543 ± 0.019

Data are from Marquès-Bonet *et al.* (2004).

Koenigswald, 1982)]. Pericentric inversions, such as those between humans and chimpanzees, are common among the great apes, including those in chromosomes 2 and 3 that differentiate the Sumatran and Borneo subspecies of orangutans; polymorphic pericentric inversions occur within each one of the two subspecies, for instance, in chromosome 9 (de Boer and Seuánez, 1982).

SPECIATION IN DROSOPHILA

D. pseudoobscura is a widely distributed Nearctic species common in temperate forests throughout the western third of North America, extending from British Columbia to Guatemala and from the Pacific to the Great Plains. A geographically isolated subspecies, *D. p. bogotana,* lives in the altiplano near Bogota, Colombia. More narrowly distributed than *D. pseudoobscura* is its sibling species, *D. persimilis*, common in the temperate forests of the American northwest, from British Columbia to southern

California. Throughout this territory, the two species are sympatric and abundant, but *D. persimilis* tends to be more common in cooler, moister environments. *D. persimilis* and *D. pseudoobscura* are reproductively isolated by a cascade of reproductive isolating mechanisms that include sexual isolation, hybrid male sterility, hybrid male courtship dysfunction, and hybrid backcross inviability. Hybrids have been found in nature at extremely low frequencies (Dobzhansky, 1973; Powell, 1997), but they evince that there is a possibility for gene introgression across the species by means of backcrosses of hybrid females to the parental species. Interspecific hybrid females are fertile, even though males are sterile.

The nuclear genome of the species is distributed in five large chromosome arms and one very small chromosome. The X chromosome is metacentric, incorporating two chromosome arms. The two species differ by fixed large paracentric inversions on the left arm of chromosome X and on chromosome 2 (Anderson *et al.*, 1977; Moore and Taylor, 1986). The genes associated with isolating mechanisms between *D. pseudoobscura* and *D. persimilis* are largely located on these chromosome inversions, with no evidence of major barriers decreasing the potential for gene flow across the other major chromosome arms, just as would be expected according to the suppressed-recombination model of speciation (Noor *et al.*, 2001a,b). Fixed inversion polymorphisms inhibiting recombination would facilitate genic differentiation along the inverted segments, where genes promoting reproductive isolation between the species would have gradually accumulated. Consistent with this interpretation is the observation that the reproductive isolation factors between the allopatric pair, *D. persimilis* and *D. p. bogotana,* are located in the colinear regions of their chromosomes rather than on the inverted segments (Brown *et al.*, 2004), demonstrating that the protection against recombination provided by chromosomal inversions has played no particular role in promoting reproductive isolation between these two populations, which are not geographically contiguous or overlapping.

The divergence between *D. pseudoobscura* and *D. persimilis* is estimated, by molecular and other information, to have occurred some 500,000 years ago; the separation between the two subspecies *D. p. pseudoobscura* and *D. p. bogotana* is dated ≈150,000 years ago, when a propagule of the species somehow colonized the altiplano near Bogota (Brown *et al.*, 2004; Wang *et al.*, 1997). Machado *et al.* (2002) have examined DNA sequence variation at 11 loci in a large sample of strains geographically representative of the three taxa. Their results are consistent with gene flow between the largely sympatric *D. pseudoobscura* and *D. persimilis* but not between the allopatric *D. p. pseudoobscura* and *D. p. bogotana.* The evidence indicates that most of the gene flow is ancient rather than recent. Most importantly, as predicted by the suppressed-recombination model

of speciation, genomic regions associated with reproductive isolation between *D. pseudoobscura* and *D. persimilis* show less evidence of gene flow than regions not so associated.

SPECIATION IN MALARIA'S MOSQUITOES

Fig. 4.2 shows the geographic distribution of seven species of the *A. gambiae* complex, and Fig. 4.3 shows their phylogeny. Epidemiologically most important are *A. gambiae* and *Anopheles arabiensis*. Where the geographic distributions of these two species overlap, there is competitive exclusion between them, with *A. gambiae* prevailing in the rain forests and other messic habitats and *A. arabiensis* prevailing in xeric habitats. The seven species are siblings that are morphologically nearly indistinguishable, but they differ in genetic and ecological attributes, including breeding sites, as well as egg configuration and some subtle morphological traits. Chromosome rearrangements are common, so that the species are identified by their chromosome configurations, primarily inversion polymorphisms of the second and third chromosomes, which are particularly extensive in *A. gambiae* and *A. arabiensis* and are distinctive for each species (Coluzzi and Bradley, 1999; Coluzzi *et al.*, 2002).

A. arabiensis is considered the most likely ancestral species, which may have originated in the Middle East and reached Africa through the Arabian peninsula. Two sources of evidence support *A. arabiensis* as the ancestral species: It is the only member of the complex present in the Horn of Africa and in the Arabian peninsula, and it exhibits a fixed second-chromosome arrangement (labeled 2La) (Coluzzi *et al.*, 2002), which is thought to be ancestral because it is also present in other species groups such as the *Anopheles subpictus* complex, where it is fixed in at least one of the siblings. Various sources of evidence indicate that *A. arabiensis* was originally zoophilic and exophilic but acquired anthropophily and domesticity secondarily in Sudan and West Africa, the regions where this species exhibits its most extensive chromosome polymorphisms. *A. arabiensis* may have first dispersed in East Africa starting >6,000 years ago and soon reached Madagascar, where it remains zoophilic and exophilic, having failed there to adapt to human environments perhaps because low human density has not provided suitable selective pressure for this adaptation. In East Africa, where human density is higher, *A. arabiensis* gradually became anthropophilic, although never to the extent of *A. gambiae.*

Chromosome inversion patterns indicate that *A. arabiensis* gave rise to *Anopheles quadriannulatus,* from which it notably differs by three X-chromosome inversions, where factors are located that account for the reproductive isolation between the two species and which has retained the ancestral condition of being zoophilic and exophilic. *A. quadriannulatus*

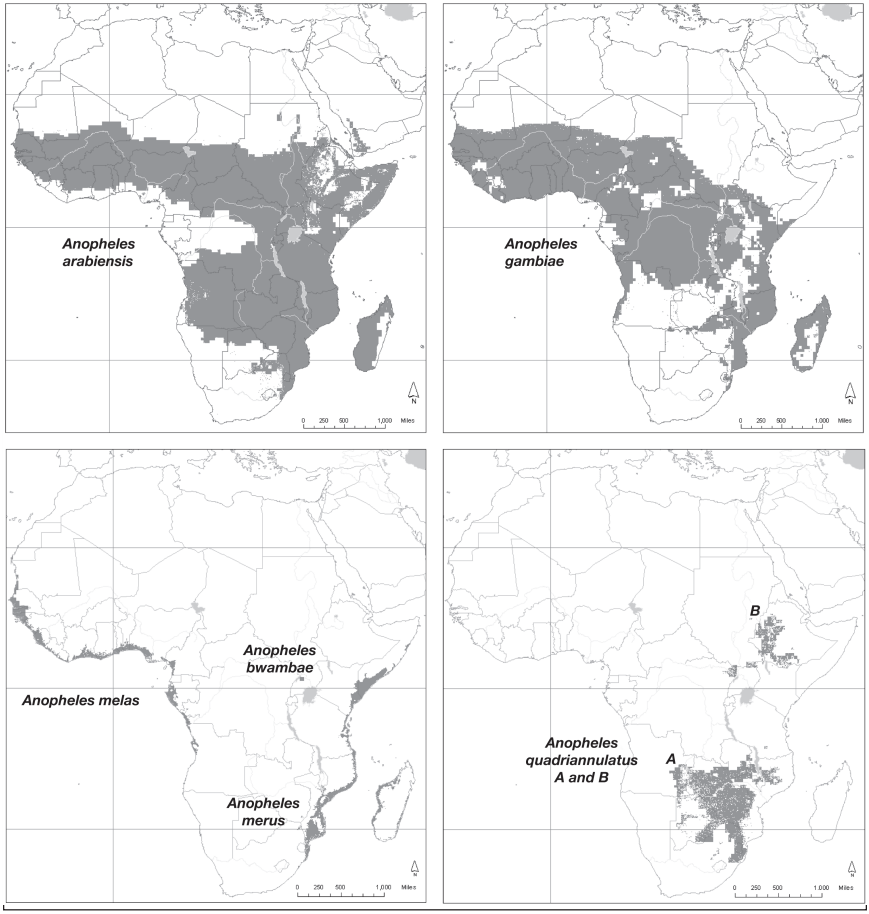


FIGURE 4.2 Geographic distribution of *A. gambiae* and six other closely related species. *A. arabiensis*, descended from a *Pyretophorus* species from the Arabian peninsula, is the likely ancestral species of the complex (see Fig. 4.3). Originally, *A. arabiensis* was zoophilic and exophilic, but it became anthropophilic and domestic by gradual adaptation to the human environment in Sudan and western Africa. *A. quadriannulatus A* and *A. quadriannulatus B* retain the original zoophily and exophily of their ancestral homonymous species, which also gave rise to *A. bwambae* and *A. melas*, and to *A. gambiae*, the most effective vector of malignant human malaria. *A. gambiae* and its strong anthropophily evolved <4000 B.P. with human invasion of the rain forest and introduction of slash-and-burn agriculture. In western Africa, *A. gambiae* is well represented in the Sahel region, extending up to 18° N, also the northern limit of *A. arabiensis*. In the Sudan, *A. arabiensis*, but not *A. gambiae*, is found along the river Nile upwards to the Egyptian border. Genetic data indicate that *A. merus* descends from *A. gambiae* and became adapted to breed in brackish, tide-dependent pools independently of *A. melas*.

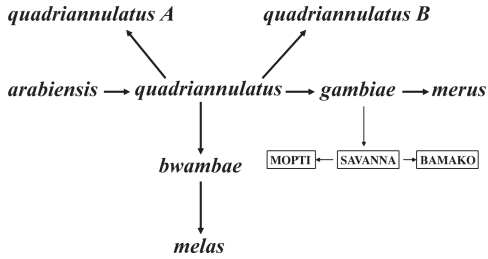


FIGURE 4.3 Phylogeny of seven species and three incipient species related to *A. gambiae*. The likely ancestral species is *A. arabiensis*, which differs from *A. quadriannulatus* by three X-chromosome inversions and differs from *A. gambiae* by two additional X-chromosome inversions. Reproductive factors among these species are primarily located on the X chromosome. Several incipient species can be recognized that are related to *A. arabiensis*, *A. melas*, or *A. gambiae*. Three incipient species related to *A. gambiae*, labeled Mopti, Savanna, and Bamako, are shown.

has split into two species, *A. quadriannulatus A* in southern Africa and *A. quadriannulatus B* in Ethiopia, which have homosequential chromosomes (although two polymorphic inversions are present in species *A*). These two allopatric species represent relics of the ancestral species, which genically diverged from each other after their geographic distribution became discontinuous. Two lineages originated from *A. quadriannulatus:* one giving rise to *Anopheles bwambae*, with a restricted geographic presence in northeast Uganda, and *Anopheles melas*, with a narrow distribution along the western coast of Africa, and a second, far more important lineage giving rise to *A. gambiae.*

The origins of *A. gambiae* can be traced to the late Neolithic, <4000 B.P. *A. gambiae* exhibits the primitive chromosome arrangement 2R (used within the complex as the standard of reference), which is adapted to the African rain forest, where, nevertheless, *A. gambiae* can only breed in environments modified by human agriculture, given that the larvae are "eliophilic," requiring sunlight for breeding (Coluzzi *et al.*, 2002). Agriculture was introduced in Africa ≈8000 B.P., imported from Mesopotamia into the lower Nile valley, but the forest remained for a long time impenetrable, without any traces of agricultural activity up to 4000 years B.P. Extensive penetration of the forest began ≈3000 B.P., made possible by climate change and the temporary "savannization" of much of the central African rain forest, a process which began ≈2800 B.P. and lasted ≈5 centuries (Willis *et al.*, 2004). When the forest belt regained its original range of distribution, ≈2300 B.P., it was invaded by Bantu agriculturalists who adopted "slash-and-burn" agricultural techniques. Increase in rainfall and

the return of the forest were determinant factors for the spreading of the tze-tze fly (*Glossina*), vector of the lethal animal trypanosomiasis (*Trypanosoma brucei*), which decimated cattle, thus promoting the adaptation of *A. quadriannulatus* to humans, who became an easy host for the blood meal and, more importantly, caused the opening of the forest canopy. These conditions promoted strong selection for anthropophily and domesticity, which facilitated the evolution of *A. gambiae* (Coluzzi, 1982; Willis *et al.*, 2004). There was a consequent increase in the rate of human infection by *P. falciparum*, which in turn led to strong selection for highly virulent strains, such as are the current forms of this parasite. Genetic evidence indicates that the expansion of malignant *P. falciparum* in Africa and throughout the world tropics occurred only within the last few thousand years (Hartl, 2004; Rich *et al.*, 1998; Tishkoff *et al.*, 2001).

A. gambiae differs from its ancestral *A. quadriannulatus* by two X-chromosome inversions, where factors responsible for their reproductive isolation are located (as is the case for the three X-chromosome inversions fixed between *A. arabiensis* and *A. quadriannulatus*; see above). *A. gambiae* has a nearly continentwide distribution in sub-Saharan African and is highly anthropophilic, like *A. arabiensis*, with which it accounts for most of the transmission in Africa of *P. falciparum* (but surpasses it). Further speciation is occurring within *A. gambiae* (as well as within *A. arabiensis* and *A. melas*). In southern Mali and northern Guinea, there are three chromosomally distinct populations of *A. gambiae* that are partially sympatric or parapatric and manifest enough assortative mating to qualify as incipient species. These populations are named "Savanna," "Mopti," and "Bamako" (Fig. 4.3). Chromosome-rearrangement evidence indicates that *A. gambiae* is the ancestor of *A. merus*, a species adapted to breed in tide-dependent pools.

A suppressed-recombination model of speciation was proposed by Coluzzi (1982) to account for the speciation patterns elucidated in the *A. gambiae* complex of species, particularly the successive adaptation of new species to environments quite different from the xeric conditions to which *A. arabiensis* was originally suited, such new environments as the rain forest or brackish waters. In outline, the essential components of the model are as follows (Coluzzi, 1982). A chromosomally monomorphic population colonizes a suitable environment and expands in numbers. The population will encounter adaptively marginal conditions, whether at the periphery of its distribution or at different times as a consequence of the seasons or other climatic and ecological oscillations. Mutations that increase adaptation to these marginal conditions will be favored wherever such conditions exist, yielding new "ecotypes" [i.e., "intraspecific groups having distinctive characters that result from the selective pressure of the local environment" (Lincoln *et al.*, 1998)]. As the prevalence of the marginal

conditions oscillates, so will the subpopulations adapted to them, resulting in population flushes and crashes. Breeding between individuals from a subpopulation with those of the central population will tend to diffuse the alleles that are distinctively adapted to the marginal conditions, except for alleles that may have been captured within a chromosome rearrangement. Interbreeding between the central and the marginal population in parapatric zones of contact will thus homogenize their genetic makeup, except for the alleles protected by the chromosome rearrangements, where new adaptive alleles will accumulate, including those that will promote reproductive isolation. Reproductive isolation will gradually evolve, yielding incipient speciation and eventually full species.

Coluzzi (1982) explored the variety of environmental and genetic conditions that may yield various possible outcomes, the most significant of which are (*i*) speciation, as just described, (*ii*) extinction of the distinctive genotypes adapted to the marginal conditions, and (*iii*) incorporation of the rearranged chromosomes into the main population as an adaptive polymorphism, particularly when the heterozygotes exhibit overdominance. Speciation would facilitate adaptation to the originally marginal conditions, leading eventually to full exploitation of new environments or ecological niches. The fixed X-chromosome inversions that differentiate *A. arabiensis*, *A. quadriannulatus*, and *A. gambiae* include reproductive isolation factors between these species and thus would have been instrumental in the speciation process, according to the suppressed-recombination model. Fixed and polymorphic inversions in other chromosomes, particularly on the right arm of chromosome 2, characterize genetically distinct subpopulations that have adapted to different regions and niches. Some are associated with incipient speciation processes that still prevail within this young species complex (see Fig. 4.2 and Coluzzi *et al.*, 2002).

The suppressed-recombination model of chromosome speciation predicts that sympatric sister species will be more different with respect to fixed chromosome rearrangements than allopatric sister species. Such is the state of affairs prevailing among the sibling species of the *A. gambiae* complex. Fixed rearrangements occur between sympatric sister species, but not between the two allopatric species *A. quadriannulatus A* and *A. quadriannulatus B*, relics of a widely distributed species that genically diverged allopatrically after their geographic distribution became discontinuous.

The suppressed-recombination speciation model also predicts that while gene exchange persists between the diverging populations, genes protected by the rearrangements will accumulate allelic differences faster than genes in the colinear chromosomes, where gene flow occurs between populations. The recent origin of seven species of the *A. gambiae* complex and the continuing processes of incipient speciation throughout the com-

plex provide an ideal system to test this prediction. Hybrid females are fertile between all of the species pairs of the complex. Therefore, the potential for gene flow persists up to the present, although hybrids are rarely detected in nature (0.02–0.76%) (Temu, 1997; Touré *et al.*, 1998). In any case, the increased genic differentiation between the rearranged chromosomes that would have taken place during the speciation process should be detectable, given that it has occurred within the last few thousand years, rather than hundreds of thousands (as between *D. pseudoobscura* and *D. persimilis*) or millions of years (as between humans and chimpanzees) ago.

The currently available evidence is limited but consistent with the model. An investigation of DNA sequence variation in four gene loci (one from each of chromosomes X and 2 and two from chromosome 3) sampled from multiple specimens shows considerable similarity of gene polymorphisms and even haplotypes between the two autosomal chromosomes of *A. gambiae* and *A. arabiensis* (Besansky *et al.*, 2003), which share polymorphic inversions. At the X chromosome locus, however, there are fixed nucleotide differences and greater overall nucleotide differentiation between *A. gambiae* and *A. arabiensis*. These results are consistent with the suppressed-recombination model because the X chromosomes of the two species are fixed for different inversions, where reproductive factors are located.

The shared nucleotide polymorphisms in the autosomal chromosomes may, however, be ancestral rather than originated by recent gene flow between the two species. An investigation of the *ND5* locus of mitochondrial DNA shows that ancestral haplotypes persist in the two species, *A. gambiae* and *A. arabiensis* (Donnelly *et al.*, 2004). Nevertheless, comparison of allopatric and sympatric populations suggests locale-specific unidirectional introgression from *A. arabiensis* into *A. gambiae*. Indeed, the acquisition by *A. gambiae* of alleles from the more arid-adapted *A. arabiensis* may have contributed to its spread and ecological dominance (Besansky *et al.*, 2003). A more definitive test of the suppressed-recombination model of chromosomal speciation would call for estimation of the K_A/K_S ratio in 20 or more genes distributed among the three chromosomes of these two species, as well as among the other five species and additional incipient species.

It may be worth adding that among the nearly 500 known members of the genus *Anopheles*, there are no fewer than 170 cryptic taxa belonging to 30 complexes of closely related species (Harbach, 2004). Most siblings are outcomes of recent speciation processes detected by paracentric inversions, mostly involving the X chromosome, as well as ribosomal DNA sequences. The chromosome rearrangements act as mechanisms of crossover suppression for reorganized regulatory units of gene expression.

These recently evolved complexes of *Anopheles* taxa are splendid materials for testing models of speciation mediated by chromosome rearrangements.

REFERENCES

Anderson, W. W., Ayala, F. J. & Michod, R. E. (1977) Chromosomal and allozymic diagnosis of three species of *Drosophila*: *Drosophila pseudoobscura, D. persimilis*, and *D. miranda*. *J. Hered.* **68,** 70–74.

Besansky, N. J., Krzywinski, J., Lhemann, T., Simard, F., Kern, M., Mukabayire, O., Fontenille, D., Touré, Y. & Sagnon, N. F. (2003) Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: Evidence from multilocus DNA sequence variation. *Proc. Natl. Acad. Sci. USA* **100,** 10818–10823.

Brown, K. M., Burk, L. M., Henagan, L. M. & Noor, M. A. F. (2004) A test of the chromosomal rearrangement model of speciation in *Drosophila pseudoobscura***.** *Evolution* **58,** 1856–1860.

Cáceres, M., Lachuer, J., Zapala, M. A., Redmond, J. C., Kudo, L., Geschwind, D. H., Lockhart, D. J., Preuss, T. M. & Barlow, C. (2003) Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl. Acad. Sci. USA* **100,** 13030–13035.

Coluzzi, M. (1982) Spatial distribution of chromosomal inversions and speciation in anopheline mosquitoes. In *Mechanisms of Speciation* (Liss, New York), pp. 143–153.

Coluzzi, M. & Bradley, D., eds. (1999) *The Malaria Challenge After One Hundred Years of Malariology* (Lombardo Editore, Rome). *Parassitologia* **41**(1–3).

Coluzzi, M., Sabatini, A., della Torre, A., Di Deco, M. A. & Tetrarca, V. (2002) A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* **298,** 1415–1418.

de Boer, L. E. M. & Seuánez, H. N. (1982) The chromosomes of the oran utan and their relevance to the conservation of the species. In *The Orang Utan: Its Biology and Conservation*, ed. de Boer, L. E. M. (W. Junk, The Hague, The Netherlands), pp. 135–170.

Dobzhansky, T. (1935a) A critique of the species concept in biology. *Philos. Sci.* **2,** 344–355.

Dobzhansky, T. (1935b) *Drosophila Miranda*, a new species. *Genetics* **20,** 377–391.

Dobzhansky, T. (1937a) Genetic nature of species differences. *Am. Nat.* **71,** 404–420.

Dobzhansky, T. (1937b) *Genetics and the Origin of Species* (Columbia Univ. Press, New York).

Dobzhansky, T. (1970) *Genetics of the Evolutionary Process* (Columbia Univ. Press, New York).

Dobzhansky, T. (1973) Is there gene exchange between *Drosophila pseudoobscura* and *Drosophila persimilis* in their natural habitats? *Am. Nat.* **107,** 312–314.

Donnelly, M. J., Pinto, J., Girod, R., Besansky, N. J. & Lehmann, T. (2004) Revisiting the role of introgression vs shared ancestral polymorphisms as key processes shaping genetic diversity in the recently separated sibling species of the *Anopheles gambiae* complex. *Heredity* **92,** 61–68.

Ghiselin, M. (1974) A radical solution to the species problem. *Syst. Zool.* **23,** 536–544.

Harbach, R. E. (2004) The classification of genus *Anopheles* (Diptera : Culicidae): A working hypothesis of phylogenetic relationships**.** *Bull. Entomol. Res.* **94,** 537–553.

Hartl, D. L. (2004) The origin of malaria: Mixed messages from genetic diversity**.** *Nat. Rev. Microbiol.* **2,** 15–22.

Hey, J. (2003) Speciation and inversions: Chimps and humans. *BioEssays* **25,** 825–828.

Hull, D. L. (1977) The ontological status of species as evolutionary units. In *Foundational Problems in Special Sciences*, eds. Butts, R. & Hintikka, J. (Kluwer, Dordrecht, The Netherlands), pp. 91–102.

Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, U.K.).

Lincoln, R. J., Boxshall, G. A. & Clark, P. F. (1998) *A Dictionary of Ecology, Evolution, and Systematics* (Cambridge Univ. Press, Cambridge, U.K.), 2nd Ed.

Lu, J., Li, W. H. & Wu, C. I. (2003) Comment on "Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes." *Science* **302,** 988.

Machado, C. A., Kliman, R. M., Markert, J. A. & Hey, J. (2002) Inferring the history of speciation from multilocus DNA sequence data: The case of *Drosophila pseudoobscura* and close relatives. *Mol. Biol. Evol.* **19,** 472–488.

Marquès-Bonet, T., Cáceres, M., Bertranpetit, J., Preuss, T. M., Thomas, J. W. & Navarro, A. (2004) Chromosomal rearrangements and the genomic distribution of gene-expression divergence in humans and chimpanzees. *Trends Genet.* **20,** 524–529.

Mayr, E. (1942) *Systematics and the Origin of Species* (Columbia Univ. Press, New York).

Mayr, E. (1976) Is the species a class or an individual? *Syst. Zool.* **25,** 192.

Mayr, E. (1987) The ontological status of species: Scientific progress and philosophical terminology. *Biol. Philos.* **2,** 146–166.

Moore, B. C. & Taylor, C. E. (1986) Drosophila of Southern California. 3. Gene arrangements of *Drosophila persimilis*. *J. Hered.* **77,** 313–323.

Navarro, A. & Barton, N. H. (2003a) Accumulating postzygotic isolation genes in parapatry: A new twist on chromosomal speciation. *Evolution* **57,** 447–459.

Navarro, A. & Barton, N. H. (2003b) Chromosomal speciation and molecular divergence—Accelerated evolution in rearranged chromosomes. *Science* **300,** 321–324.

Noor, M. A. F., Grams, K. L., Bertucci, L. A., Almendarez, Y., Reiland, J. & Smith, K. R. (2001a) The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. *Evolution* **55,** 512–521.

Noor, M. A. F., Grams, K. L., Bertucci, L. A. & Reiland, J. (2001b) Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci. USA* **98,** 12084–12088.

Orr, H. A. & Turelli, M. (2001) The evolution of postzygotic isolation: Accumulating Dobzhansky–Muller incompatibilities. *Evolution* **55,** 1085–1094.

Powell, J. R. (1997) *Progress and Prospects in Evolutionary Biology: The* Drosophila *Model* (Oxford Univ. Press, New York).

Rich, S. M., Licht, M. C., Hudson, R. R. & Ayala, F. J. (1998) Malaria's Eve: Evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum. Proc. Natl. Acad. Sci. USA* **95,** 4425–4430.

Rieseberg, L. H. (2001) Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16,** 351–358.

Rieseberg, L. H. & Livingstone, K. (2003) Evolution—Chromosomal speciation in primates. *Science* **300,** 267–268.

Rieseberg, L. H., Vanfossen, C. & Desrochers, A. M. (1995) Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* **375,** 313–316.

Spirito, F. (2000) The role of chromosomal change in speciation. In *Endless Forms*, eds. Howard, D. J. & Berlocher, S. H. (Oxford Univ. Press, New York), pp. 320–329.

Temu, E. A. (1997) Detection of hybrids in natural populations of the *Anopheles gambiae* complex by the rDNA-based, PCR method. *Ann. Trop. Med. Parasitol.* **91,** 963–965.

Tishkoff, S. A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J., *et al.* (2001) Haplotype diversity and linkage disequilibrium at human G6PD: Recent origin of alleles that confer malarial resistance. *Science* **293,** 455–462.

Touré, Y. T., Petrarca, V., Traore, S. F., Coulibaly, A., Maiga, H. M., Sankare, O., Sow, M., DiDecco, M. A. & Coluzzi, M. (1998) The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia* (*Rome*) **40,** 477–511.

von Koenigswald, G. H. R. (1982) Distribution and evolution of the orang utan, *Pongo pygmaeus* (Hoppius). In *The Orang Utan: Its Biology and Conservation*, ed. de Boer, L. E. M. (W. Junk, The Hague, The Netherlands), pp. 1–15.

Wang, R. L., Wakeley, J. & Hey, J. (1997) Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* **147,** 1091–1106.

White, M. J. D. (1968) Models of speciation. *Science* **159,** 1065–1070.

White, M. J. D. (1978) *Modes of Speciation* (Freeman, San Francisco).

Willis, K. J., Gillson, L. & Brncic, T. M. (2004) How "virgin" is virgin rainforest? *Science* **304,** 402–403.

Yunis, J. J. & Prakash, O. (1982) The origin of man—A chromosomal pictorial legacy. *Science* **215,** 1525–1530.

Zhang, J., Wang, X. & Podlaha, O. (2004) Testing the chromosomal speciation hypothesis for humans and chimpanzees. *Genome Res.* **14,** 845–851.

5

Developmental Plasticity and the Origin of Species Differences

MARY JANE WEST-EBERHARD*

Speciation is the origin of reproductive isolation and divergence between populations, according to the "biological species concept" of Mayr. Studies of reproductive isolation have dominated research on speciation, leaving the origin of species differences relatively poorly understood. Here, I argue that the origin of species differences, and of novel phenotypes in general, involves the reorganization of ancestral phenotypes (developmental recombination) followed by the genetic accommodation of change. Because selection acts on phenotypes, not directly on genotypes or genes, novel traits can originate by environmental induction as well as mutation, then undergo selection and genetic accommodation fueled by standing genetic variation or by subsequent mutation and genetic recombination. Insofar as phenotypic novelties arise from adaptive developmental plasticity, they are not "random" variants, because their initial form reflects adaptive responses with an evolutionary history, even though they are initiated by mutations or novel environmental factors that are random with respect to (future) adaptation. Change in trait frequency involves genetic accommodation of the threshold or liability for expression of a novel trait, a process that follows rather than directs phenotypic change. Contrary to common belief, environmentally initiated novelties may have greater evolutionary potential than muta-

*Smithsonian Tropical Research Institute, Unit 2511, APO AA 34020-9511. Present address: Escuela de Biología, Universidad de Costa Rica, Ciudad Universitaria, Costa Rica.

tionally induced ones. Thus, genes are probably more often followers than leaders in evolutionary change. Species differences can originate before reproductive isolation and contribute to the process of speciation itself. Therefore, the genetics of speciation can profit from studies of changes in gene expression as well as changes in gene frequency and genetic isolation.

The evolution of reproductive isolation is a defining characteristic of speciation. Reproductive isolation contributes to the diversification of species by creating genetically independent lineages, the branches of a phylogenetic tree. Each branching point of the tree of life is a speciation event. However, reproductive isolation alone does not create a new branch, because by itself it cannot produce the phenotypic divergence represented by the angular departure of a branch from the ancestral form. In the book celebrated by this colloquium, *Systematics and the Origin of Species* (Mayr, 1942), Ernst Mayr called phenotypic divergence between populations "the other aspect of speciation." Mayr wrote that speciation has two parts: "One part . . . is the establishment of discontinuities," or reproductive isolation. "The other aspect is the establishment of diversity and divergence, that is the origin of new characters. . ." (Mayr, 1942, p. 23). The origin of species differences, not reproductive isolation, were the main focus of Darwin's book *On the Origin of Species by Means of Natural Selection* (Darwin, 1858).

This second aspect of speciation, the origin of new characters, is the subject I address here. In particular, I will pursue Mayr's suggestion that "the workings of this process," the origin of new characters or novel phenotypic traits, "can best be studied if we analyze variation" (Mayr, 1942, p. 23). I will take a close look at the origins of variation, starting with two simple questions. (*i*) Where does the variation, or the variant that makes a new trait, come from? (*ii*) What gets this second, divergence part of speciation, the origin of species differences, started?

THE NATURE OF SELECTION AND SELECTABLE VARIATION

The evolutionary synthesis of the mid-20th century, sometimes called the "Neo-Darwinian Synthesis," has been characterized as a synthesis of Darwinism and genetics, with genetic mutation seen as the source of new selectable variation. "The 'genetical theory of natural selection,' the theory that evolution proceeds by natural selection of 'random' mutations, . . . is the basis of the 'neo-Darwinian synthesis'" (Leigh, 1987, p. 187). Consistent with this theory, natural selection, or fitness differences (differential reproductive success), is sometimes defined in terms of genotypes rather than phenotypes (e.g., Orr, 2005; see also review in West-Eberhard, 2003,

Chapter 1). However, the synthesis was not a monolithic affair. Mayr always insisted that the individual phenotype, not the genotype or the gene, is the object of selection (Mayr, 1942, 1963, 2004).

Although the genetic emphasis has been widely adopted, it is an approach that creates problems for understanding the origins of novel traits. The root of the problems is a concept of selection that, mistakenly, requires genetic variation. If selection requires genetic variation, then novel selectable variation must be genetic in nature; hence, mutation is seen as the primary source of evolutionary novelties. However, Darwinian evolution, the origin and evolution of phenotypic traits by natural selection, cannot possibly proceed by natural selection acting directly on mutations or genes. Except for the alleles that carry out their competitive battles within the germ cells for access to the germ line, in processes like meiotic drive, natural selection does not concern reproduction by genes themselves. Most genes under selection depend for their differential propagation on the differential reproduction of the bodies that contain them. That is, genes can replicate themselves, but only within organisms. To spread within populations, they depend on their ability to affect the reproduction of their bearers; they depend on their effects on phenotypes. Therefore, selection should be seen as acting on phenotypes (Mayr, 1963), and selectable variation means phenotypic variation, whether it has a genetic component or not. It is adaptive evolution, or a genetic *response* to selection, that requires genetic variation among the selected entities, not selection (differential reproductive success) itself. The question of a genetic response to selection on environmentally induced traits is discussed below.

Are these trivial, or merely semantic, matters of definition? Different definitions of selection imply different conclusions regarding fundamental issues, such as identification of the units of selection, the importance of development, and how adaptive evolution works. If it is the developmentally organized and environmentally sensitive phenotype that is the object of selection, as argued here, then certain facts that are "surprising" under the conventional mutation-selection idea of adaptive evolution (Orr, 2005, p. 119) are easily understood and expected (see West-Eberhard, 2003; also see below). Examples include the occurrence of extensive morphological evolution with only a modest number of genetic changes and small genetic changes that have a large effect on the phenotype or fitness (Mayr, 2004). If it is the phenotype, not the genotype, that is the object of selection, then selection can proceed for generations without genetic variation and without an evolutionary effect, as long as there is developmentally significant environmental variation. Then, should genetic variation affecting these traits arise, e.g., due to mutation or genetic recombination, it would immediately have an evolutionary effect. Selection on the pheno-

type means that directional selection can persist over longer time scales than predicted by concepts that see selection as requiring genetic variation. Traits that fail to respond to selection on a short time scale, due to paucity or depletion of genetic variation (e.g., under artificial selection), may undergo evolutionary change over long time scales in nature.

Most important for studies of species diversification, the phenotypic definition of selection permits a more complete analysis of the origins of new traits. If selectable variation is seen to be phenotypic variation, then the scope for the origins of novelty has to be broadened to include environmentally induced phenotypic variation. Phenotype development, which responds to both genomic and environmental inputs, is the source of selectable variation. This analysis brings development, largely omitted from evolutionary biology during the synthesis era (Hamburger, 1980), to the forefront of evolutionary biology as the source of the variation that fuels natural selection and adaptive evolution.

In this discussion, I look beyond mutation to seek the origins of selectable variation in the developmental plasticity of organisms (for a more extensive discussion, see West-Eberhard, 2003). I argue that the origin of species differences can be explained, and the synthesis of Darwinism with genetics can be improved, by invoking two concepts: developmental recombination and genetic accommodation. Developmental recombination, or developmental reorganization of the ancestral phenotype (West-Eberhard, 2003), explains where new variants come from: they come from the preexisting phenotype, which is developmentally plastic and therefore subject to reorganization to produce novel variants when stimulated to do so by new inputs from the genome or the environment. Genetic accommodation, or genetic change in the regulation or form of a novel trait (West-Eberhard, 2003), is the process by which new developmental variants become established within populations and species because of genetic evolution by selection on phenotypic variation when it has a genetic component.

Here, I use a broad concept of selection that encompasses both natural and sexual or social selection. I will not extensively discuss these different contexts of selection that are important in driving speciation-related divergence (e.g., see Etges and Noor, 2002; Schluter, 2000; West-Eberhard, 2003). Instead, I examine the very beginnings of traits and ask how they get started in populations or species.

THE ORIGIN OF DIVERGENCE: SEQUENCE OF EVENTS

A large body of evidence (West-Eberhard, 2003) indicates that regardless of selective context the origin of species differences under natural selection occurs as follows:

1. The origin of a new direction of adaptive evolution starts with a population of variably responsive, developmentally plastic organisms. That is, before the advent of a novel trait, there is a population of individuals that are already variable, and differentially responsive, or capable of producing phenotypic variants under the influence of new inputs from the genome and the environment. Variability in responsiveness is due partly to genetic variation and partly to variations in the developmental plasticity of phenotype structure, physiology, and behavior that arise during development and may be influenced by environmental factors, including maternal effects that reflect genetic and environmental variation present in previous generations. Genetic variation and developmental plasticity are fundamental properties of all living things: all individual organisms, with the exception of mutation-free clones, have distinctive genomes, and all of them have phenotypes that respond to genomic and environmental inputs. By "responsiveness" and "developmental plasticity," I do not mean just phenotypic plasticity in the way that term is usually used, to mean only responsiveness to the external environment. Rather, I include responsiveness to the action of genes, which may modify the internal environment of other genes and phenotypic elements within cells, with effects that extend outward to higher levels of organization and responsiveness. Any new input, whether it comes from the genome, like a mutation, or from the external environment, like a temperature change, a pathogen, or a parental opinion, has a developmental effect only if the preexisting phenotype is responsive to it. Without developmental plasticity, the bare genes and the impositions of the environment would have no effect and no importance for evolution.

2. Developmental recombination occurs in a population of individuals because of a new, or newly recurrent, input. A new input from the genome, such as a positively selected mutation, or from the environment of the affected individuals, causes a reorganization of the phenotype, or "developmental recombination." Given the variable developmental plasticity of different individuals, this process produces a population of novel variable phenotypes, providing material for selection.

3. Genetic accommodation may follow. If the resultant phenotypic variation has a fitness effect, that is, it correlates with the survival or reproductive success of the affected individuals, then selection (differential reproduction of individuals or other reproducing entities with different phenotypes) occurs. If the phenotypic variation has a genetic component, selection leads to "genetic accommodation," that is, adaptive evolution that involves gene-frequency change. Genetic accommodation of regulation adjusts the frequency, timing, and circumstances of the novel response (e.g., by adjusting the threshold for its expression), and genetic accommodation of form refines the characteristics and efficiency of the newly expressed trait.

This view of adaptive evolution is conventional in depicting adaptive evolution as phenotypic change that involves gene frequency change under selection. It departs only slightly, but importantly, from the mutation-selection version of adaptive evolution: although novelties may be induced by mutation they need not be; novelties may be induced by environmental factors. In either case, the genetic accommodation of novelty need not await mutation as long as there is a standing pool of genetic variation. As I discuss below, such variation is likely to be sufficient to support a response to selection on virtually any novel trait.

DEVELOPMENTAL RECOMBINATION

In developmental recombination, phenotypic traits are expressed in new or distinctive combinations during ontogeny, or undergo correlated quantitative change in dimensions. In the most easily visualized examples, elements of the phenotype controlled by switches are turned off or on in novel combinations. In tropical vines of the genus *Monstera*, for example, a single individual, e.g., of *Monstera dubia*, produces several sequential leaf forms during its ontogeny, and the leaf forms observed in the genus occur in different sequences and combinations in different species, with some species producing several and others only one (Madison, 1977; West-Eberhard, 2003). That is, the leaf forms have been developmentally duplicated, deleted, and recombined in a multitude of ways during the evolution of the genus *Monstera*, giving rise to a variety of species-specific ontogenies. This example illustrates how switch mechanisms can participate in the origins of novelty by recombination of ancestral phenotypic traits. Developmental switches, including decision points in behavior, physiology, and morphology, contribute to the modularity and dissociability of the phenotypic subunits we call "traits" (West-Eberhard, 2003), and this modularity, to the degree that it is not constrained by pleiotropic effects of component elements, may allow traits to be expressed in different combinations during development and evolution.

A common kind of developmental recombination is cross-sexual transfer, or the transfer of trait expression from one sex to the other (Iltis, 1983; West-Eberhard, 2003; Woodroffe and Vincent, 1994; Wynne-Edwards and Reburn, 2000). The hypothesis that cross-sexual transfer has produced the origin of a novel phenotype is particularly subject to tests because the hormonal mechanisms responsible often can be experimentally manipulated within the same or closely related species. Hormonal correlates of cross-sexual transfer have been studied in some species of birds and mammals where both males and females express parental care (reviews in West-Eberhard, 2003; Wynne-Edwards and Reburn,

2000). In birds, there appears to be a testosteronemediated tradeoff between male investment in aggressiveness and parental behavior: male parental care is associated with relatively low testosterone levels (Ketterson and Nolan, 1992), and in role-reversed species such as sandpipers, where males incubate the eggs, incubation is accompanied by a sharp increase in prolactin (Oring *et al.*, 1986a,b), as occurs in incubating females (Beach, 1961). In California mice (*Peromyscus californicus*), in which males show all of the parental behavior shown by mothers except lactation, male prolactin levels are similar to those of females soon after parturition (Gubernick and Nelson, 1989). Similarly, as in birds, male dwarf hamsters (*Phodopus* sp.) of species showing parental care have elevated prolactin and reduced testosterone during the lactation period, whereas congeneric species in which males do not express male parental care do not show such hormonal changes (Reburn and Wynne-Edwards, 1999). A key group for research on evolution by cross-sexual transfer in mammals is the voles (*Microtus* species), where there is both intraspecific and interspecific variation in the expression of parental care by the two sexes in a variety of ecological and social circumstances (reviewed in West-Eberhard, 2003).

Cross-sexual transfer also occurs in plants. It has long been recognized as the basis for the origin of the maize ear, which involved the feminization of the male flower, or tassel, of wild teosinte, the ancestor of maize (reviewed in West-Eberhard, 2003). The lateral branches of teosinte were shortened, and the terminal male inflorescence became (or was replaced by) the female ear of maize. This example is especially instructive: the major differences between teosinte and maize are largely explained by the influence of only four or five genetic loci (reviewed in West-Eberhard, 2003, p. 267; see also Doebley *et al.*, 1995). Leading researchers on the evolution of maize hypothesize that the transition began with environmental induction of branch shortening, followed by genetic assimilation of the short-branched morphology (Doebley *et al.*, 1995; Iltis, 1983). A "catastrophic sexual transmutation" hypothesis (Iltis, 1983) drew attention to the contribution of developmental sources of the origin of the distinctive maize phenotype. However, it is impossible to tell, from current information, whether the loci now identified as involved in the change are the result of genetic accommodation based on alleles at low frequency in natural populations of teosinte or are products of later mutations, whose phenotypic effects may have been modified (amplified or reduced) by genetic accommodation. Not all novel elements of the maize phenotype originated simultaneously; there is evidence that increased softness of the glumes originated in Panama (Piperno and Pearsall, 1998) rather than in the Balsas valley of Mexico that was the cradle of maize evolution (reviewed in West-Eberhard, 2003).

A different kind of developmental recombination is represented by the famous two-legged goat described by the Dutch morphologist Slijper (1942), in which a correlated shift in morphology and behavior accommodated an induced abnormality, leading to the well coordinated production of a complex and individually advantageous adjustment, producing a novel phenotype with little or no genetic change (Rachootin and Thomson, 1981). Slijper's two-legged goat was born with a congenital defect of the front legs so that it could not walk on all fours, and so it learned to walk and run by using its hind legs alone. Then, when it died an accidental death, Slijper dissected it and documented remarkable changes in muscle and bone, including striking changes in the bones of the hind legs; the leg muscles, including a greatly thickened and elongated gluteal tongue and an innovative arrangement of small tendons, a modified shape of the thoracic skeleton, and extensive modifications of the pelvis (West-Eberhard, 2003, p. 53).

It is not known whether the bipedal goat's abnormal front legs were due to a genetic or an environmentally induced defect, but in either case the inducer acted as a novel switch mechanism that in effect controlled the expression of a whole suite of correlated and adaptive changes in behavior, muscle, and bone. Even though the event that caused these changes was random with respect to adaptation, the phenotypic result was not a random variant. Rather, it was an adaptive accommodation of a random input, the result of pushing to extremes developmental plasticity in behavior, muscle, and bone that had already been subjected to a long history of selection and adaptive evolution. Similar effects on behavior and morphology are quite common in quadripedal mammals, including primates, forced or trained to walk upright (West-Eberhard, 2003, p. 42, Figure 3.12 on bipedal baboon; Hirasaki *et al.*, 2004; see also descriptions of a bipedal macaque in Waldman, 2004, and a bipedal dog in KFOR, 2003). These observations raise the possibility that the two-legged-goat effect, or "phenotypic accommodation" (West-Eberhard, 2003, 2005), has played a role in the evolution of bipedal locomotion in vertebrates, including humans, as suggested by Slijper, who noted that some of the novel morphological features of the two-legged goat resembled those of kangaroos and of other bipedal species such as orangutans (Slijper, 1942). Japanese macaques experimentally taught to walk upright develop humanlike gait characteristics (Hirasaki *et al.*, 2004), suggesting that the evolution of bipedalism in humans might not be as difficult or as large an evolutionary step as some anthropologists have believed. The distinctive anatomical features of humans compared with other primates that are associated with bipedal running include changes in muscle mass, tendon length, and thorax and pelvis shape (Bramble and Lieberman, 2004), the same features that underwent striking alterations in the bipedal goat

(Slijper, 1942). It is highly likely that developmental plasticity contributed to the species-specific morphological changes associated with the evolution of human bipedal walking and running.

Developmental recombination that can result in evolutionary divergence can occur at all levels of organization. At the molecular level, the modular structure of protein molecules, which parallels the modular organization of switch-controlled phenotypic development, facilitates reorganization. Proteins are composed of domains associated with the exon (expressed) regions of DNA that produce them. Some domains are known to be associated with particular biochemical functions, in a manner that parallels the functional and structural modularity of other aspects of the phenotype (West-Eberhard, 2003). The fibronectin family of proteins is a good example of a set of related proteins, where the nine domains or subunits that compose them are duplicated and organized in different combinations to form different molecules with distinctive functions (Holland and Blake, 1990).

DEVELOPMENTAL RECOMBINATION AND PARALLEL SPECIES PAIRS

Recurrent phenotypes, similar or identical phenotypic traits with discontinuous phylogenetic distributions, are quite common in a wide diversity of taxa (West-Eberhard, 2003). Their similarity is sometimes attributed to parallel evolution, the independent origin of phenotypic similarity due to selection and adaptative change in similar environmental conditions. Interpretation of the origin of species differences in stickleback offers a good example. In three-spined stickleback (*Gasterosteus aculeatus*), a complex of closely related species, there is a recurrent set of species pairs, with one member of the pair a slender, large-eyed "limnetic" form, which typically lives in the water column and feeds on plankton, and the other a stockier, smaller-eyed "benthic" form, typically a bottom dweller. All of these parallel species pairs are derived from anadromous ancestors, which migrated between the sea and the freshwater of rivers and lakes.

According to the way biologists usually think about the evolution of similarity, the recurrence of species pairs like these are the result of parallel evolution by natural selection on reproductively isolated breeding populations in parallel environments (Schluter and Nagel, 1995). By that interpretation, the anadromous ancestor gave rise repeatedly to the parallel forms by recurrent independent adaptation to parallel ecological conditions within lakes. The benthic and limnetic body types are invented over and over again independently, due to natural selection in a pair of ecological conditions that are common in lakes.

Developmental recombination offers an alternative explanation for

parallelism in stickleback (Foster and Baker, 2004; West-Eberhard, 2003). The recurrence of this pair of forms suggests that there may be something about the development of their common ancestor that enables it to give rise to these two particular forms readily, by means of altered expression of ancestral traits. Research on the ontogeny of anadromous stickleback (Andrews, 1999) has revealed that individuals are limnetic when young. They have a slender body form and live in the water column, where they feed on plankton like a limnetic species. Older individuals look and behave like the benthic form, with a stockier body and bottom-feeding habits. So the ancestral population occupies both of the habitats observed in the descendent species pairs and exhibits both phenotypes at different times during its life cycle, a pattern that suggests that the different recurrent forms may have originated not by parallel evolution but by altered timing (heterochrony) in the expression of previously evolved adaptive traits (West-Eberhard, 2003). By the heterochrony hypothesis, the limnetic-form species have juvenilized adults, and the benthic-form have full-sized adults like those of the ancestral form.

The morphology and behavior of stickleback are highly plastic, with feeding behavior influenced by learning (Hart and Gill, 1994). Feedback between morphology, food, and habitat choice, and learned feeding specializations would speed divergence between limnetic and benthic forms. When individuals of a sympatric benthic and limnetic pair of species were reared on each others' diets, their morphologies changed toward increased resemblance to the species whose diet they had experienced (Day *et al.*, 1994).

This discussion is not to deny the importance of natural selection in contrasting habitats for the evolution of species differences and their possible effects on the origin of reproductive isolation (Schluter, 2000; also see below). Rather, it is to urge a deeper look at the question of how such differences originate in lineages with a developmentally and adaptively versatile ancestral phenotype. Obviously, selection in their respective habitats would play an important role in the fixation and elaboration of the divergent descendent stickleback forms, but their parallel morphologies and behaviors may have originated before speciation. The plasticity hypothesis is supported by the fact that in all of the fish genera with replicate speciation between limnetic and benthic species pairs, the limnetic–benthic interspecific alternatives occur as intraspecific alternative phenotypes as well, in at least some populations, suggesting that they, too, could have originated as intraspecific developmental variants. These genera include, in addition to stickleback, lampreys (*Lampetra*), arctic charr (*Salvelinus*), Pacific salmon (*Oncorhynchus*), lake whitefish (*Coregonus*), trout (*Salmo*), and smelt (*Osmerus*) (reviewed in West-Eberhard, 2003).

Replicate speciation has produced three parallel species pairs of 13- and 17-year periodical cicadas (*Magicicada* species), with one of the 17-

year populations (*M. septendecim*) giving rise to a 13-year species (*M. neotredecim*), forming a species pair (*septendecim–neotredecim*) within a species pair (*tredecim–septendecim*) (Marshall and Cooley, 2000). The recurrent phylogenetically abrupt switches between these two life cycles suggests that, as in the replicate species pairs of fishes, some ancestral developmental mechanism with a 4-year periodicity has repeatedly influenced life-cycle divergence between the species pairs of periodical cicadas (Grant, 2005; West-Eberhard, 2003), although the physiological mechanism is poorly understood.

There are a great many examples of phylogenetic recurrence in a wide diversity of animals and plants (West-Eberhard, 2003). Phylogenetically separate, recurrent phenotypes show that a common type of developmental recombination is the reexpression of phenotypes that have been lost because of developmental deletion, alteration of a regulatory mechanism without extensive alteration of other aspects of the developmental capacity to produce the lost form. As a rule of thumb, recurrent parallel forms suggest ancestral developmental plasticity for producing both forms. This phenomenon may explain many of the parallelisms and homoplasies that are so commonly discovered in systematics and phylogenetics, but it is a hypothesis that needs to be tested case by case using detailed comparative studies of phenotypic variation and its developmental basis.

GENE-EXPRESSION CONSEQUENCES OF DEVELOPMENTAL RECOMBINATION

Individual development can be visualized as a series of branching pathways. Each branch point is a developmental decision, or switch point, governed by some regulatory apparatus, and each switch point defines a modular trait. Developmental recombination implies the origin or deletion of a branch and a new or lost modular trait. It is important to realize that the novel regulatory response and the novel trait originate simultaneously. Their origins are, in fact, inseparable events: you cannot have a change in the phenotype, a novel phenotypic state, without an altered developmental pathway.

In terms of gene expression, developmental recombination means that a set of ancestral genes are now coexpressed, or their products used, in a new combination or a new context, and the ancestral regulatory mechanisms are now triggered by a new inducer or an old one in some new sequence or environmental context.

Although the subject of developmental regulation may bring to mind "regulatory genes" or "master control genes," such as hox genes, regulatory mechanisms controlling the expression of many of the traits I have discussed here must often be polygenic in nature, because they may in-

volve complex aspects of the phenotype, including neural and sensory equipment, hormone systems, and complexly responsive tissues and organs. Polygenic complexity of regulation increases the likelihood that there will be genetic variation in the ability to respond and, thus, that genetic accommodation will occur. Similarly, there is likely to be polygenic variation in the dimensions and subcomponents of a trait that are newly expressed together because of the developmental recombination of ancestral traits. It is these two sets of coexpressed genes that are exposed as sets to selection: those that modify the regulation of expression of a trait and those that are activated as a set, or whose products are used together as a set, to affect the form of a trait.

GENETIC ACCOMMODATION

Genetic accommodation is simply quantitative genetic change in the frequency of genes that affect the regulation or form of a new trait. If a novelty has been induced by a mutation, then the mutation would be, at least initially, a gene of major effect on a polygenic regulatory mechanism. If regulation is polygenic, as just described, then the effect of that mutant gene would be subject to genetic modification involving multiple loci.

The sensitivity of a regulatory mechanism can be adjusted either up or down. If a novel trait is favored by selection, then genetic accommodation is expected to lower the threshold for its production or increase liability to pass the threshold, and genetic assimilation or fixation of the trait may occur. If the trait is selected against, then genetic accommodation would raise the threshold or reduce the liability for its expression until it is not expressed at all. Still another possibility is that a novel trait may persist indefinitely as an adaptive alternative phenotype. The important point here is that the genetic accommodation of regulation can determine the frequency of the trait in a population. Frequency of expression does not depend on the frequency of the inducer (mutation or environmental factor) alone.

For these reasons I consider genes followers, not leaders, in adaptive evolution. A very large body of evidence (West-Eberhard, 2003) shows that phenotypic novelty is largely reorganizational rather than a product of innovative genes. Even if reorganization was initiated by a mutation, a gene of major effect on regulation, selection would lead to genetic accommodation, that is, genetic change that follows, and is directed by, the reorganized condition of the phenotype. Some authors have expressed this pattern as "phenotype precedes genotype" (Palmer, 2004). This description applies best to genetic assimilation, a special case of genetic accommodation that begins with environmental induction and proceeds toward fixation of the novel trait (Waddington, 1953).

Of course, genetic mutation must ultimately fuel the genetic variation that permits adaptive evolution. However, genetic accommodation need not await mutation. There are two kinds of evidence that the standing genetic variation (Orr and Betancourt, 2001) of particular responses is probably usually sufficient to support genetic accommodation without mutation. A large accumulation of data on protein polymorphisms has shown that genetic variation is common in natural populations (Lewontin, 1974); and virtually every trait subjected to artificial selection shows a response to selection (references in West-Eberhard, 2003). A rare class of exceptions occurs under artificial selection for directional (consistently right or left) asymmetry of various traits (ommatidia number, wing folding, eye size, and thoracic bristle number) in *Drosophila* (Coyne, 1987; Maynard Smith and Sondhi, 1960; Purnell and Thompson, 1973; Tuinstra *et al.*, 1990). Nonetheless, directional (consistent right or left) asymmetry has evolved repeatedly in insects (e.g., see Schuh and Slater, 1995, on genitalia and other abdominal structures) and in other organisms (Palmer, 2004), suggesting that even this category of lack of response to selection can be overcome by variation and selection during longer evolutionary time scales. The phenotypic definition of selection helps to explain some of these cases, in which lack of genetic variation initially may have blocked a response to positive selection, and yet the trait eventually evolves to fixation. Because selection (differential reproductive success) of phenotypes does not require genetic variation, directional selection can persist generation after generation, favoring either the right or left form under environmentally influenced fluctuating asymmetry ("antisymmetry"), a state known to precede the evolution of some examples of directional symmetry (Palmer, 2004). Then an evolutionary response to selection would occur as soon as favorable genetic variations arise, e.g., due to mutation. Thus, although standing genetic variation usually must be sufficient to produce a response to selection (Orr and Betancourt, 2001), genetic accommodation may in some cases, like that of directional asymmetry, await mutation (Palmer, 2004).

EVOLUTIONARY POTENTIAL OF ENVIRONMENTALLY INDUCED CHANGE

Biologists are inclined to doubt the evolutionary importance of environmentally induced traits because it is not immediately obvious how they can be inherited in subsequent generations. Initiation by mutation is intuitively more appealing because it solves the problem of novelty and heritability in one stroke. However, environmentally induced variants are heritable as well, insofar as the ability to respond by producing them is heritable (that is, genetically variable). The responsiveness of organ-

isms to environmental influence involves mechanisms that are likely to be genetically complex and therefore subject to genetic variation at multiple loci, as just discussed.

A mutant gene may seem more dependable, that is, more likely to persist across generations, than a novel environmental factor. However, it takes only a little reflection on the nature of development to appreciate the dependability of environmental factors. All organisms depend on the cross-generational presence of large numbers of highly specific environmental inputs: particular foods, vitamins, hosts, symbionts, parental behaviors, and specific regimes of temperature, humidity, oxygen, or light. Such environmental elements are as essential and as dependably present as are particular genes; some, such as photoperiod and atmospheric elements like oxygen and carbon dioxide, are more dependably present than any gene in particular habitats and zones, so we forget that these environmental factors constitute powerful inducers and essential raw materials whose geographically variable states can induce developmental novelties as populations colonize new areas.

Some environmental elements act as developmental building blocks and signals quite comparable to the products of genes. For example, DNA microarray studies have shown that environmentally supplied bacteria in the digestive tract of zebrafish regulate the expression of 212 genes. Without these bacteria, many of which produce highly specific host responses, the developing fish die (Rawls *et al.*, 2004). Given the evidence, familiar to everyone, that numerous environmental inputs are consistently supplied (essential) during normal development, the skepticism of biologists regarding the reliability of environmental factors relative to that of genes has to rank among the oddest blind spots of biological thought.[†]

Contrary to the notion that mutational novelties have superior evolutionary potential, there are strong arguments for the greater evolutionary potential of environmentally induced novelties. An environmental factor can affect numerous individuals at once, whereas a mutation initially can affect only one. The larger the population affected, the greater the likelihood that an environmentally induced novelty occurs in a favorable genetic, phenotypic, or selectively advantageous environment in at least some subpopulation of the individuals affected, and the larger the probability of genetic variation that can result in an evolutionary response to selection. Furthermore, environmentally induced phenotypes can persist over generations even when disadvantageous because, in contrast to mu-

[†]The assumption that genetic inputs are more reliable than environmental ones may be further challenged by future research on gene expression. Recent findings suggest that processes like gene transcription, involving small numbers of molecules, are especially subject to randomness and noise (Raser and O'Shea, 2004).

tations, they cannot readily be eliminated by selection. The superior persistence of environmentally induced traits allows time for genetic accommodation and adaptive evolutionary change. For example, undersized and "starvation" forms are exceedingly common in nature, even though disadvantageous, because developing individuals often cannot escape environmental variation in food supply. As a result, evolved specializations to small size, such as nonfighting morphologies and behaviors of small males (West-Eberhard, 2003), and other striking size-associated adaptations (Schmidt-Nielsen, 1984) are common in nature.

Population-wide environmental induction and genetic accommodation sometimes occur when populations colonize islands, where new environmental stimuli and opportunities repeatedly induce novel phenotypes, such as learned foraging techniques, which then subject the population to selection on associated morphology, behaviors, and diet-associated physiology (Price *et al.*, 2003; West-Eberhard, 2003) and when habitat change forces dietary change in ingested carotenoids, with effects on the plumage colors and associated evolved biochemistry of birds (Price *et al.*, 2003).

THE ORIGIN OF SPECIES DIFFERENCES: BEFORE OR AFTER REPRODUCTIVE ISOLATION?

Trait origin by developmental recombination predicts several properties of species and their genetics as follows:

(*i*) In evolution by developmental recombination, the same genes are used over and over in different contexts and combinations. This process should contribute to the maintenance of the same or homologous genes over long periods of evolutionary time and would help account for finding similar genes in distantly related species or the observed conservation of genes across long periods of evolutionary change;

(*ii*) Small genetic distances between species with strikingly different phenotypes are expected because extensively reorganized new phenotypes can occur with little genetic change;

(*iii*) Homoplasy and parallelism are expected to be common among related species, because developmental flexibility can give rise repeatedly to the same kinds of variation; and

(*iv*) Phenotypic differences that eventually distinguish species may often arise before the advent of reproductive isolation between them, because the origin and maintenance of more than one developmental pathway can occur within a population; the evolution of a divergent novelty does not require gene-pool divergence, only developmental-pathway and gene-expression divergence (West-Eberhard, 2003).

This last point is important for students of systematics and speciation because it means that some phenotypic divergence assumed to mark species may in fact represent intraspecific alternative phenotypes or, in paleontology, morphotypes assumed to be species when they are in fact complex alternative forms that represent gene-expression, not genetic, differences between individuals. More importantly for the process of speciation, divergent developmental pathways within species enable the exploitation of different conditions and resources by members of the same species as adaptive options, and assortative mating by developmentally similar individuals can then contribute to speciation, whether in sympatry or in geographically isolated populations. In either setting, selection for a single alternative would speed the specialization of an increasingly monomorphic subpopulation, because an approach to phenotypic fixation is expected to accelerate the (genetic) evolution of the fixed form (Clarke, 1966; West-Eberhard, 2003). This acceleration would contribute to the evolution of reproductive isolation between populations with different alternative phenotypes, insofar as genetic divergence contributes to the likelihood of pre- or postzygotic reproductive incompatibility between them (for tests of this largely unexamined condition, see Funk *et al.*, 2002).

Consistent with the hypothesis that ancestral developmental alternatives can precede and contribute to speciation, there are many species differences that parallel differences between alternative phenotypes within closely related species (Schlichting, 2004; West-Eberhard, 1989, 2003). Multiple kinds of evidence support the hypothesis that species differences originated before reproductive isolation in a variety of organisms, including buttercups, butterflies, aphids, migrant fishes and birds, and socially parasitic ants (reviewed in West-Eberhard, 2003). Intraspecific divergence in host-specific behaviors and lifehistory characteristics has repeatedly been suggested as possibly contributing to speciation in apple maggot flies (*Rhagoletis*) (Carson, 1989; Walsh, 1864; West-Eberhard, 2003), but the possible contribution of preisolation developmental divergence in the form of a host-associated polyphenism or polymorphism has never been systematically investigated in that genus. Despite the experimental evidence for speciationrelated preisolation phenotypic divergence in host-switching phytophagous insects, given long ago by Walsh (1864) and mentioned by Bush (1994), speciation research has focused primarily on genetic divergence accompanying or following breeding isolation in *Rhagoletis* and other organisms.

Research like that of Schluter and associates (2000, 1995) shows how traits such as body size, which have diverged under natural selection, can contribute to the origin of reproductive isolation. Variation in body size is often associated with the evolution of condion-sensitive, faculatatively expressed alternative phenotypes within species, including in fishes and other

organisms, such as ants, where size differences have been implicated in the origins of reproductive isolation between contrasting phenotypes (summary in West-Eberhard, 2003). Such phenotypes, like those of geographic isolates, evolve under natural and sexual or social selection, but with their divergence originally mediated by developmental plasticity (West-Eberhard, 2003). Indeed, speciation-related divergence is arguably facilitated, that is, more rapid and readily divergent, within species than between geographic isolates (Etges and Noor, 2002; West-Eberhard, 1989, 2003). Therefore, hypotheses such as the "ecological speciation" hypothesis (Schluter, 2000), which seek to associate preisolation divergence under selection with the origin of reproductive isolation, whether in sympatry or allopatry, may find some of their best support in taxa containing marked intraspecific alternative phenotypes ("ecophenotypes," polyphenisms, polyethisms, etc.).

At present, it is impossible to evaluate the contribution of preisolation divergence by means of developmental plasticity to the evolution of reproductive isolation because genetic studies have been designed to detect the extent and breeding consequences of differences between populations, by means of studies of genetic distances, hybridization, etc. (e.g., see review in Coyne, 1987) rather than their possible sources in patterns of gene expression within populations. Preisolation divergence may turn out to be another "excellent but embarrassing example of not being able to find what you are not looking for" (Mann, 2004), because there is plenty of indirect evidence that developmental plasticity has been important for the origins of species differences. Not only are species differences often parallel to intraspecific alternative phenotypes, as already mentioned, but the rampant speciation and associated phenotypic diversification that characterize some of the most spectacular adaptive radiations known can be linked to particular kinds of developmental plasticity. The beak and trophic diversification of Darwin's finches in the Galapagos Islands has involved learned associations between beak size and shape and feeding preferences; the niche diversification of African lake cichlids is associated with dietary flexibility in mouth morphology and behavior; and the larval habitat diversification of Hawaiian *Drosophila* may have involved biochemical versatility within species (reviewed in West-Eberhard, 2003).

FUTURE OF GENETIC STUDIES OF SPECIATION

Inspired by Dobzhansky and the Darwinian evolutionary geneticists Fisher, Haldane, and Wright, as well as by Mayr's biological species concept (Mayr, 1963) and other concepts that also emphasize the role of genetic isolation in promoting genetic divergence (Coyne and Orr, 2004;

Drés and Mallet, 2002), research on species differences inspired by the synthesis has focused on the genetics of reproductive isolation between populations. This approach has produced many insights regarding the process of speciation, but it has created a kind of selective vision that may sometimes overlook the potential contribution of preisolation phenotypic divergence by means of developmental plasticity and its possible consequences for assortative mating and reproductive isolation. Ironically, this emphasis on genetic isolation may impede understanding of the causes of speciation because important genepool or genotypic-cluster (Coyne and Orr, 2004) differences may come to exist only when reproductive isolation is already well advanced. So geneticists may end up describing the results of speciation rather than its causes.

Lack of attention to developmental phenomena in relation to speciation promises to change, because genomic studies of speciation can now contemplate gene-expression as well as genefrequency data (e.g., Foster and Baker, 2004, on the stickleback model system and Michalak and Noor, 2003, on *Drosophila*). Research on patterns of gene expression makes it possible to pinpoint the (expressed) loci that are actually subject to selection in the evolution of species differences, beginning with differences that arise because of developmental recombination without reproductive isolation. Comparative genomics has the potential to illuminate the contribution of developmental-genetic processes to speciation. Are the alternative sets of genes that are differentially expressed in particular forms in an intraspecific polyphenism, polymorphism, or life-stage difference, like those observed in some host-specific insects or during the ontogeny of stickleback, the same sets of genes that characterize differences between recently derived host races or species whose phenotypes parallel those forms? Is there a burst of genetic change in the modifiers of form when a particular form in a polyphenic population approaches fixation, as predicted by theoretical models (West-Eberhard, 2003)? Finally, does the accompanying phenotypic change contribute to the evolution of reproductive isolation?

As genomic libraries expand and the associated techniques for research on gene expression become increasingly accessible to speciation biologists (e.g., see Drés and Mallet, 2002; Foster and Baker, 2004), the interests of geneticists will increasingly converge with those of organismic biologists interested mainly in phenotypes (Stearns and Magwene, 2003). The meeting ground of intermediate processes, such as hormone physiology, neurobiology, and subcellular signaling process (Larsen, 2004), topics long estranged from direct involvement in speciation studies and other areas of evolutionary biology, then will be of increasing interest, because these are the processes that link environmental, genetic, and phenotypic variation to selection and evolution through their mediation

of gene expression. Progress in understanding the developmental nature of variation supplies a missing piece of the synthesis begun by Dobzhansky and Mayr.

ACKNOWLEDGMENTS

I thank William Eberhard, Jody Hey, Mohamed Noor, John Skoyles, and an anonymous reviewer for helpful comments and references.

REFERENCES

Andrews, C. A. (1999) Ontogenetic Ecomorphology of the Threespine Stickleback (*Gasterosteus aculeatus*). Ph.D. thesis (State Univ. N.Y., Stony Brook, NY).

Beach, F. A. (1961) *Hormones and Behavior* (Cooper Square, New York).

Bramble, D. M. & Lieberman, D. E. (2004) Endurance running and the evolution of *Homo. Nature* **433,** 345–352.

Bush, G. L. (1994) Sympatric speciation in animals: New wine in old bottles. *Trends Ecol. Evol.* **9,** 285–288.

Carson, H. L. (1989) Sympatric pest. *Nature* **338,** 304.

Clarke, B. (1966) The evolution of morph-ratio clines. *Am. Nat.* **100,** 389–402.

Coyne, J. A. (1987) Lack of response to selection for directional asymmetry in *Drosophila melanogaster. J. Hered.* **78,** 119.

Coyne, J. A. & Orr, A. (2004) *Speciation* (Sinauer, Sunderland, MA).

Darwin, C. (1858) *On the Origin of Species by Means of Natural Selection* (Murray, London).

Day, T., Pritchard, J. & Schluter, D. (1994) Ecology and genetics of phenotypic plasticity: A comparison of two sticklebacks. *Evolution* **48,** 1723–1734.

Doebley, J., Stec, A. & Gustus, C. (1995) *Teosinte branched1* and the origin of maize: Evidence for epistasis and the evolution of dominance. *Genetics* **141,** 333–346.

Drés, M. & Mallet, J. (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philos. Trans. R. Soc. London B* **357,** 471–492.

Etges, W. J. & Noor, M. A. F., eds. (2002) *Genetics of Mate Choice: From Sexual Selection to Sexual Isolation* (Kluwer, London).

Foster, S. A. & Baker, J. A. (2004) Evolution in parallel: new insights from a classic system. *Trends Ecol. Evol.* **19,** 456–459.

Funk, D. J., Filchak, K. E. & Feder, J. L. (2002) Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetica,* **116,** 251–267.

Grant, P. R. (2005) The priming of periodical cicada life cycles. *Trends Ecol. Evol.* **20,** 169–174.

Gubernick, D. J. & Nelson, R. J. (1989) Prolactin and paternal behavior in the biparental California mouse, *Peromyscus californicus. Horm. Behav.* **23,** 203–210.

Hamburger, V. (1980) Embryology. In *The Evolutionary Synthesis,* eds. Mayr, E. & Provine, W. B. (Harvard Univ. Press, Cambridge, MA), pp. 96–112.

Hart, P. J. P. & Gill, A. B. (1994) Evolution of foraging behaviour in the threespine stickleback. In *The Evolutionary Biology of the Threespine Stickleback,* eds. Bell, M. A. & Foster, S. A. (Oxford Univ. Press, New York), pp. 207–239.

Hirasaki, E., Ogihara, N., Hamada, Y., Kumakura, H. & Nakatsukasa, M. (2004) Do highly trained monkeys walk like humans? A kinematic study of bipedal locomotion in bipedally trained Japanese macaques. *J. Hum. Evol.* **46,** 739–750.

Holland, S. K. & Blake, C. C. F. (1990) Proteins, exons, and molecular evolution. In *Intervening Sequences in Evolution and Development*, eds. Stone, E. M. & Schwartz, R. J. (Oxford Univ. Press, New York), pp. 10–42.

Iltis, H. (1983) From teosinte to maize: the catastrophic sexual transmutation. *Science* **222,** 886–895.

Ketterson, E. D. & Nolan, V., Jr. (1992) Hormones and life histories: An integrative approach. *Am. Nat.* **140,** S33–S62.

KFOR News. Available at www.kfor.com/global/story.asp?s=1333882. Accessed June 24, 2003.

Larsen, E. W. (2004) A view of phenotypic plasticity from molecules to morphogenesis. In *Environment, Development, and Evolution*, eds. Hall, B. K., Pearson, R. D. & Müller, G. B. (MIT Press, Cambridge, MA), pp. 117–123.

Leigh, E. G., Jr. (1987) Ronald Fisher and the development of evolutionary theory. II. Influences of new variation on evolutionary process. *Oxford Surv. Evol. Biol.* **4,** 212–263.

Lewontin, R. C. (1974) *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, New York).

Madison, M. (1977) A revision of *Monstera* (Araceae). In *Contributions from the Gray Herbarium* (Harvard Univ. Press, Cambridge, MA) Vol. 27, pp. 3–131.

Mann, C. C. (2004) Oldest civilization in the Americas revealed. *Science* **307,** 34–35.

Marshall, D. C. & Cooley, J. R. (2000) Reproductive character displacement and speciation in periodical cicadas, with description of a new species, 13-year *Magicicada neotredecim. Evolution* **54,** 1313–1335.

Maynard Smith, J. & Sondhi, K. C. (1960) The genetics of a pattern. *Genetics* **45,** 1039–1050.

Mayr, E. (1942) *Systematics and the Origin of Species* (Columbia Univ. Press, New York).

Mayr, E. (1963) *Animal Species and Evolution* (Harvard Univ. Press, Cambridge, MA).

Mayr, E. (2004) 80 years of watching the evolutionary scenery. *Science* **305,** 46–47.

Michalak, P. & Noor, M. A. (2003) Genome-wide patterns of expression in *Drosophila* pure species and hybrid males. *Mol. Biol. Evol.* **20,** 1070–1076.

Oring, L. W., Fivizzani, A. J. & El Halawani, M. E. (1986a) Changes in plasma prolactin associated with laying and hatch in the spotted sandpiper. *Auk* **103,** 820–822.

Oring, L. W., Fivizzani, A. J. & El Halawani, M. E. (1986b) Seasonal changes in prolactin and luteinizing hormone in the polyandrous spotted sandpiper, *Actitis macularia*. *Gen. Comp. Endocrinol.* **62,** 394–403.

Orr, H. A. (2005) The genetic theory of adaptation: A brief history. *Nat. Rev. Genet.* **6,** 119–127.

Orr, H. A. & Betancourt, A. J. (2001) Haldane's seive and evolution from the standing variation. *Genetics* **157,** 875–884.

Palmer, A. R. (2004) Symmetry breaking and the evolution of development. *Science* **306,** 828–833.

Piperno, D. R. & Pearsall, D. M. (1998) *The Origins of Agriculture in the Lowland Neotropics* (Academic, New York).

Price, T. D., Qvarnström, A. & Irwin, D. E. (2003) The role of phenotypic plasticity in driving genetic evolution. *Proc. R. Soc. London Ser. B*, **270,** 1433–1440.

Purnell, D. J. & Thompson, J. N. J. (1973) Selection for asymmetrical bias in a behavioral character of *Drosophila melanogaster*. *Heredity* **31,** 401–405.

Rachootin, S. P. & Thomson, K. S. (1981) Epigenetics, paleontology, and evolution. *Proc. Int. Congress Syst. Evol. Biol.* **2,** 181–193.

Raser, J. M. & O'Shea, E. K. (2004) Control of stochasticity in eukaryote expression. *Science* **304**, 1811–1814.

Rawls, J. F., Samuel, B. S. & Gordon, J. I. (2004) Gnotobiotic zebrafish reveal evolutionarily conserved responses to the gut microbiota. *Proc. Natl. Acad. Sci. USA* **101,** 4596–4601.

Reburn, C. J. & Wynne-Edwards, K. E. (1999) Hormonal changes in males of a naturally biparental and a uniparental mammal. *Horm. Behav.* **35,** 163–176.

Schlichting, C. D. (2004) The role of phenotypic plasticity in diversification. In *Phenotypic Plasticity: Functional and Conceptual Approaches*, eds. DeWitt, T. J. & Scheiner, S. M. (Oxford Univ. Press, Oxford), pp. 191–200.

Schluter, D. (2000) *The Ecology of Adaptive Radiation* (Oxford Univ. Press, New York).

Schluter, D. & Nagel, L. M. (1995) Parallel speciation by natural selection. *Am. Nat.* **146,** 292–301.

Schmidt-Nielsen, K. (1984) *Scaling: Why Is Animal Size So Important?* (Cambridge Univ. Press, Cambridge, U.K.).

Schuh, R. T. & Slater, J. Z. (1995) *True Bugs of the World (Hemiptera: Heteroptera)* (Cornell Univ. Press, Ithaca, NY).

Slijper, E. J. (1942) Biologic-anatomical investigations on the bipedal gait and upright posture in mammals, with special reference to a little goat, born without forelegs. *Proc. Koninklijke Nederlandse Akademie Van Wetenschappen* **45,** pp. 288–295, 407–415.

Stearns, S. C. & Magwene, P. (2003) The naturalist in a world of genomics. *Am. Nat.* **161,** 171–180.

Tuinstra, E. J., DeJong, G. & Scharloo, W. (1990) Lack of response to family selection for directional asymmetry in *Drosophila melanogaster*: Left and right are not distinguished in development. *Proc. R. Soc. London Ser. B* **231,** 146–152.

Waddington, C. H. (1953) Genetic assimilation of an acquired character. *Evolution* **7,** 118–126.

Waldman, D. *MSNBC News* (Associated Press). Available at www.msnbc.msn.com/id/5479501. Accessed July 21, 2004.

Walsh, B. D. (1864) On phytophagic varieties and phytophagic species. *Proc. Entomol. Soc. Philadelphia* **3,** 403–430.

West-Eberhard, M. J. (1989) Phenotypic plasticity and the origins of diversity. *Ann. Rev. Ecol. Syst.* **20,** 239–278.

West-Eberhard, M. J. (2003) *Developmental Plasticity and Evolution* (Oxford Univ. Press, New York).

West-Eberhard, M. J. (2005) Phenotypic accommodation: adaptive innovation due to developmental plasticity. *J. Exp. Zool. B Mol. Dev. Evol.*, in press.

Woodroffe, R. & Vincent, A. (1994) Mother's little helpers: patterns of male care in mammals. *Trends Ecol. Evol.* **9,** 294–297.

Wynne-Edwards, K. & Reburn, C. J. (2000) Behavioral endocrinology of mammalian fatherhood. *Trends Ecol. Evol.* **16,** 464–523.

Part II

DISCERNING RECENT DIVERGENCE

One of Mayr's achievements that is not always counted is that he made the difficult questions on evolutionary divergence seem accessible. By laying out clear scenarios for a seemingly intractable process in which evolutionary factors interact with the geographic circumstances of populations to cause divergence, he fueled the interest and enthusiasm of generations of evolutionary biologists. The modern fruits of this enthusiasm are the many studies that detail reconstruction of recent cases where evolution has given rise to new species. The progress, and rapid pace of progress, in this field is clearly shown in the paper by Scott Edwards *et al.,* "Speciation in Birds: Genes, Geography, and Sexual Selection" (Chapter 6), which outlines how present-day studies on speciation in birds are gaining the genetic and theoretical sophistication that had formerly only been associated with the model *Drosophila* systems. Topics such as the role of sexual selection and the frequency of sympatric speciation are now being addressed genetically in a number of avian systems.

Among the most intricate of speciation puzzles are those involving obligate mutualistic relationships. If one species of a mutualistic assemblage diverges into two, must the other species of the assemblage follow along? This is the question that arises for figs and their pollinating wasps, and it is addressed by Carlos Machado, Nancy Robbins, Tom Gilbert, and Allen Herre in "Critical Review of Host Specificity and Its Coevolutionary Implications in the Fig/Fig-Wasp Mutualism" (Chapter 7). Each of the 750 or so species of fig depends upon fig wasps for pollination; the

wasps, in turn, require the ovaries of the figs as oviposition sites (Wiebes, 1979). Strong reciprocal species specificity suggests that when individual fig species or individual wasp species undergo speciation, they do so in tandem with their mutualistic partner (cospeciation). But Machado *et al.*'s phylogenetic and population genetic study (Chapter 7) shows that the history has not been this straightforward and that host switches by wasps, and possibly species hybridization by figs, have created partly independent phylogenetic histories of figs and wasps.

In Mayr's world view, new species arise under allopatry, and, after that, as divergence accrues, the geographic ranges of related species may later come to overlap. In this way, related but divergent species may be sympatric, in contrast to most closely related species, which are expected to have disjunct, allopatric distributions. This sequence of events was outlined explicitly by Mayr in a 1954 paper on the biogeography of sea urchins (Mayr, 1954). Stephen Palumbi and Harilaos Lessios, in "Evolutionary Animation: How Do Molecular Phylogenies Compare to Mayr's Reconstruction of Speciation Patterns in the Sea?" (Chapter 8), have returned to this same Echinoid system and reconsidered Mayr's synthesis using DNA sequence data. They find that although the pattern described by Mayr still largely applies, rapidly evolving gamete recognition proteins play a strong role in reproductive isolation. In contrast, Mayr had envisioned the evolution of reproductive isolation by a more genome-wide steady accumulation of substitutions.

For many biologists, the question of whether geographic separation is strictly necessary for speciation (i.e., the question of whether sympatric or parapatric speciation occurs) comes into sharpest focus with the case of *Rhagoletis pomonella*. This is the apple maggot fly that has diverged into two host races (apple and hawthorne), apparently under geographic sympatry and aided by the different fruiting times of the two hosts (Filchak *et al.*, 2000). Mayr's former student Guy Bush discovered the history of sympatric divergence in *Rhagoletis*, and it has long been a standard component of the debates on the prevalence of sympatric speciation. Now we learn from Guy Bush's former student Jeffrey Feder and his colleagues, in "Mayr, Dobzhansky, and Bush and the Complexities of Sympatric Speciation in *Rhagoletis*" (Chapter 9), that the sympatric divergence that occurred within U.S. populations may have been facilitated by genetic variation that came in by means of gene flow from largely separated populations in Mexico.

The question of sympatric speciation has also been much discussed in the context of the highly speciose cichlid fishes from the great African lakes: Victoria, Malawi, and Tanganyika (Mayr, 1984). Particularly in the cases of Lakes Malawi and Victoria, which are relatively young, it is a wonder how hundreds of species could form within confined bodies of water within <1 million years. Yong-Jin Won, Arjun Sivasundar, Yong

Wang, and Jody Hey, in "On the Origin of Lake Malawi Cichlid Species: A Population Genetic Analysis of Divergence" (Chapter 10), take a close look at a group of rock-dwelling species from Lake Malawi. To gain resolution, they used a new type of genetic marker that includes a microsatellite and linked sequence and a new Bayesian method for fitting complex models of divergence (Hey *et al.*, 2004). The results suggest that some of these species have formed within the past few thousands years and that gene exchange is ongoing between species at some loci.

REFERENCES

Filchak, K. E., Roethele, J. B. & Feder, J. L. (2000) Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*. *Nature* **407,** 739–742.

Hey, J., Won, Y.-J., Sivasundar, A., Nielsen, R. & Markert, J. A. (2004) Using nuclear haplotypes with microsatellites to study gene flow between recently separated Cichlid species. *Mol. Ecol.* **13,** 909–919.

Mayr, E. (1954) Geographic speciation in tropical Echinoids. *Evolution* **8,** 1–18.

Mayr, E. (1984) Evolution of fish species flocks: A commentary. In *Evolution of Fish Species Flocks*, eds. Echelle, A. A. & Kornfield, I. (Univ. of Maine Press, Orono, ME), pp. 3–11.

Wiebes, J. T. (1979) Co-evolution of figs and their insect pollinators. *Annu. Rev. Ecol. Syst.* **10,** 1–12.

6

Speciation in Birds: Genes, Geography, and Sexual Selection

SCOTT V. EDWARDS,*§ SARAH B. KINGAN,*§
JENNIFER D. CALKINS,†§
CHRISTOPHER N. BALAKRISHNAN,‡§
W. BRYAN JENNINGS,*§ WILLIE J. SWANSON,†§
AND MICHAEL D. SORENSON‡§

Molecular studies of speciation in birds over the last three decades have been dominated by a focus on the geography, ecology, and timing of speciation, a tradition traceable to Mayr's *Systematics and the Origin of Species*. However, in the recent years, interest in the behavioral and molecular mechanisms of speciation in birds has increased, building in part on the older traditions and observations from domesticated species. The result is that many of the same mechanisms proffered for model lineages such as *Drosophila*—mechanisms such as genetic incompatibilities, reinforcement, and sexual selection—are now being seriously entertained for birds, albeit with much lower resolution. The recent completion of a draft sequence of the chicken genome, and an abundance of single nucleotide polymorphisms on the autosomes and sex chromosomes, will dramatically accelerate research on the molecular mechanisms of avian speciation over the next few years. The challenge for ornithologists is now to inform well studied examples of speciation in nature with increased molecular resolution—to clone speciation genes if they

*Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138; †Department of Genome Sciences, University of Washington, Seattle, WA 98195; and ‡Department of Biology, Boston University, Boston, MA 02215.

§S.V.E., S.B.K., J.D.C., C.N.B., W.B.J., W.J.S., and M.D.S. contributed equally to this work.

exist—and thereby evaluate the relative roles of extrinsic, intrinsic, deterministic, and stochastic causes for avian diversification.

Presumably, it was only a distant dream of mid-century participants in the Modern Synthesis that one day genetic analysis would come to dominate so completely the analysis of speciation in a clade such as birds. That genetics would be a useful tool in the analysis of speciation in model organisms had been evident since the early days of *Drosophila* genetics. Yet Mayr's *Systematics and the Origin of Species* (1942), published 11 years before the discovery of the structure of DNA, was a treasure-trove of speciation stories not of logistically tractable species with easily sampled and manipulated populations; rather, this book focused on speciation stories from the distant South Pacific, on what were, even for ornithologists, virtually inaccessible taxa with ranges straddling some of the most remote and challenging habitats of the planet. The allure of the exotic continues for ornithologists: Mayr and Diamond have recently undertaken a complete taxonomic and biogeographic revision of the birds of the Solomon Islands (2001), and detailed molecular phylogeographic tests of several speciation stories in this assemblage are finally underway (Filardi, 2003; Filardi and Smith, 2005; Smith, 2003). Indeed, the role of molecular techniques introduced to ornithology with the first allozyme surveys of avian populations nearly 35 years ago has been primarily to inform the *geography* and *timing* of speciation, thereby emphasizing extrinsic aspects of the speciation process: species delimitation, allopatric speciation, ecological divergence, bottlenecks, and the role of the Pleistocene (Avise, 2000; Barrowclough, 1983). Those mechanisms of avian speciation described by Mayr in terms of *internal* factors—for example, speciation resulting from so-called "genetic revolutions" (Mayr, 1963)—were often vague and, in the case of genetic revolutions, have been largely discredited (Barton and Charlesworth, 1987).

In the last 10 years, however, there has been a renewed interest in the behavioral, cognitive, and even molecular mechanisms of speciation in birds. This renaissance, spearheaded largely by recent reviews by Price (Irwin and Price, 1999; Price, 1998, 2002; Price and Bouvier, 2002), builds in part on the ancient tradition of avian husbandry and domestication, and in part on theoretical models suggesting a role for diverse behavioral factors in bird speciation, including sexual selection, sexual imprinting, learning, reinforcement, and genetic incompatibilities. Although biogeographic analyses still largely support allopatric speciation models (Coyne and Price, 2000), recent years have also witnessed the first serious attempts to document sympatric speciation in birds (Grant and Grant, 1979;

Sorenson *et al.*, 2003), and the frequency of cases of sympatric speciation and divergence due to hybrid incompatibilities or reinforcement is an open question (Coyne and Orr, 2004). Now, a draft of the complete chicken genome and >2.8 million chicken SNPs have been determined roughly 60 years after Mayr's landmark book (International Chicken Genome Sequencing Consortium, 2004; International Chicken Polymorphism Map Consortium, 2004). This treasure-trove of genes and genetic markers will no doubt spur rapid advances in both the geography and genetics of speciation in birds. This article reviews recent studies of extrinsic and intrinsic aspects of speciation in birds, focusing specifically on systematic and mechanistic issues that challenge the universality of the allopatric speciation paradigm.

GENE TREES, SPECIES DELIMITATION, AND PHYLOGEOGRAPHIC PATTERNS

Genetic data are serving an ever-increasingly important role in the delimitation of species, yet considerable controversy remains over which criteria to apply to this age-old problem (Avise and Ball, 1990; Coyne and Orr, 2004; Cracraft, 1983; Hey, 2001; Sites and Marshall, 2003; Wiens and Penkrot, 2002). Indeed, determining which of myriad species delimitation methods and species definitions is most appropriate for one's focal taxa remains one of the paramount challenges in systematics, with important consequences for evolutionary biologists as well as for conservation biologists (Crandall *et al.*, 2000; Moritz, 2002; Sites and Crandall, 1997). The fact that many avian sister taxa occur in allopatry—particularly in Gondwanan continents such as South America and Australia (Bates *et al.*, 1998; Cracraft, 1991)—makes interpretation of biogeographic histories more straightforward but can also make attempts at species delimitation particularly challenging. It has long been recognized that the biological species concept (BSC), with its emphasis on reproductive isolation, is inapplicable in many allopatric situations because there is no opportunity to test for reproductive isolation, rendering the concept arbitrary (Zink and McKitrick, 1995). Species concepts emphasizing genetic clustering of forms can be equally arbitrary (reviewed in Irwin and Price, 1999). Diagnosibility—the ability to delimit and identify distinguishing character states for a given collection of individuals or taxa, usually but not always in a phylogenetic context—has been proffered as a general consideration when delimiting species (Cracraft, 1983). Although diagnosibility is sometimes construed as being equivalent to "fixed" character or genetic differences between taxa, alternate fixations are not a requirement for diagnosibility. The rise in sophisticated statistical genetic algorithms and large-scale multilocus analyses of variation in birds and other taxa con-

firm that it is trivial to diagnose species, subspecies, or even populations, even in the presence of abundant shared polymorphisms; a recent study of native and introduced house finch populations in North America was able to readily diagnose populations that had been separated for 50–100 years (Wang *et al.*, 2003), even in the absence of fixed genetic or morphological differences. A naïve evaluation of the genetic patterns in this study would have reasonably inferred the existence of three diagnosable "species" of native and introduced House Finches within the continental United States and Hawaii. A similar ease of diagnosibility using large-scale multilocus approaches is observed for geographic populations of humans (Tishkoff and Kidd, 2004). Thus, diagnosibility will become highly problematic as the resolving power of multilocus approaches increases; the specter of statistical significance without biological significance will be perennial.

The criterion of monophyly advocated by proponents of the phylogenetic species concept (PSC) also suffers from arbitrariness, particularly given the disproportionate emphasis on mtDNA. It is not surprising that a number of workers have advocated the use of mtDNA in species delimitation, given its rapid attainment of reciprocal monophyly (relative to nuclear genes) and frequent ability to diagnose populations (Moore, 1994; Moritz, 1994, 2002; Wiens and Penkrot, 2002). Reciprocal monophyly of mtDNA is attractive because it is an objective criterion that can be applied to any animal system. However, use of mtDNA alone to delimit species has been criticized, because the complete organismal history is not captured and because many other loci in the genome may not exhibit reciprocal monophyly (Hudson and Coyne, 2002; Sites and Crandall, 1997; Wiens and Penkrot, 2002).

Another gene-tree-based criterion for species delimitation calls for finding reciprocal monophyly among a large majority (≈95%) of sampled nuclear genes (Avise and Ball, 1990; Baum and Shaw, 1995). Because nuclear DNA (nDNA) achieves reciprocal monophyly much slower than mtDNA, this criterion is considered conservative (Hudson and Coyne, 2002; Sites and Crandall, 1997). However, to date, and certainly at the time this species concept was put forward, there have been no avian data sets consisting of multiple independent gene trees with which to test the utility of this approach. A recent study of speciation in three Australian grassfinches (*Poephila*) using 30 independent nuclear loci provides a test of this concept. One of the taxa (*Poephila cincta*) examined is phenotypically and geographically very distinct from the other two, and its species status has never been questioned (Schodde and Mason, 1999); the two allopatric western lineages of long-tailed finches (*Poephila acuticauda/hecki*) are distinguished by fewer characters but are nonetheless diagnosable morphologically and have been designated phylogenetic species (Cracraft, 1986). Coalescent estimates of population divergence times sug-

gest a split of ≈0.5 million years ago for the two western lineages, and >0.7 million years ago for the basal split with *cincta*. Given the dynamics of nuclear genes under reasonable assumptions of ancestral population sizes for birds, it is therefore not surprising that considerable heterogeneity and conflict among the gene trees in *Poephila* was observed, even with regard to alleles sampled from the divergent *cincta* lineage (Fig. 6.1*a*) (Nei, 1987; Wakeley and Hey, 1997). There is reason to believe that these conflicts are the result of incomplete lineage sorting, rather than of hybridization or gene flow (Avise, 2000). Application of the criterion of genealogical concordance among ≈95% of sampled gene trees would result in lumping all three taxa together despite their considerable temporal and phenotypic divergence. We suggest that concordance among multiple nuclear genes will rarely be achieved among avian lineages that are considered good species by multiple other criteria; as in *Drosophila* (Ting *et al.*, 2000), if the characters leading to diagnosibility diverge by natural selection, they may outpace the well known but slow progression of the neutral nuclear genome from paraphyly to polyphyly to reciprocal monophyly (Neigel and Avise, 1986).

Although the incidence of reciprocal monophyly among genes and organisms may vary considerably, it is instructive to examine further the efficacy of this criterion in delimiting species of *Poephila*. The probability that a gene tree will match the species tree in a three-species scenario of divergence and isolation without gene flow has been known for some time (Hudson, 1992; Nei, 1987) and is determined by T, the ratio of time elapsed between speciation events to the ancestral effective population size (Fig. 6.1*b*). T must therefore equal or exceed 2.6 if there is to be significant concordance between the gene trees and species tree. The multilocus *Poephila* data suggests a value for T of 0.4, implying a substantial ancestral population size relative to divergence time, also the likely cause for the lack of concordance among gene trees. The probability of incongruence between gene and species trees for the *Poephila* loci are not strictly equivalent to the genealogical concordance among loci described by Avise and Ball (1990), because in the *Poephila* study only a single allele was sampled per species; thus, some of the loci exhibiting congruence with the species tree (Topology 1, Fig. 6.1*a*) may in fact exhibit genealogical incongruence, manifested as incomplete lineage sorting, upon sampling of further alleles. Nonetheless, these data, the first substantial sampling of multiple gene trees for an avian species, suggest that species arising rapidly or having ancestors with large effective population sizes will not satisfy the concordant genealogies criterion even though they are reasonable species under the BSC or morphological PSC. For avian species, finding any nuclear gene that has achieved reciprocal monophyly, whether by

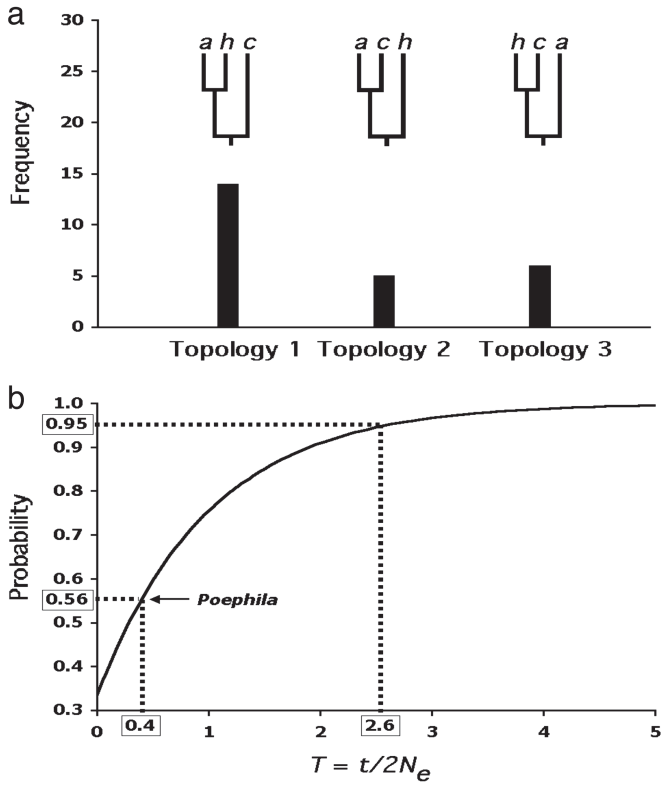


FIGURE 6.1 Conflicts between gene and species trees in *Poephila* finches, and their implications. (*a*) Frequency distribution of all three possible gene trees encountered in a survey of nuclear DNA sequence variation among three species of Australian grassfinches (*Poephila*). Branch tips are labeled as follows: a, *P. acuticauda*; h, *P. hecki*; c, *P. cincta*. *P. cincta* vs. *P. acuticauda/hecki* is a well established divergence across the Carpentarian barrier in northern Australia and is supported by numerous species-level phenotypic differences; *P. hecki* vs. *acuticauda* are allopatric but represent a more recent taxonomic split which is regarded by some (Cracraft, 1986) as phylogenetic species. In the genetic survey, a single allele was sampled from each of the three taxa for thirty presumably unlinked loci. Twenty-five gene trees exhibiting all three possible topologies were unambiguously reconstructed. Gene trees for five loci could not be resolved. Topology 1 reflects the presumed species tree. (*b*) Unconditional probabilities of a gene tree being congruent with the species tree over a range of values for T (after Hudson, 1992). The curve is based on the equation for three species [$P_{(\text{congruent gene tree})} = 1 - 2/3e^{-T}$ (Hudson, 1992; Nei, 1987), where T is equal to $t/2N_e$, t is defined as the time between speciation events (in generations), and $2Ne$ is twice the size of the effective population size of the basal ancestor. When $T = 0$, the topology of a gene tree of three sampled alleles is expected to be random with respect to the species tree

neutral or selective means (akin to finding a fixed diagnostic phenotypic trait), may be a reasonable criterion for delimiting species.

It is certain that molecular approaches will continue to play an important role in species delimitation. However, the battle over species delimitation between the nuclear and mitochondrial genomes, and between the BSC and PSC, will have no victor. Nuclear gene trees will not provide enough phylogenetic resolution to satisfy avian systematists. Furthermore, the high levels of recombination detected in the first surveys of avian SNP variation appear prohibitive for standard phylogenetic analysis (Edwards and Dillon, 2004; International Chicken Polymorphism Map Consortium, 2004). On the other hand, in our view, maternally inherited mtDNA can never capture enough of a species history to delimit species on its own. Although mtDNA will frequently deliver clean phylogeographic breaks within avian species, these breaks need not have their origin in reproductive isolation (Irwin, 2002) and may in some cases be driven by natural selection (Hudson and Turelli, 2003). These same species showing clean mitochondrial breaks will frequently look very messy with regard to nuclear gene splits, as decades of allozyme analyses have already confirmed. We suggest that, despite its disproportionate contribution to revealing phylogeographic patterns and its ability to reflect cessation of female gene flow more rapidly than nuclear genes, mtDNA should not have priority over nuclear genes in avian species delimitation. Ours is a generation of avian systematists raised primarily on single locus analyses of avian phylogeny and divergence. But nuclear genes can and should be interrogated in questions of avian taxonomy, even if the interpretation of nuclear histories and the contrast with mtDNA histories will be challenging.

The Role of Song in Allopatric and Sympatric Speciation

As suggested by Mayr (1942), divergence of populations in allopatry appears to be the dominant mode of speciation in birds. Molecular phylogenies have provided new opportunities for testing alternative geographic models, and avian sister taxa generally meet the expectation of having allopatric distributions, whether because of vicariance or dispersal (Drovetski *et al.*, 2004; Pereira and Baker, 2004; Voelker, 1999; Zink *et al.*, 2000). Current distributions, however, do not necessarily reflect the geo-

(i.e., probability of 0.33). Note that T must be at least as large as 2.6 for a 0.95 probability of congruence, yet the empirically derived probability for *Poephila* finches of 0.56 results in a T value of only 0.4. The *Poephila* tree congruence probability is based on the fact that 14 of 25 independent gene trees matched the presumed species tree observed by W.B.J. and S.V.E. (unpublished work).

graphic context of speciation given the potential for dispersal and range expansion in birds (Chesser and Zink, 1994). When expectations are derived from a model incorporating random changes in geographic range, phylogenetic data suggest allopatric speciation as the predominant mode in several avian groups, with sympatry due to range changes after speciation (Barraclough and Vogler, 2000). Greater asymmetry in range size for recently evolved species also implies a role for peripatric speciation (Barraclough and Vogler, 2000), as suggested in other recent avian studies (Drovetski and Ronquist, 2003; Johnson and Cicero, 2002; Omland, 1997). The lability of geographic ranges, however, ultimately limits the power of phylogenetic approaches to distinguish between alternative geographic models of speciation (Losos and Glor, 2004).

Greater insight into avian speciation is perhaps gained by focusing on the processes of population divergence and mechanisms of reproductive isolation, particularly in closely related taxa and/or diverging populations. Both ecological and sexual selection may contribute to rapid morphological and behavioral divergence in allopatric or parapatric populations (Rasner *et al.*, 2004; Uy and Borgia, 2000; Yeh, 2004). Whether these changes lead to speciation, however, depends on the evolution of reproductive isolation before or after secondary contact. Avian species typically retain hybrid viability and fertility for millions of years after speciation, reflecting a general lack of intrinsic isolating mechanisms among closely related species (Price and Bouvier, 2002). Although ecological and/or sexual selection against hybrids may help to maintain species boundaries, reproductive isolation in birds will often depend on prezygotic mechanisms. Thus, divergence in characters involved in mate choice, such as song, plumage, and behavioral displays, likely play a central role in avian speciation. The role of song is particularly interesting given the multiple factors influencing vocal evolution and the potential for rapid change through learning and cultural evolution.

Gradual divergence of song in allopatric populations may result in reproductive isolation upon secondary contact. Although generally difficult to observe, this process is captured in present day populations of the greenish warbler (*Phylloscopus trochiloides*) complex. The range of this Old World species forms a narrow ring around the southern margin of the Tibetan plateau with eastern and western populations extending northward, expanding longitudinally, and meeting in Siberia. Genetic composition and songs change gradually through the nearly continuous ring of intergrading populations, but eastern and western populations in the north are reproductively isolated because of differences in song (Irwin *et al.*, 2001b). Parallel sexual selection for increased song complexity in northern latitudes has apparently resulted in the stochastic divergence of songs in eastern and western populations (Irwin, 2000). The gradual intergrada-

tion through intermediate populations in this ring species are analogous to the gradual changes that might occur over time in geographically isolated populations diverging in a similar manner (Irwin *et al.*, 2001a).

Songs may diverge as a direct result of habitat-dependent selection or indirectly as a consequence of morphological adaptations, such as those related to foraging. It has long been recognized that different types of vocalizations vary in their quality as signals in different habitats, and recent studies suggest a role for habitat-dependent selection in population divergence and reproductive isolation. Two subspecies of song sparrows (*Melospiza melodia*) differ both in song characteristics and preferred habitat, with *Melospiza melodia hermanii* occupying more dense vegetation than *Melospiza melodia fallax* and singing a lower frequency song with more widely spaced elements, a pattern consistent with acoustic adaptation (Patten *et al.*, 2004). Playback experiments further indicate that both males and females show greater response to homotypic songs, suggesting a role for song in reproductive isolation and the consequent development of significant genetic differentiation between the subspecies (Patten *et al.*, 2004). Similar habitat-dependent vocal divergence accompanies morphological differentiation in the little greenbul (*Andropodus virens*) and may promote population differentiation across ecological gradients (Slabbekoorn and Smith, 2002b).

Ecological selection on other characters also may result in correlated vocal evolution that contributes to prezygotic reproductive isolation. In Darwin's finches, bill morphology and vocal characteristics are correlated because of physical constraints on sound production, perhaps contributing to the diversification of these species (Podos, 2001). In contrast, the black-bellied seedcracker (*Pyrenestes ostrinus*), which shows a similar pattern of divergent selection on bill morphology, shows no effect of bill size on vocal performance, contributing to the conclusion that bill size variation in this species reflects only intraspecific niche polymorphism and not incipient speciation (Slabbekoorn and Smith, 2000; Smith, 1993).

As a learned behavior in many birds, song is subject to rapid cultural evolution in which stochastic innovations or errors in copying are spread as individuals learn songs from their parents and/or neighbors (Grant and Grant, 1996; Payne, 1996). Song learning may also increase the rate at which genetic predispositions to learn or prefer certain songs evolve in allopatry (Lachlan and Servedio, 2004). Song learning, however, may sometimes inhibit reproductive isolation upon secondary contact if there has been only minimal divergence in the capacity to learn particular songs and/or the morphological structures affecting sound production (Slabbekoorn and Smith, 2002a). A well known example of reproductive character displacement involves the *Ficedula* flycatchers in which a sexually selected male plumage trait shows greater divergence in sympatry

than in allopatry (Saetre *et al.*, 1997). Results were mixed, however, in a recent analysis of vocal divergence in this system (Haavie *et al.*, 2004). In sympatry, songs of pied flycatchers *Ficedula hypoleuca* have converged on those of collared flycatchers *Ficedula albicollis* because of heterospecific copying and singing of mixed songs. Although collared flycatcher songs in the zone of contact have diverged away from pied flycatcher songs typical of allopatric populations, the net effect of these changes is greater song similarity in sympatry.

Allopatric divergence of songs among suboscines and other birds in which differences in song are genetically determined may evolve more slowly but should also contribute to reproductive isolation. In a large comparative analysis of antbirds (Thamnophilidae), Seddon (2005) found evidence of vocal divergence both as a correlated effect of morphological evolution and as a response to habitat-dependent selection on signal transmission. In addition, among trios of closely related species, sympatric forms exhibited striking vocal divergence in comparison to allopatric taxa, providing strong evidence for reproductive character displacement and the role of song in reproductive isolation. Despite divergence of song in allopatry, individuals of different species may recognize each other as potential mates upon secondary contact, leading to hybridization. If song differences are genetically determined, hybrid offspring may have intermediate songs (de Kort *et al.*, 2002).

Given divergence in vocalizations, the development of song preferences through sexual imprinting may contribute to reproductive isolation even without genetic evolution of female preferences (Irwin and Price, 1999; Price, 1998). In Darwin's finches, for example, cultural inheritance clearly plays a greater role than bill morphology in determining songs and song preferences (Grant and Grant, 1996, 1997) and is critical in promoting reproductive isolation after secondary contact of populations that have diverged in feeding adaptations in allopatry (Grant and Grant, 2002). Ecological selection against hybrid individuals also helped maintain species boundaries, at least before changes associated with El Niño events in the 1980s (Grant and Grant, 1998). In the past 20 years, however, increased fitness of hybrids has resulted in substantial genetic introgression from *Geospiza fortis* to *Geospiza scandens* on Daphne Major (Grant *et al.*, 2004). Although learned songs and song preferences are strong determinants of pair formation in these species, reproductive isolation is imperfect because of constraints on mate choice imposed by asymmetries in population size and operational sex ratios as well as infrequent cases of individuals misimprinting on heterospecific songs (Grant and Grant, 1996, 1997, 1998).

Song learning and sexual imprinting explain the recent diversification of brood parasitic indigobirds (genus *Vidua*), the best and perhaps

only example of sympatric speciation in birds (Fig. 6.2). Male indigobirds include mimicry of host song in a repertoire that also includes species-specific indigobird vocalizations learned from other male indigobirds mimicking the same host (Payne *et al.*, 1998). Likewise, female preferences for both male song and host nests result from imprinting on the host (Payne *et al.*, 2000). Thus, speciation in indigobirds begins with reproductive isolation as a consequence of host colonization and only then proceeds to divergence in other characters, including host-specific mimicry of mouth markings and colors by indigobird chicks. Indigobird species within a region show a pattern of incomplete but significant genetic differentiation (Sorenson *et al.*, 2003) but also genetic continuity across intermediate spatial scales (Sefc *et al.*, 2005), a pattern consistent with recent sympatric speciation and current reproductive isolation (Via, 2001). As in Darwin's finches, song learning likely plays a role in hybridization between indigobird species. When females parasitize the host of another indigobird species (Payne *et al.*, 1993), their offspring are likely to hybridize in the subsequent generation because they have imprinted on the songs of the alternate host (Payne and Sorenson, 2004). The frequency of misimprinting and hybridization, however, appears to be lower in indigobirds than among the *Geospiza* ground finches (Grant *et al.*, 2004; Payne *et al.*, 1993).

In birds generally, the importance of prezygotic isolating mechanisms may allow for rapid speciation, whereas the slow development of intrinsic postzygotic isolation will facilitate continuing hybridization. Closely related taxa may therefore be strongly differentiated at only a small number of loci influenced by divergent ecological or sexual selection. Loci "that can be shown to cause some degree of ecological, sexual or postmating isolation between young, or even nascent, species" are good candidates for speciation genes (Wu and Ting, 2004). Finding such genes and understanding the genetics of avian speciation are challenging but increasingly realistic objectives as genomic resources and molecular methods continue to evolve.

Sex Chromosomes and Avian Speciation

The architects of the Modern Synthesis laid the foundation for a body of work that has resulted in two "rules" of speciation that directly implicate sex chromosomes: Haldane's rule and the large X(Z)-effect (Coyne and Orr, 1989). Haldane's rule is the preferential sterility or inviability of the heterogametic sex in hybrid crosses, when a sex-biased fitness loss in hybrids occurs. This phenomenon is found across diverse taxa, including butterflies and birds in which the female is heterogametic (ZW) (Haldane, 1922). Although the X chromosome was experimentally implicated in

Haldane's rule and hybrid male sterility in *Drosophila* decades ago (Dobzansky, 1936), it was not until recently that this large effect of the hemizygous sex chromosome was documented for birds, using genetic data from natural hybrid zones and domesticated species (Price, 2002; Saetre *et al.*, 2003). In this section, we review empirical and theoretical work that explores these two rules of speciation in birds.

The phenomenon of Haldane's rule describes patterns of postzygotic incompatibilities in hybrids and is likely caused by negative epistatic interactions between loci derived from divergent parental genomes (Coyne and Orr, 2004). Heterogametic hybrids are more severely affected by these interactions because, unlike the homogametic sex, they fully express recessive sex-linked genes. Interestingly, avian and Lepidopteran F1 hybrid females may suffer from an additional source of negative epistasis between parental genomes, namely that between the maternally derived mitochondria or cytoplasm and the paternally derived Z chromosomes (Presgraves, 2002). There is debate over the extent to which Haldane's rule is driven by interactions among sometimes rapidly diverging sex chromosomes *per se*, or whether it is the peculiar dominance patterns exhibited by the hemizygous sex chromosomes that underlie the rule. Support for Haldane's rule is excellent in birds based on experimental studies of hybrid fitness in ducks (Tubaro and Lijtmaer, 2002), pigeons and doves (Lijtmaer *et al.*, 2003), and many other avian taxa (Price and Bouvier, 2002).

Price and Bouvier (2002) characterized patterns of postzygotic incompatibilities in birds using published data from 254 hybrid crosses and found that the order in which incompatibilities accumulate with increasing species divergence differs between birds and other taxa, a pattern that informs the causes of Haldane's rule in birds (Fig. 6.3). In *Drosophila*, male sterility appears at early stages of divergence, followed in turn by male inviability, then female sterility, and finally female inviability. By contrast, avian incompatibilities accumulate in the following order: female sterility, male sterility, female inviability, and male inviability (Price and Bouvier, 2002). Thus, in birds, homogametic (male) sterility evolves at earlier stages than does homogametic (female) sterility in *Drosophila* (Coyne and Orr, 1989). The appearance of homogametic (male) sterility before heterogametic (female) inviability in birds may reflect a general trend, regardless of sex-chromosome system, of the rapid evolution of male reproductive genes via sexual selection, resulting in high divergence between species at these loci (Wu and Davis, 1993). However, the rapid evolution of male reproductive genes via sexual selection—so-called "faster-male evolution" or the "sexual selection model" of Wu and Davis (1993)—works in opposition to Haldane's rule in birds because this particular force will negatively affect the homogametic sex (males), not the

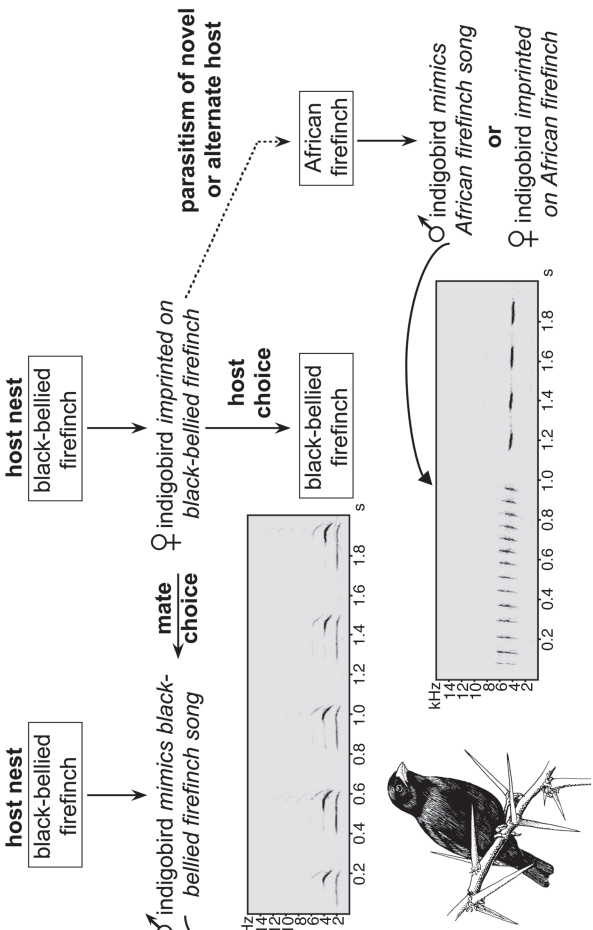


FIGURE 6.2 Behavioral imprinting maintains host specificity and genetic cohesion of indigobird species but also provides a mechanism for rapid speciation when new hosts are colonized. Male indigobirds mimic the songs of their hosts, whereas females use song cues to choose both their mates and the nests they parasitize. Rarely, females lay in the nest of a novel or alternate host. The resulting offspring imprint on the novel host and are therefore reproductively isolated from their parent population. Indigobird drawing by Karen Klitz.

heterogametic sex (females), in ZW species. By contrast, sexual selection accentuates the pattern for sterility in XY species, such as *Drosophila*, perhaps explaining the preponderance of cases of hybrid male sterility in the genus (Coyne and Orr, 2004).

Support for sexual selection as a modulator of Haldane's rule in birds was first detected in the relative proportions of cases of Haldane's rule for sterility and inviability (Kirkpatrick and Hall, 2004). Specifically, the majority of cases of Haldane's rule in XY species involve heterogametic sterility, whereas Wu and Davis (1993) initially found as many cases for heterogametic inviability as for sterility in birds and other ZW species. This pattern is expected when traits expressed only in males are evolving rapidly because it results in male hybrid sterility in early stages of divergence, resulting in fewer cases of Haldane's rule for sterility. However, in the larger avian data set of Price and Bouvier (2002), five times as many cases of Haldane's rule involve sterility as inviability, a pattern that does not support sexual selection as a modulator of Haldane's rule. Direct evidence for a role for sexually selected genes in avian speciation, such as an examination of the molecular evolution of avian reproductive genes (see next section), is a necessary but not sufficient condition for their role Haldane's rule in birds. Additionally, the domain of such sexually selected genes needs to be determined: Do they include sexually selected traits such as plumage and song that are not directly involved in the physical production of gametes? Because males and females of many bird species exhibit correlated evolution of plumage and other sexually selected traits, the specific expression patterns of genes related to both reproduction and morphology need to be investigated to determine how such traits will influence Haldane's rule.

The rate of evolution of sex chromosomes has important implications for their role in avian speciation. In general, when rapidly evolving loci diverge between species, incompatibilities between these and other loci can arise when parental genomes come together in hybrids. In the same way that rapid evolution of sexual traits in males can produce incompatibilities in hybrids at loci involved in male sexual traits—whether autosomal or sex-linked—a faster rate of evolution of the sex chromosomes, combined with effects of dominance (Coyne and Orr, 2004), is one hypothesis for the large X(Z)-effect in low-fitness hybrids. So-called "faster-X(Z) evolution" owing to selection on favorable mutations on the hemizygous X(Z) when they are partially recessive, manifested as more rapid substitution rates on the X chromosome than on autosomes, was demonstrated theoretically by Charlesworth *et al.* (1987). More recently, Kirkpatrick and Hall (2004) modeled the relative rates of evolution of sex chromosomes and autosomes while accounting for interactions between mode of inheritance and the intrinsically higher mutation rate in males

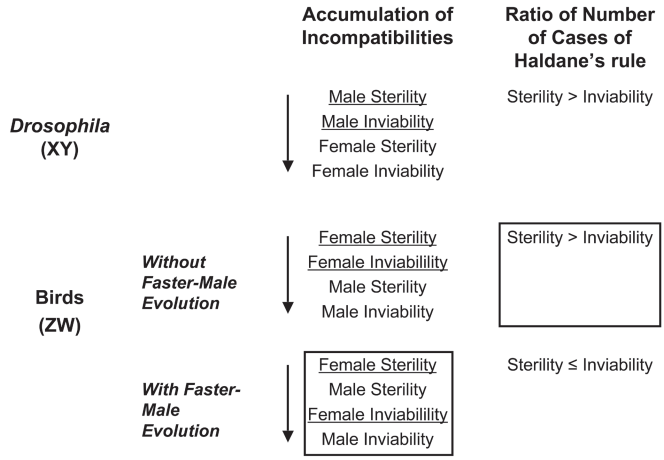


FIGURE 6.3 Summary of Haldane's rule in birds. Predictions are based on the presence or absence of faster-male evolution (sexual selection hypothesis), which changes the order of accumulation of incompatibilities in WZ species. The predicted order of accumulation of incompatibilities for the case without sexual selection is based on the pattern in *Drosophila*. The data observed by Price and Bouvier (2002) are boxed. The heterogametic sex is underlined.

known as "male-driven evolution." Male-driven evolution (not to be confused with faster-male evolution) has been documented in birds, mammals, and plants, most likely because of a greater number of cell divisions in the male than in the female germ line (Hurst and Ellegren, 1998); critically, such studies have documented faster neutral evolution on the avian Z but not faster adaptive, presumably nonsynonymous evolution, as predicted by Charlesworth *et al.* (1987). Values of α, the ratio of male to female mutation rates, have been estimated from sequence data and range from $\approx$1.8 to $\approx$6.4 in both birds and mammals (reviewed in Wu and Ting, 2004). The avian Z chromosome is expected to have a higher neutral mutation rate than the autosomes because it passes through the male germ line twice as often as the female germ line, whereas avian autosomes spend equal time in both germ lines. In contrast to predictions of hemizygous sex chromosome evolution in mammals, where mutations must be strongly recessive for the X to evolve faster than autosomes (Charlesworth *et al.*, 1987), the higher mutational flux on the Z of birds is predicted to produce faster rates of Z chromosome evolution relative to autosomes even when mutations are strongly dominant (Kirkpatrick and Hall, 2004). Perhaps not surprisingly, support for fast-X evolution is inconsistent (Betancourt *et al.*, 2002; Countermann *et al.*, 2004), whereas

evidence of fast-Z evolution is more compelling (Axelsson *et al.*, 2003; Bartosch-Harlid *et al.*, 2003; Carmichael *et al.*, 2000; Ellegren and Fridolffson, 1997; Kahn and Quinn, 1999). Sequence data flowing from the chicken genome project have already facilitated confirmation of faster-Z evolution in birds, both at the intra- and interspecific levels (Axelsson *et al.*, 2003).

The first empirical inquiries into the role of sex chromosomes in the speciation process of natural avian groups have focused on the well studied system provided by hybridization in Old World *Ficedula* flycatchers. Saetre *et al.* (2003) found that Z-linked single nucleotide polymorphism (SNP) markers in *F. hypoleuca* and *F. albicollis* showed little evidence for introgression, whereas substantial introgression was documented for autosomal SNPs. The sex chromosomes had a large effect on the fertility of hybrids: Among birds with heterozygous sex chromosomes (one from each parental species), 7 of 7 females were sterile, as opposed to 3 of 11 males; a pattern that is consistent with Haldane's rule, although whether interactions among Z loci or between the Z and W or autosomes is the cause remains unclear. The possibility of linkage of loci involved in pre- and postzygotic isolation in this system (Saetre *et al.*, 2003) motivated the development of a novel model of reinforcement (Servedio and Saetre, 2003). The model focused on the evolution of linkages among a male trait locus, a female preference locus (collectively referred to as "mating loci"), and two postzygotic incompatibility loci. Consistent with previous theory, prezygotic isolation is reinforced through tight linkage between the mating loci and incompatibility loci. As incompatibility loci diverge between the two populations, causing a decrease in the fitness of hybrids, the frequency of assortative mating increases, thereby reducing the occurrence of interspecific matings. The tight linkage between the mating loci and the incompatibility loci creates a positive feedback loop because the frequency of linked incompatibility loci increases in tandem with the loci causing population-specific mating as a result of genetic hitchhiking. The positive feedback loop is enhanced when the linkage group occurs on a sex chromosome for two reasons. First, if as assumed in the model crossing over does not occur between sex chromosomes in the heterogametic sex, hitchhiking is enhanced. (This lack of crossing over does not strictly represent the case in nature because, as in mammals, many birds possess a pseudoautosomal region of varying size in which crossing over does in fact occur.) Second, recessive mutations are more exposed to positive selection in the heterogametic sex and may rapidly sweep through the population. The model reveals a number of interesting features about the dynamics of the *Ficedula* hybrid zone (Saetre *et al.*, 2003). However, it is unclear how well the *Ficedula* hybrid zone represents the diversity of avian hybrid zones. Many avian hybrids show little evidence of fitness

loss (Grant and Grant, 1992) and may even enjoy an ecological advantage over parental species (Good *et al.*, 2000), conditions that do not favor reinforcement of prezygotic isolation.

Studies in birds and other taxa indicate that sex chromosomes may disproportionately harbor genes related to sex and reproduction (Reinhold, 1998; Wang *et al.*, 2001). Literature on a variety of domesticated avian species suggests that 22% of traits that distinguish breeds, including likely targets of sexual selection in natural populations such as plumage and vocalizations, are sex-linked (Price, 2002), an excess when one considers that the Z chromosome comprises ≈2.7% of the chicken genome, by our estimate using data from the chicken genome project (International Chicken Polymorphism Map Consortium, 2004). Avian speciation is commonly demonstrated to involve prezygotic isolation in traits such as song or display (Grant and Grant, 1997; see previous section), making the genes that underlie these traits promising candidate speciation genes (Wu and Ting, 2004). The dawn of the genomic era for ornithologists provides exciting opportunities to study the genomic composition of avian sex chromosomes and will allow a better understanding of the complex interaction between their gene content, gene expression patterns, and rate of evolution in the context of speciation.

CRYPTIC MATE CHOICE AND CONFLICT: A ROLE FOR REPRODUCTIVE PROTEINS IN AVIAN SPECIATION?

Birds provide an abundance of examples of intense sexual selection through cryptic female choice and sperm competition (Birkhead and Møller, 1992). The dramatic evolutionary consequences of this competition have been documented at the molecular level in mammals and invertebrates, through genes collectively known as reproductive proteins. Rapid evolution of reproductive proteins has been documented in a diverse array of taxonomic groups ranging from humans to corn and is thought to be a component of the speciation process (Swanson and Vacquier, 2002). However, to our knowledge there are no examples of this process from birds. One potential protein is the female-specific gene HNTW, which, although its function is unknown, shows adaptive evolution (Ceplitis and Ellegren, 2004). With the recent draft release of the chicken genome (International Chicken Genome Sequencing Consortium, 2004) we can expect that the molecular evolution of avian reproductive genes will come under intense scrutiny, a development that is particularly exciting because birds offer unique opportunities to study the selective forces driving the rapid evolution of reproductive proteins.

Reproductive genes can be split into two classes. First, there are gamete recognition proteins on the surface of gametes that are directly in-

volved in sperm–egg interaction. The classic example of a rapidly evolving gamete recognition protein is the abalone sperm protein lysin, perhaps the most rapidly diverging protein yet discovered (Lee and Vacquier, 1992). Lysin acts to dissolve the egg vitelline envelope, a process that demonstrates species specificity. In mammals it has been demonstrated that sperm and egg molecules are among the most diverse found within the genome, with a minimum of 10 reproductive genes showing evidence of adaptive evolution (Swanson *et al.*, 2003b). One such gene is the mammalian egg coat protein ZP3. The region within this protein undergoing adaptive evolution corresponds to experimentally determined binding sites (Swanson *et al.*, 2001a), suggesting that the rapid evolution relates to fertilization.

The second class of reproductive proteins exhibiting rapid evolutionary change are not directly involved in surface recognition of the gametes. These include components of seminal fluid (Kingan *et al.*, 2003), pheromones and protamines. In *Drosophila* seminal fluid, an estimated 10% of the genes show the signatures of adaptive evolution (Swanson *et al.*, 2003a). Many of these genes, called accessory gland proteins (ACPs) act to manipulate female reproductive behavior, thus increasing male fitness (Wolfner, 1997). In primates, seminogellin II (SEM2), a major component of seminal fluid, shows rapid adaptive evolution. This protein is involved in copulatory plug formation in rodents, and in primates its rate of evolution shows a correlation with mating system (Dorus *et al.*, 2004).

The recent comparison of predicted genes in chicken genome with the human genome supports the pattern of divergence found in reproductive proteins across taxa (International Chicken Genome Sequencing Consortium, 2004). Genes implicated in reproduction appear less conserved between chicken and human than genes involved in typical "housekeeping" functions. For example, among genes classified into 10 different tissue specificities, those expressed in the testis showed the most divergence: 65% sequence conservation compared with the mean of 75% across all genes. Many of these genes, such as ZP3, have orthologues in birds and would be important candidates for targets of natural selection. Other potential target genes in birds include seminal fluid proteins. The prediction of natural selection on such genes in other species can be inferred from reproductive observations. Adkins-Reagan (1999) documented a viscous mucoprotein produced by Japanese quail (*Corturnix japonica*) thought to increase the probability of fertilization when a hard-shelled egg is in the uterus. The origin of such viscosity must have a basis in protein evolution, although the target loci have not yet been identified.

Cryptic female choice, sperm competition, and sexual conflict are three nonexclusive hypotheses for the forces driving the rapid evolution of these proteins (Swanson and Vacquier, 2002). Cryptic female choice of

reproductive proteins involves the "preference" of male proteins on the surface of the sperm or in the seminal fluid by egg coat proteins, egg proteins, or proteins in the female reproductive tract, examples of which come mostly from invertebrates (Palumbi, 1999). The ability of many birds to store sperm provides a ready mechanism for cryptic female choice. Sperm competition involves the direct competition between sperm of different males providing a potent source selection acting to improve sperm motility (Birkhean and Møller, 1992). Sexual conflict occurs when the reproductive goals of the sexes differ and is thought to drive rapid coevolutionary arms races in reproductive proteins at the molecular level (Gavrilits, 2000; Gowaty, 1996; Rice, 1996). Testing the various hypotheses for the rapid evolution of reproductive proteins is particularly promising in birds. In no other taxonomic group is so much known about the diversity of mating systems and the natural history of female preferences driving trait differences (Parker and Burley, 1998). Combining evolutionary data on reproductive proteins with predictions of sperm competition and mate choice from behavioral studies, as has been attempted with the SEM2 locus in primates (Dorus *et al.*, 2004), is a promising avenue of research in birds. Furthermore, reproductive traits unique to birds, such as physiological polyspermy, the ability of multiple sperm to penetrate the egg without rendering it inviable, permit testing of specific hypothesis underlying rapid adaptive protein evolution. Also, variation across avian species in particular traits such as the presence and type of intromittent organs, from penises to cloacal protuberances to the absence of any intromittent organ, allows for hypothesis-testing that would be difficult in taxonomic groups without such variation. Conversely, the study of reproductive protein evolution in birds might help clarify the roles of these traits on reproductive evolution and reinforcement. Avian mating systems are thought to play an important role in the speciation process (Jennions and Petrie, 2000), and reproductive proteins provide a convenient link between these two arenas.

CONCLUSION

Rich natural histories, diverse biogeographies, and complex character traits and mating systems have made birds central to the formulation of many speciation theories. Now, these and other ideas are more amenable to direct testing with the increased molecular access provided by the chicken genome and by new genomic technologies and resources. These new tools will increase the resolving power of both phylogeographic analysis of speciation and of interactions among diverging genomes. Largescale EST surveys, bacterial artificial chromosome (BAC) libraries, and other genetic resources for dimorphic species such as zebra finches,

turkeys, and Japanese quail are available, and high-resolution linkage maps will soon follow. We can expect the information from these model avian species to inform the analysis of speciation within their respective clades and beyond (Edwards *et al.*, 2005). Linking of these genetic analyses with predictions from theory and application to natural populations will make for exciting times to come in avian speciation studies.

ACKNOWLEDGMENTS

We thank K. De Quieroz, J. Hey, H. A. Orr, C. Filardi, and J. Wakeley for helpful discussion and R. Payne, G. P. Saeter, T. Price, M. Servedio, and J. Feder for very constructive comments on the manuscript. This work was supported in part by National Science Foundation grants (to S.V.E. and M.D.S.). W.J.S. was supported by National Institutes of Health Grant HD42563-01A1 and National Science Foundation Grant DEB-0410112.

REFERENCES

Adkins-Reagan, E. (1999) Foam produced by male coturnix quail what is its function? *Auk* **116,** 184–193.

Avise, J. C. (2000) *Phylogeography: The History and Formation of Species* (Harvard Univ. Press, Cambridge, MA).

Avise, J. C. & Ball, R. M. (1990) Principles of genealogical concordance in species concepts and biological taxonomy. *Oxf. Surv. Evol. Biol.* **7,** 45–67.

Axelsson, E., Smith, N. G., Sundstrom, H., Berlin, S. & Ellegren, H. (2004) Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. *Mol. Biol. Evol.* **21,** 1538–1547.

Barraclough, T. G. & Vogler, A. P. (2000) Detecting the geographical pattern of speciation from species-level phylogenies. *Am. Nat.* **155,** 419–434.

Barrowclough, G. F. (1983) In *Perspectives in Ornithology: Essays Presented for the Centennial of the American Ornithologists' Union*, eds. Brush, A. H. & Clark, G. A., Jr. (Cambridge Univ. Press, New York), pp. 223–261.

Barton, N. H. & Charlesworth, B. (1987) Genetic revolutions, founder effects and speciation. *Annu. Rev. Ecol. Syst.* **15,** 133–164.

Bartosch-Harlid, A., Berlin, S., Smith, N. G., Moller, A. P. & Ellegren, H. (2003) Life history and the male mutation bias. *Evolution (Lawrence, Kans.)* **57,** 2398–2406.

Bates, J. M., Hackett, S. J. & Cracraft, J. (1998) Area-relationships in the Neotropical lowlands: An hypothesis based on raw distributions of Passerine birds. *J. Biogeogr.* **25,** 783–793.

Baum, D. & Shaw, K. L. (1995) In *Experimental and Molecular Approaches to Plant Biosystematics*, eds. Hoch, P. C. & Stephenson, A. C. (Missouri Botanical Garden, St. Louis), pp. 289–303.

Betancourt, A. J., Presgraves, D. C. & Swanson, W. J. (2002) A test for faster X evolution in *Drosophila*. *Mol. Biol. Evol.* 1816–1819.

Birkhead, T. R. & Møller, A. P. (1992) *Sperm Competition in Birds; Evolutionary Causes and Consequences* (Academic, London).

Carmichael, A. N., Fridolfsson, A. K., Halverson, J. & Ellegren, H. (2000) Male-biased mutation rates revealed from Z and W chromosome-linked ATP synthase alpha-subunit (ATP5A1) sequences in birds. *J. Mol. Evol.* **50,** 443–447.

Ceplitis, H. & Ellegren, H. (2004) Adaptive molecular evolution of HINTW, a female-specific gene in birds. *Mol. Biol. Evol.* **21,** 249–254.

Charlesworth, B., Coyne, J. A. & Barton, N. H. (1987) The relative rates of evolution of sex-chromosomes and autosomes. *Am. Nat.* **130,** 113–146.

Chesser, R. T. & Zink, R. M. (1994) Modes of speciation in birds—A test of Lynchs method. *Evolution (Lawrence, Kans.)* **48,** 490–497.

Countermann, B. A., Ortiz-Barrientos, D. & Noor, M. A. F. (2004) Using comparative genomic data to test for fast-X evolution. *Evolution (Lawrence, Kans.)* 656–660.

Coyne, J. A. & Orr, H. A. (1989) In *Speciation and Its Consequences*, eds. Otte, D. & Endler, J. A. (Sinauer, Sunderland, MA), pp. 180–207.

Coyne, J. A. & Orr, H. A. (2004) *Speciation* (Sinauer, Sunderland, MA).

Coyne, J. A. & Price, T. D. (2000) Little evidence for sympatric speciation in island birds. *Evolution (Lawrence, Kans.)* **54,** 2166–2171.

Cracraft, J. (1983) Species concepts and speciation analysis. *Curr. Ornithol.* **1,** 159–187.

Cracraft, J. (1986) Origin and evolution of continental biotas: speciation and historical congruence within the Australian avifauna. *Evolution (Lawrence, Kans.)* **40,** 977–996.

Cracraft, J. (1991) Patterns of diversification within continental biotas: Hierarchical congruence among the areas of endemism of Australian vertebrates. *Aust. Syst. Bot.* **4,** 211–227.

Crandall, K. A., Bininda-Emonds, O. R. P., Mace, G. M. & Wayne, R. K. (2000) Considering evolutionary processes in conservation biology. *Trends Ecol. Evol.* **15,** 290–295.

de Kort, S. R., den Hartog, P. M. & ten Cate, C. (2002) Diverge or merge? The effect of sympatric occurrence on the territorial vocalizations of the vinaceous dove *Streptopelia vinacea* and the ring-necked dove *S. capicola*. *J. Avian Biol.* **33,** 150–158.

Dobzansky, T. (1936) Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* **21**.

Dorus, S., Evans, P. D., Wyckoff, G. J., Choi, S. S. & Lahn, B. T. (2004) Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nat. Genet.* **36,** 1326–1329.

Drovetski, S. V. & Ronquist, F. (2003) Plio-Pleistocene climatic oscilations, Holarctic biogeography and speciation in an avian subfamily. *J. Biogeogr.* **30,** 1173–1181.

Drovetski, S. V., Zink, R. M., Rohwer, S., Fadeev, I. V., Nesterov, E. V., Karagodin, I., Koblik, E. A. & Redkin, Y. A. (2004) Complex biogeographic history of a Holarctic passerine. *Proc. R. Soc. London Ser B* **271,** 545–551.

Edwards, S. V. & Dillon, M. (2004) Hitchhiking and recombination in birds: Evidence from Mhc-linked and unlinked loci in red-winged blackbirds *(Agelaius phoeniceus). Genet. Res.* **84,** 175–192.

Edwards, S. V., Jennings, W. B. & Shedlock, A. M. (2005) Phylogenetics of modern birds in the era of genomics. *Proc. R. Soc. London Ser. B*, in press.

Ellegren, H. & Fridolfsson, A. K. (1997) Male-driven evolution of DNA sequences in birds. *Nat. Genet.* **17,** 182–184.

Filardi, C. E. (2003). Ph.D. dissertation (Univ. of Washington, Seattle).

Filardi, C. E. & Smith, C. E. (2005) Molecular phylogenetics of monarch flycatchers (genus *Monarcha*) with emphasis on Solomon Island endemics. *Mol. Phylogenet. Evol.*, in press.

Gavrilets, S. (2000) Rapid evolution of reproductive barriers driven by sexual conflict. *Nature* **403,** 886–889.

Good, T. P., Ellis, J. C., Annett, C. A. & Pierotti, R. (2000) Bounded hybrid superiority in an avian hybrid zone: Effects of mate, diet, and habitat choice. *Evolution (Lawrence, Kans.)* **54,** 1774–1783.

Gowaty, P. A. (1996) In *Partnerships in Birds*, ed. Black, J. M. (Oxford Univ. Press, Oxford), pp. 21–52.

Grant, B. R. & Grant, P. R. (1979) Darwin's finches: Population variation and sympatric speciation. *Proc. Natl. Acad. Sci. USA* **76,** 2359–2363.

Grant, B. R. & Grant, P. R. (1996) Cultural inheritance of song and its role in the evolution of Darwin's finches. *Evolution (Lawrence, Kans.)* **50,** 2471–2487.

Grant, B. R. & Grant, P. R. (1998) In *Endless Forms: Species and Speciation*, eds. Howard, D. J. & Berlocher, S. H. (Oxford Univ. Press, Oxford).

Grant, B. R. & Grant, P. R. (2002) Simulating secondary contact in allopatric speciation: An empirical test of premating isolation. *Biol. J. Linn. Soc.* **76,** 545–556.

Grant, P. R. & Grant, B. R. (1992) Hybridization of bird species. *Science* **256,** 193–197.

Grant, P. R. & Grant, B. R. (1997) Hybridization, sexual imprinting, and mate choice. *Am. Nat.* **149,** 1–28.

Grant, P. R., Grant, B. R., Markert, J. A., Keller, L. F. & Petren, K. (2004) Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution (Lawrence, Kans.)* **58,** 1588–1599.

Haavie, J., Borge, T., Bures, S., Garamszegi, L. Z., Lampe, H. M., Moreno, J., Qvarnstrom, A., Torok, J. & Saetre, G. P. (2004) Flycatcher song in allopatry and sympatry—Convergence, divergence and reinforcement. *J. Evol. Biol.* **17,** 227–237.

Haldane, J. B. S. (1922) Sex ratio and unisexual sterility in animal hybrids. *J. Genet.* **12,** 101–109.

Hey, J. (2001) *Genes, Categories, and Species: The Evolutionary and Cognitive Causes of the Species Problem* (Oxford Univ. Press, New York).

Hudson, R. R. (1992) Gene trees, species trees, and the segregation of ancestral alleles. *Genetics* **131,** 509–513.

Hudson, R. R. & Coyne, J. A. (2002) Mathematical consequences of the genealogical species concept. *Evolution (Lawrence, Kans.)* **56,** 1557–1565.

Hudson, R. R. & Turelli, M. (2003) Stochasticity overrules the "three-times rule": Genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution (Lawrence, Kans.)* **57,** 182–190.

Hurst, L. D. & Ellegren, H. (1998) Sex biases in the mutation rate. *Trends Genet.* **14,** 446–452.

International Chicken Genome Sequencing Consortium. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432,** 695–716.

International Chicken Polymorphism Map Consortium. (2004) A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* **432,** 717–722.

Irwin, D. E. (2000) Song variation in an avian ring species. *Evolution (Lawrence, Kans.)* **54,** 998–1010.

Irwin, D. E. (2002) Phylogeographic breaks without geographic barriers to gene flow. *Evolution (Lawrence, Kans.)* **56,** 2383–2394.

Irwin, D. E. & Price, T. D. (1999) Sexual imprinting, learning and speciation. *Heredity* **82,** 347–354.

Irwin, D. E., Bensch, S. & Price, T. D. (2001a) Speciation in a ring. *Nature* **409,** 333–337.

Irwin, D. E., Irwin, J. H. & Price, T. D. (2001b) Ring species as bridges between microevolution and speciation. *Genetica* **112,** 223–243.

Jennions, M. D. & Petrie, M. (2000) Why do females mate multiply? A review of the genetic benefits. *Biol. Rev.* **75,** 21–64.

Johnson, N. K. & Cicero, C. (2002) The role of ecologic diversification in sibling speciation of Empidonax flycatchers *(Tyrannidae):* Multigene evidence from mtDNA. *Mol. Ecol.* **11,** 2065–2081.

Kahn, N. W. & Quinn, T. W. (1999) Male-driven evolution among Eoaves? A test of the replicative division hypothesis in a heterogametic female (ZW) system. *J. Mol. Evol.* **49,** 750–759.

Kingan, S. B., Tatar, M. & Rand, D. M. (2003) Reduced polymorphism in the Chimpanzee semen coagulating protein semenogelin I. *J. Mol. Evol.* **57,** 159–169.

Kirkpatrick, M. & Hall, D. W. (2004) Male-biased mutation, sex linkage, and the rate of adaptive evolution. *Evolution (Lawrence, Kans.)* **58,** 437–440.

Lachlan, R. F. & Servedio, M. R. (2004) Song learning accelerates allopatric speciation. *Evolution (Lawrence, Kans.)* **58,** 2049–2063.

Lee, Y. H. & Vacquier, V. D. (1992) The divergence of species-specific abalone sperm lysins is promoted by positive Darwinian selection. *Biol. Bull. (Woods Hole, MA)* **182,** 97–104.

Lijtmaer, D. A., Mahler, B. & Tubaro, P. L. (2003) Hybridization and postzygotic isolation patterns in pigeons and doves. *Evolution (Lawrence, Kans.)* **57,** 1411–1418.

Losos, J. B. & Glor, R. E. (2003) Phylogenetic comparative methods and the geography of speciation. *Trends Ecol. Evol.* **18,** 220–227.

Mayr, E. (1942) *Systematics and the Origin of Species* (Columbia Univ. Press, New York).

Mayr, E. (1963) *Animal Species and Evolution* (Harvard Univ. Press, Cambridge, MA).

Mayr, E. & Diamond, J. (2001) *The Birds of Northern Melanesia: Speciation Ecology, and Biogeography* (Oxford Univ. Press, New York).

Moore, W. S. (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution (Lawrence, Kans.)* **49,** 718–726.

Moritz, C. (1994) Applications of mitochondrial DNA analysis on conservation: A critical review. *Mol. Ecol.* **3,** 401–411.

Moritz, C. (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Syst. Biol.* **51,** 238–254.

Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).

Neigel, J. E. & Avise, J. C. (1986) In *Evolutionary Processes and Theory*, eds. Karlin, S. & Nevo, E. (Academic, New York), pp. 515–534.

Omland, K. E. (1997) Examining two standard assumptions of ancestral reconstructions: Repeated loss of dichromatism in dabbling ducks *(Anatini). Evolution (Lawrence, Kans.)* **51,** 1636–1646.

Palumbi, S. R. (1999) All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proc. Natl. Acad. Sci. USA* **96,** 12632–12637.

Parker, P. G. & Burley, N. T. (1998) *Avian Reproductive Tactics: Female and Male Perspectives* (Am. Ornithol. Union, Washington, DC).

Patten, M. A., Rotenberry, J. T. & Zuk, M. (2004) Habitat selection, acoustic adaptation, and the evolution of reproductive isolation. *Evolution (Lawrence, Kans.)* **58,** 2144–2155.

Payne, R. B. (1996) In *Ecology and Evolution of Acoustic Communication in Birds*, eds. Kroodsma, D. E. & Miller, E. H. (Cornell Univ. Press, Ithaca, NY), pp. 198–220.

Payne, R. B. & Sorenson, M. D. (2004) Behavioral and genetic identification of a hybrid *Vidua*: Maternal origin and mate choice in a brood-parasitic finch. *Auk* **121,** 156–161.

Payne, R. B., Payne, L. L., Nhlane, M. E. D. & Hustler, K. (1993) Species status and distribution of the parasitic indigo-birds *Vidua* in east and southern Africa. *Proc. VIII Pan-Afr. Ornithol. Congr.* 40–52.

Payne, R. B., Payne, L. L. & Woods, J. L. (1998) Song learning in brood parasitic indigobirds *Vidua chalybeata:* Song mimicry of the host species. *Anim. Behav.* **55,** 1537–1553.

Payne, R. B., Payne, L. L., Woods, J. L. & Sorenson, M. D. (2000) Imprinting and the origin of parasite-host species associations in brood-parasitic indigobirds, *Vidua chalybeata. Anim. Behav.* **59,** 69–81.

Pereira, S. L. & Baker, A. J. (2004) Vicariant speciation of curassows (Aves, Cracidae): A hypothesis based on mitochondrial DNA phylogeny. *Auk* **121,** 682–694.

Podos, J. (2001) Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature* **409,** 185–188.

Presgraves, D. C. (2002) Patterns of postzygotic isolation in *Lepidoptera. Evolution (Lawrence, Kans.)* **56,** 1168–1183.

Price, T. D. (1998) Sexual selection and natural selection in bird speciation. *Philos. Trans. R. Soc. London B* **353,** 251–260.

Price, T. D. (2002) Domesticated birds as a model for the genetics of speciation via sexual selection. *Genetica* **116,** 311–327.

Price, T. D. & Bouvier, M. M. (2002) The evolution of F1 postzygotic incompatibilities in birds. *Evolution (Lawrence, Kans.)* **56** 2083–2089.

Rasner, C. A., Yeh, P., Eggert, L. S., Hunt, K. E., Woodruff, D. S. & Price, T. D. (2004) Genetic and morphological evolution following a founder event in the dark-eyed junco, *Junco hyemalis thurberi. Mol. Ecol.* **13,** 671–681.

Reinhold, K. (1998) Sex linkage among genes controlling sexually selected traits. *Behav. Ecol. Sociobiol.* **44,** 1–7.

Rice, W. R. (1996) Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature* **381,** 232–234.

Saetre, G. P., Moum, T., Bures, S., Kral, M., Adamjan, M. & Moreno, J. (1997) A sexually selected character displacement in flycatchers reinforces premating isolation. *Nature* **387,** 589–592.

Saetre, G.-P., Borge, T., Lindroos, K., Haavie, J., Sheldon, B. C., Primmer, C. & Syvänen, A.-C. (2003) Sex chromosome evolution and speciation in *Ficedula* flycatchers. *Proc. R. Soc. London Ser. B* **270,** 53–59.

Schodde, R. & Mason, I. J. (1999) *The Directory of Australian Birds* (CSIRO, Canberra, Australia).

Seddon, N. (2005) Ecological adaptation and species recognition drives vocal evolution in Neotropical suboscine birds. *Evolution (Lawrence, Kans.)* **59,** 97–112.

Sefc, K. M., Payne, R. B. & Sorenson, M. D. (2005) Genetic continuity of brood parasitic indigobird species. *Mol. Ecol.*, in press.

Servedio, M. R. & Sætre, G. P. (2003) Speciation as a positive feedback loop between postzygotic and prezygotic barriers to gene flow. *Proc. R. Soc. London Ser. B* **270,** 1473–1479.

Sites, J. W. & Crandall, K. A. (1997) Testing species boundaries in biodiversity studies. *Conserv. Biol.* **11,** 1289–1297.

Sites, J. W. & Marshall, J. C. (2003) Delimiting species: A Renaissance issue in systematic biology. *Trends Ecol. Evol.* **18,** 462–470.

Slabbekoorn, H. & Smith, T. B. (2000) Does bill size polymorphism affect courtship song characteristics in the African finch *Pyrenestes ostrinus? Biol. J. Linn. Soc.* **71,** 737–753.

Slabbekoorn, H. & Smith, T. B. (2002a) Bird song, ecology and speciation. *Philos. Trans. R. Soc. London Ser. B* **357,** 493–503.

Slabbekoorn, H. & Smith, T. B. (2002b) Habitat-dependent song divergence in the little greenbul: An analysis of environmental selection pressures on acoustic signals. *Evolution (Lawrence, Kans.)* **56,** 1849–1858.

Smith, C. E. (2003) Ph.D. dissertation (Univ. of Washington, Seattle).

Smith, T. B. (1993) Disruptive selection and the genetic-basis of bill size polymorphism in the African finch *Pyrenestes. Nature* **363,** 618–620.

Sorenson, M. D., Sefc, K. M. & Payne, R. B. (2003) Speciation by host switch in brood parasitic indigobirds. *Nature* **424,** 928–931.

Swanson, W. J. & Vacquier, V. D. (2002) Rapid evolution of reproductive proteins. *Nat. Rev. Genet.* **3,** 137–144.

Swanson, W. J., Clark, A. G., Waldrip-Dail, H. M., Wolfner, M. F. & Aquadro, C. F. (2001a) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **98,** 7375–7379.

Swanson, W. J., Yang, Z., Wolfner, M. F. & Aquadro, C. F. (2001b) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc. Natl. Acad. Sci. USA* **98,** 2509–2514.

Swanson, W. J., Nielsen, R. & Yang, Q. (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* **20,** 18–20.

Ting, C. T., Tsaur, S. C. & Wu, C. I. (2000) The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proc. Natl. Acad. Sci. USA* **97,** 5313–5316.

Tishkoff, S. A. & Kidd, K. K. (2004) Implications of biogeography of human populations for "race" and medicine. *Nat. Genet.* **36,** S21–S27.

Tubaro, P. L. & Lijtmaer, D. A. (2002) Hybridization patterns and the evolution of reproductive isolation in ducks. *Biol. J. Linn. Soc.* **77,** 193–200.

Uy, J. A. C. & Borgia, G. (2000) Sexual selection drives rapid divergence in bowerbird display traits. *Evolution (Lawrence, Kans.)* **54,** 273–278.

Via, S. (2001) Sympatric speciation in animals: The ugly duckling grows up. *Trends Ecol. Evol.* **16,** 381–390.

Voelker, G. (1999) Dispersal, vicariance, and clocks: Historical biogeography and speciation in a cosmopolitan passerine genus *(Anthus : motacillidae). Evolution (Lawrence, Kans.)* **53,** 1536–1552.

Wakeley, J. & Hey, J. (1997) Estimating ancestral population parameters. *Genetics* **145,** 847–855.

Wang, P. J., McCarrey, J. R., Yang, F. & Page, D. C. (2001) An abundance of X-linked genes expressed in spermatogonia. *Nat. Genet.* **27,** 422–426.

Wang, Z., Hill, G. E., Baker, A. J. & Edwards, S. V. (2003) Reconciling actual and inferred population histories in the house finch *(Carpodacus mexicanus)* by AFLP analysis. *Evolution (Lawrence, Kans.)* **57,** 2852–2864.

Wiens, J. J. & Penkrot, T. A. (2002) Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (Sceloporus). *Syst. Biol.* **51,** 69–91.

Wolfner, M. F. (1997) Tokens of love: Functions and regulation of *Drosophila* male accessory gland products. *Insect Biochem. Mol. Biol.* **27,** 179–192.

Wu, C. I. & Davis, A. W. (1993) Evolution of postmating reproductive isolation: The composite nature of Haldane's rule and its genetic bases. *Am. Nat.* **142,** 187–212.

Wu, C. I. & Ting, C. T. (2004) Genes and speciation. *Nat. Rev. Genet.* **5,** 114–122.

Yeh, P. J. (2004) Rapid evolution of a sexually selected trait following population establishment in a novel habitat. *Evolution (Lawrence, Kans.)* **58,** 166–174.

Zink, R. M. & McKitrick, M. C. (1995) The debate over species concepts and its implications for ornithology. *Auk* **112,** 701–719.

Zink, R. M., Blackwell-Rago, R. C. & Ronquist, F. (2000) The shifting roles of dispersal and vicariance in biogeography. *Proc. R. Soc. London Ser. B* **267,** 497–503.

7
Critical Review of Host Specificity and Its Coevolutionary Implications in the Fig/Fig-Wasp Mutualism

CARLOS A. MACHADO,* NANCY ROBBINS,† M. THOMAS P. GILBERT,* AND EDWARD ALLEN HERRE†

Figs (*Ficus* spp., Moraceae) and their pollinating wasps (Agaonidae, Chalcidoidea) constitute perhaps the most tightly integrated pollination mutualism that is known. Figs are characterized by extraordinarily high global and local species diversity. It has been proposed that the diversification of this mutualism has occurred through strict-sense coadaptation and cospeciation between pairs of fig and wasp species that are associated in highly specific one-to-one relationships. However, existing studies cast doubt on the generality of this proposition. Here, we review our current knowledge of the evolutionary history of the fig/fig-wasp mutualism. We critically examine the idea that codivergence between figs and their pollinators has been dominated by strict-sense cospeciation. We present phylogenetic and population genetic data from neotropical fig and fig wasp species that suggest that a more accurate model for diversification in this mutualism is that of groups of genetically well defined wasp species coevolving with genetically less well defined (frequently hybridizing) groups of figs. Last, we use our results to assess previously proposed hypotheses on models of speciation in this mutualism.

*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721; and †Smithsonian Tropical Research Institute, Unit 0948, APO AA 34002-0948.

Abbreviations: COI–II, cytochrome oxidase I/II; ML, maximum likelihood.

Data deposition: The nucleotide sequences reported in this paper have been deposited in the GenBank database (accession nos. AY967870–AY968016).

Both ecologically and evolutionarily, mutualisms represent one of the most influential of all biological interactions, with fundamental consequences for the evolution and maintenance of biotic diversity (Boucher, 1985; Bronstein, 2001; Douglas, 1994; Herre *et al.*, 1999; Margulis and Fester, 1991; Maynard Smith and Szathmáry, 1995; Thompson, 1994). The long-term stability of mutualisms poses a considerable and, as yet, not fully resolved challenge to evolutionary theory. However, the obvious fact of long-term stability coupled with the proliferation and diversification of many mutualisms raises a set of interesting questions concerning coadaptation and speciation among the partners in the interaction. The obligate mutualisms between flowering plants and their insect pollinators (Corner, 1952; Kato *et al.*, 2003; Pellmyr, 2003; Wiebes, 1979) constitute fascinating extreme cases of interspecific mutualisms. Most obligate plant–pollinator mutualisms show high levels of reciprocal species or taxon specificity. Usually, the insect requires the plant for food or other substances to complete its life cycle successfully, and the plant requires the insect for pollination. Further, it is the insect's recognition and choice of hosts that determine the patterns of host gene flow. Although there are relatively few cases of obligate pollination mutualisms (Corner, 1952; Kato *et al.*, 2003; Pellmyr, 2003; Wiebes, 1979), these few cases are often marked by high to extreme speciation and diversification in both partners, raising the question of how host specificity and control of gene flow affects patterns of speciation in one or both partners.

Figs (*Ficus* spp., Moraceae) and their pollinating wasps (Agaonidae, Chalcidoidea) constitute perhaps the most tightly integrated pollination mutualism that is known (Cook and Rasplus, 2003; Corner, 1952; Galil and Eisikowitch, 1968; Janzen, 1979; Ramirez, 1970a; Weiblen, 2002; Wiebes, 1979). *Ficus* is one of the most diverse genera of flowering plants in number of species and growth and life forms (Berg and Wiebes, 1992; Harrison, 2005). The nearly 750 described species of *Ficus* (Berg, 1989) occur worldwide in tropical and subtropical regions, and they are considered ''keystone'' species in tropical forests because of their year-round production of fruit that is essential to a large number of frugivores (Kalko *et al.*, 1996; Korine *et al.*, 2000; McKey, 1989; Terborgh, 1986). Figs depend on minute, pollen-bearing female wasps to pollinate the flowers and thereby initiate seed production (Corner, 1952; Eisikowitch, 1968; Galil, 1977; Galil and Ramirez, 1974; Herre, 1989, 1999; Herre and West, 1997). The mated female wasps, in turn, depend on the developing fig inflorescence for the production of their offspring, because each wasp larva consumes the contents of one would-be seed. The cycle begins when mated female wasps locate a receptive tree and enter the enclosed fig inflorescences (Syconia). As the females search for oviposition sites, they pollinate the flowers. Usually the foundresses die inside the syconium, and

then both their offspring and the seeds begin to develop (Galil, 1977; Herre, 1989, 1999; Herre and West, 1997; Ramirez, 1974). Last, after maturation, the offspring mate and the mated females collect pollen and fly off to find a receptive tree and begin the cycle anew.

Results from morphological studies and the notion of the high (one-to-one) species specificity of the interaction led to the proposal of strict-sense coevolution and tight cospeciation between the two groups (Berg, 1989; Berg and Wiebes, 1992; Ramirez, 1974; Wiebes, 1982, 1987). Recent molecular phylogenetic studies (Herre *et al.*, 1996; Jousselin *et al.*, 2003; Machado *et al.*, 2001; Weiblen, 2001, 2004; Yokoyama, 1995;) have provided some support for the proposition of cocladogenesis and coadaptation between recognized genera of pollinating wasps and their respective fig sections (species groups within fig subgenera). Results from those studies (as well as earlier morphological studies) have led some authors to suggest that finer-scale cospeciation of individual fig and wasp species should be widespread (Berg and Wiebes, 1992; Ramirez, 1974; Weiblen, 2002: Wiebes, 1979).

However, even at this fairly coarse level of wasp genera and fig sections, clear incongruencies among their phylogenies have been detected, directly undermining the empirical basis of strict-sense cospeciation (Jousselin *et al.*, 2003; Machado *et al.*, 2001; Weiblen, 2000, 2001). Also, existing phylogenetic studies are not adequate to clearly distinguish between strict-sense cospeciation between individual pairs of species-specific wasps and figs, and a much less specific form of broad-sense coevolution between related groups of wasps with related groups of figs. This difference is crucial with respect to understanding the actual mechanisms of coadaptation and cospeciation in the mutualism. If strict-sense cospeciation were the rule, then the evolutionary dynamics of coadaptation would be expected to take place independently in a series of genetically isolated, tightly coupled pairs of fig and pollinator wasp species. If strict-sense cospeciation were not the rule, then the dynamics of coadaptation would be quite different. Specifically, under this latter scenario, we expect a much looser coevolutionary coupling between particular pairs of figs and wasps, with host colonization, hybridization, and introgression of relatively novel genetic material across different fig species contributing to the generation of phenotypic diversity on which selection can act.

Here, we evaluate the hypothesis that strict-sense cospeciation has dominated the evolutionary history of figs and their pollinators. We review the current knowledge on the evolutionary history of the mutualism, focusing on the available evidence suggesting cocladogenesis at high taxonomic levels (i.e., between different sections of *Ficus* and their associated genera of pollinating wasps) and cospeciation at finer taxonomic scales. We present data from neotropical fig and fig wasp species that

suggest an alternative view. Specifically, we find that groups of genetically well defined pollinator wasp species coevolve in association with groups of genetically poorly defined (hybridizing) figs. Last, we link our results to previously proposed models of speciation in this mutualism that bear directly on the question of the origin of the high species diversity in these organisms.

THE COEVOLUTIONARY HISTORY OF THE MUTUALISM

The fig-wasp mutualism is ancient and diverse, originating ≈80–90 million years ago (Datwyler and Weiblen, 2004; Machado *et al.*, 2001). The outgroup to figs is not yet clear and includes both New and Old World taxa as possibilities (Datwyler and Weiblen, 2004). However, the hypothesis of a Gondwanan origin of the mutualism (Corner, 1958; Murray, 1985) is supported by the observations that the most basal group to extant figs and pollinator wasp are, respectively, the New World Ficus subgenus *Pharmacosycea* and their associated pollinators in the genus, *Tetrapus* (Berg, 1989; Herre *et al.*, 1996; Jousselin *et al.*, 2003; Machado *et al.*, 2001; Ramirez, 1974; Weiblen, 2000). The idea that extant figs and wasps have radiated from these basal New World groups is supported further by the observation that the estimated divergence times among the pollinator genera and their current geographical distributions correspond well with several features of the break up of the southern continents during the late Cretaceous period (Machado *et al.*, 2001).

The nearly 750 described species of *Ficus* have been classified in four subgenera (*Pharmacosycea*, *Urostigma*, *Sycomorus*, and *Ficus*) and 18 sections (species groups within subgenera) (Berg, 1989; Berg and Wiebes, 1992; Corner, 1965). The pollinators of figs, minute chalcid wasps from family Agaonidae, have been assigned to 20 different genera (Berg and Wiebes, 1992; Wiebes, 1994), with >300 described species (Weiblen, 2002). With two exceptions (*Ceratosolen* and *Wiebesia*), each recognized pollinator genus is generally restricted to a single taxonomic group of *Ficus* (subgenus or section). Molecular data show that fig-pollinating wasps form a clearly monophyletic group within the Chalcidoidea that is separate from other fig wasps (a large number of nonpollinating wasps are also associated with figs), supporting the idea that the pollination syndrome evolved only once during the evolution of the mutualism (Machado *et al.*, 1996; Rasplus *et al.*, 1998).

The original taxonomic studies of the fig-pollinating wasps using morphological characters (Ramirez, 1974, 1991; Wiebes, 1982, 1994) showed that related species of wasps pollinate related species of figs. This result led to the hypothesis of strict-sense coevolution between the mutualists. Also, the almost invariable finding at that time of one species of

pollinator per host fig species reinforced this idea, and it led to the inference that cospeciation should, therefore, be the dominant pattern of codivergence at both coarser (e.g., sections of figs and genera of wasps) and finer (species of both) taxonomic scales (Berg and Wiebes, 1992; Corner, 1985; Ramirez, 1974; Wiebes, 1979). Recent molecular phylogenetic studies of the major groups of fig-pollinating wasps (Machado *et al.*, 2001; Weiblen, 2001) and figs (Herre *et al.*, 1996; Jousselin *et al.*, 2003; Weiblen, 2000) have allowed mapping the morphological classification of figs into the fig-wasp phylogeny and vice versa. These studies have provided some support to the original insight from taxonomy about the strong conservation of host associations and the preponderance of cocladogenesis during the diversification of wasp genera and associated fig subgenera and sections. However, these studies have also revealed clear incongruencies between the phylogenies of figs and those of their pollinators (Fig. 7.1) (Jousselin *et al.*, 2003; Machado *et al.*, 2001; Weiblen, 2000, 2001).

In Fig. 7.1, we present a pollinator phylogeny based on cytochrome oxidase I (*COI*) data from Machado *et al.* (2001). This tree is largely consistent with results for the pollinators obtained by Weiblen (Weiblen, 2001). The corresponding fig phylogeny is a strict consensus from maximum-parsimony analyses of combined sequences of the internal and external transcribed spacers (*ITS* and *ETS*) from a large worldwide sample of *Ficus*, and is largely consistent with results from Jousselin *et al.* (2003). This phylogeny is also largely consistent with results from a different study on Indo-Australian *Ficus* taxa from Weiblen using *ITS* sequences (Weiblen, 2000). Specifically, the well resolved relationships among sections *Ficus*, *Sycidium*, *Rhizocladus*, and *Kalosyce* were incorporated from that study (Weiblen, 2000). The degree of congruency between the two cladograms shown in Fig. 7.1 is not significant ($P = 0.85$) (Page, 1994), a result that is not consistent with previous assertions about the preponderance of cocladogenesis. However, this result should be viewed cautiously because of the lack of resolution and uncertainty of relationships of several fig sections (*Urostigma* and *Oreosycea*) and pollinator genera (*Wiebesia*, *Blastophaga*, and *Platyscapa*) (Herre *et al.*, 1996; Jousselin *et al.*, 2003; Machado *et al.*, 2001; Weiblen, 2000, 2001), and to the partial sampling of wasp genera (11 of 20) and fig sections (12 of 18) in the phylogenies.

Nonetheless, despite these caveats, the phylogenetic reconstructions show that ancestral host switches have occurred at different times during the evolution of the mutualism, even when considering the strongly supported clades. For example, wasps that pollinate the figs in subgenera *Ficus* and *Sycidium* are not close relatives (Machado *et al.*, 2001; Weiblen, 2001), although both morphological (Berg, 1989; Berg and Wiebes, 1992) and molecular (Herre *et al.*, 1996; Jousselin *et al.*, 2003; Weiblen, 2000) studies indicate that *Ficus* and *Sycidium* are sister groups. Further, molecular

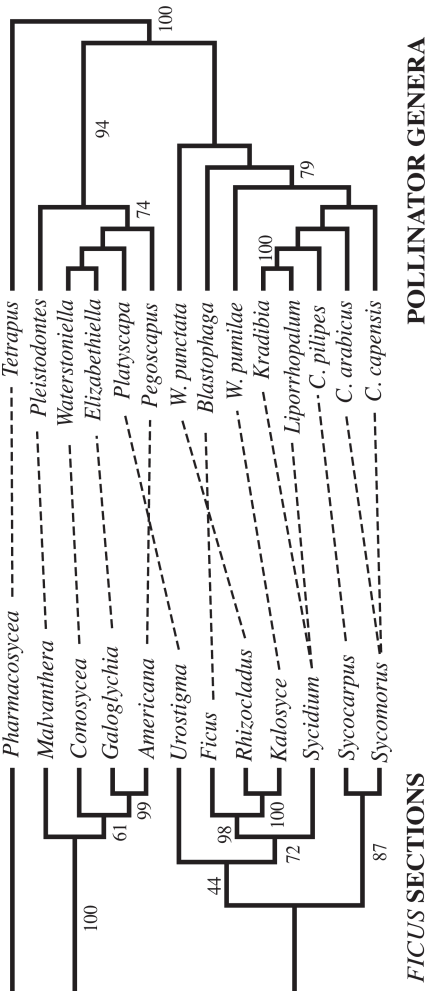


FIGURE 7.1 Tanglegram comparison of the phylogenies of sections of *Ficus* and their associated genera of pollinating wasps. Both phylogenies correspond to pruned trees from larger phylogenetic studies (Jousselin *et al.*, 2003; Machado *et al.*, 2001; Weiblen, 2000). In the pollinator phylogeny, C stands for *Ceratosolen* and W stands for *Wiebesia,* the only genera of pollinating wasps associated with more than one *Ficus* section. Numbers associated with branches are bootstrap values (>40%) that were taken directly from the studies in which the complete phylogenies were presented originally (Jousselin *et al.*, 2003; Machado *et al.*, 2001; Weiblen, 2000). They are shown to indicate the level of support for each node in the original phylogenetic studies. The two cladograms are significantly incongruent (see text). TREEMAP 1.0 was used to create this image (Page, 1994) (see text for details).

data suggest that the pollinators of *Ficus* are more closely related to the pollinators of *Urostigma* (Weiblen, 2001). Because figs from the subgenera *Ficus* and *Sycidium* are supported to be related more closely to *Sycomorus* than to *Urostigma*, the most likely scenario is that the ancestor of *Ficus* figs was colonized by the ancestor of wasps associated currently with *Urostigma* figs and that this new combination then jointly diversified. Thus, even at high taxonomic scales, the evidence indicates at least one breakdown (and possibly more) in strict cocladogenesis in the mutualism.

FINER-SCALE COEVOLUTION: IS THERE EVIDENCE FOR STRICT-SENSE COSPECIATION?

A major weakness of existing molecular phylogenetic studies in addressing the degree to which strict-sense cospeciation dominates the evolutionary dynamics of this mutualism is that they typically have concentrated on a small number of taxa that represent very ancient, distantly related taxonomic subdivisions within the genus *Ficus* and their associated genera of pollinator wasps (Herre *et al.*, 1996; Jousselin *et al.*, 2003; Machado *et al.*, 2001; Weiblen, 2001). Also, these studies have also tended to concentrate on analyses of one gene (or a few genes) from a few individuals of each species. Such sampling differentially emphasizes the products of ancient processes and inevitably tends to bias interpretation toward cospeciation, without providing a real and rigorous test of the hypothesis. Specifically, even if there were a perfect congruence of fig and wasp phylogenies at coarse scales (and available evidence suggests that there is not), such a pattern would not necessarily result from strict-sense, one-to-one cospeciation of fig and wasp species. Also, the sampling of multiple loci from multiple individuals within a species is essential to detect potentially important ongoing processes such as hybridization and introgression. Therefore, to find compelling evidence for, or against, strict-sense cospeciation as a real, ongoing process, detailed genetic sampling of multiple individuals from multiple, relatively closely related species of the same section of fig and genus of pollinator is required.

The only published coevolutionary study that has attempted to survey relationships at a finer (within section/genus) taxonomic scale (Weiblen and Bush, 2002), reconstructed molecular phylogenies for 17 Indo-Australian and two African fig species from subgenus *Sycomorus* and their pollinators (*Ceratosolen* sp.). Significant congruency between the phylogenies of the two groups supported the cospeciation hypothesis (Weiblen and Bush, 2002). However, although some of the sampled species appear to be close relatives based on their taxonomic associations (e.g., they are part of the same *Ficus* section), most species included in this sample are very distinct genetically, and speciation events appear to be

ancient (Machado *et al.*, 2001). Also, within species genetic sampling was minimal or not reported. Therefore, although that study represents a step in the right direction, it does not provide the kind of information necessary to understand species-level processes of diversification in the mutualism and to adequately test coevolutionary hypotheses.

We have begun to fill that gap by gathering detailed genetic data from a group of 17 sympatric species of figs and their associated pollinators found in the vicinity of the Panama Canal. In particular, we have collected information from multiple loci across multiple individuals within these taxa. These figs and wasps also represent two of the most divergent lineages of *Ficus* and their pollinators (Berg, 1989), and a series of extensive, long-term studies have made them the ecologically and genetically best characterized group of species from this mutualism (Herre, 1989, 1996; Herre *et al.*, 2001; Molbo *et al.*, 2003; Nason *et al.*, 1998). Free-standing neotropical monoecious fig trees (subgenus *Pharmacosycea*, section *Pharmacosycea*) are the most basal of the extant fig lineages, with 20 described species (Berg, 1989). They are pollinated by wasps of the genus *Tetrapus*, the most basal group of pollinators (Fig. 7.1). The second lineage is a derived section of strangling monoecious figs (subgenus *Urostigma*, section *Americana*) and their pollinating fig wasps (*Pegoscapus* sp.), with ≈100–120 described species (Berg, 1989). Of the 17 Panamanian fig species we have studied, 4 belong to section *Pharmacosycea* and 13 belong to section *Americana*.

We reconstructed molecular phylogenies of the 17 species of Panamanian figs and their pollinators (Fig. 7.2) using 1,587 bp of *COI–COII* from the pollinators and the following two nuclear regions from the figs: 530 bp from an intron of the triosephosphate isomerase gene (*Tpi*) and 723 bp of sequence encompassing four introns and three exons from the glycerol-3-phosphate dehydrogenase gene (*G3pdh*). The primers used for PCR and sequencing have been described elsewhere (Machado *et al.*, 2001; Simon *et al.*, 1994; Strand *et al.*, 1997). PCR products were sequenced directly in both directions by using standard protocols. If the *Tpi* or *G3pdh* sequence from an individual was polymorphic, the PCR product was cloned with the TOPO TA cloning kit (Invitrogen), and 8–10 cloned PCR fragments sequenced to identify the different haplotypes. Multiple individuals per species were sampled (see below), but the phylogenies used for reconciliation analyses (Page, 1994) were reconstructed by using either the most common haplotype per species or a randomly chosen individual sequence (in a few cases, all haplotypes had the same frequency). Results from the reconciliation analyses are similar if different combinations of haplotypes are used (data not shown). Phylogenies were reconstructed using maximum-likelihood (ML) methods and the general reversible model with rate heterogeneity (REV+Γ) using of PAUP* (version 4.0b10) (Swofford,

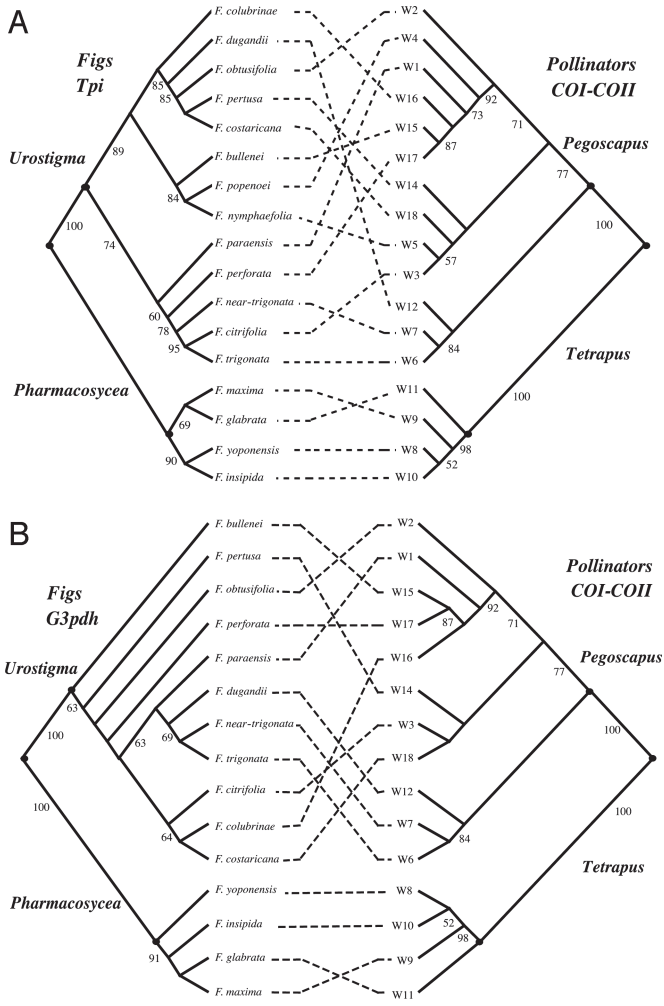


FIGURE 7.2 Tanglegram comparisons of molecular phylogenies from neotropical *Ficus* and their pollinators. Species names for the wasps are omitted for brevity and are shown as codes, with each number corresponding to a different species. All phylogenies are ML trees (see text for details). The individual sequences per species used in these reconstructions correspond to the most common haplotypes found for each species. Fig wasp sequences from fig species with multiple pollinators are from the most common species associated with a givenfighost. The pollinator phylogeny compared with the *G3pdh* phylogeny is a pruned version of the larger *COI–COII* tree compared with the *Tpi* phylogeny. Numbers below branches are bootstrap values based on 1,000 replications. The fig and pollinator phylogenies are significantly incongruent (see text).

1998). Topological comparisons were conducted by using the Shimodaira–Hasegawa test (one-tailed test) (Shimodaira and Hasegawa, 1999) with Resampling of Estimated Log Likelihoods (RELL) bootstrap (1,000 replications). The degree of match between the *Tpi* ML fig phylogeny [–log (*L*) = 1,095.06375; $\alpha = \infty$] and the *COI–COII* pollinator ML phylogeny [–log (*L*) = 8,236.58884; $\alpha = 0.572$] is not significantly different from what would be expected by chance ($P = 0.25$; Fig. 7.2*A*). Also, forcing the fig *Tpi* phylogeny to match the pollinator phylogeny generates a tree that is significantly worse than the ML tree [Δ –log (*L*) = 141.9446; $P < 0.001$]. The *G3pdh* ML phylogeny reconstructed with the REV+Γ model [–log (*L*) = 1,385.80552; $\alpha = 0.847$] and the corresponding *COI–COII* pollinator phylogenies are shown in Fig. 7.2*B*. The *G3pdh* and *COI–COII* phylogenies are incongruent ($P = 0.09$), and forcing the topology of the *G3pdh* phylogeny to match the pollinator phylogeny also generates a tree that is significantly worse than the ML tree (Δ –log (*L*) = 28.29731, $P = 0.025$). Therefore, despite the low resolution for several nodes in the fig phylogenies (Fig. 7.2), the phylogenies of two different nuclear genes show significant incongruency with the phylogenetic history of the associated pollinators.

These results suggest that cospeciation is not the dominant pattern of codivergence in these groups of sympatric neotropical figs and their pollinators. We suggest that the most likely explanation for the lack of congruency between the molecular phylogenies of the two mutualists is the recently documented presence of multiple species of pollinator per host in neotropical figs (Molbo *et al.*, 2003), which would generate incongruent phylogenetic histories because of pollinator host switches, with the resultant hybridization and possible genetic introgression across different fig species (see below).

A COMPLEX MUTUALISM UNVEILED? MULTIPLE POLLINATORS PER HOST ARE NOT THE EXCEPTION

The degree to which tight specificity of a particular wasp species to a given host species breaks down will influence the possibility of naturally occurring hybridization among different host species and affect the trajectories of fig and wasp coadaptation. The notion that the vast majority of species of fig are pollinated by its own highly specific species of pollinator (''the one-to-one rule'') has long been claimed as one of the most extraordinary aspects of this mutualism (Cook and Rasplus, 2003; Weiblen, 2002; Wiebes, 1979). Initial documented cases of breakdown of species specificity were thought to represent taxonomic artifacts (Wiebes, 1963). Nonetheless, to determine the extent to which tight cospecificity reflects a real and predominant pattern rather than an assumption requires detailed taxonomic and genetic studies.

Several recent and detailed studies have confirmed that this rule is often broken (Berg and Wiebes, 1992; Cook and Rasplus, 2003; Molbo *et al.*, 2003; Rasplus, 1996; Weiblen, 2002). Recent studies in Africa and Australia have shown that breakdowns of specificity are relatively common, occurring in approximately one third of the studied cases (24 species of *Ficus*) (Cook and Rasplus, 2003; Kerdelhué *et al.*, 1999; Lopez-Vaamonde *et al.*, 2002). In some cases, different wasp species are associated only with a host fig in different parts of its geographic range (Michaloud *et al.*, 1985, 1996; Rasplus, 1996). Importantly, it has been demonstrated in other cases that multiple pollinators associate routinely and successfully with a given host in sympatry (Compton, 1990; Kerdelhué *et al.*, 1997, 1999; Lopez-Vaamonde *et al.*, 2002; Molbo *et al.*, 2003; Ramirez, 1970a,b; Wiebes, 1994).

Moreover, several independent observations by different researchers in different geographic regions suggest the routine occurrence of interspecific hybridization in figs. Frequent cases of adult figs with hybrid morphological characteristics have been observed in Panama and South America (Berg, 1989) (W. Ramirez, personal communication; C.C. Berg, personal communication; E.A.H, unpublished data), Africa (S. Compton, personal communication; G. Michaloud, personal communication), and Indonesia (Parrish *et al.*, 2003). Reports of local pollinators breeding in introduced fig species have been reported in Florida (Ramirez, 1994), where hybrid viable seedlings were also observed. Also, Parrish *et al.* (2003) recently reported genetic evidence of hybridization among three dioecious fig species found in three Indonesian islands. The observation of multiple individuals with hybrid genetic composition at several amplified fragment length polymorphism loci was interpreted to be caused by a breakdown of pollinator specificity in the isolated island populations due to pollinator limitation. Last, a detailed study from South Africa (Compton, 1990) found that a fig tree (*Ficus lutea*) growing >100 km out of its normal range was pollinated by several different species of wasps that are not normally associated with it. Likely relevant to understanding the broader mechanisms that generate diversity in figs, this study found that the species that were most closely related to the normal pollinator of this host species showed a higher capacity to reproduce and produce seeds that would germinate (Compton, 1990, 1993; Ware and Compton, 1992).

Of all documented violations of the one-to-one rule, the most thorough sampling of genetic data comes from a group of sympatric New World fig species (Molbo *et al.*, 2003). Extensive genetic sampling of pollinators associated with a subset of the 17 sympatric neotropical fig species from Fig. 7.2, shows that the rule is broken in ≈60% of the cases (Molbo *et al.*, 2003) (Table 7.1). In some cases, the species pairs associated with the same host are each other's closest relatives, a finding that is consistent with long-term coexistence on a single host. In other cases, however, the

TABLE 7.1 Genetically Confirmed Cases of Multiple Pollinators per *Ficus* in Neotropical Species

Ficus species	No. of pollinators	Shared pollinators
F. bullenei[a]	4	*F. popenoei, F. near-trigonata*
F. near-trigonata[a]	3	*F. popenoei, F. bullenei*
F. popenoei[a]	2	*F. near-trigonata, F. bullenei*
F. perforata[a]	2	*F. colubrinae*
F. obtusifolia[a]	2	—
F. glabrata[b]	2	—
F. maxima[b]	2	—

NOTE: Other species have been sampled (>10 individuals) without revealing the presence of multiple pollinators: *F. citrifolia, F. nymphaefolia, F. trigonata, F. dugandii,* and *F. colubrinae.*
[a]*Ficus* from subgenus *Urostigma.*
[b]*Ficus* from the subgenus *Pharmacosycea.* Sample sizes for those two species are small (<10 individuals).

species pairs associated with the same host are not each other's closest relatives, indicating a host switch. Importantly, wasps that are genetically indistinguishable regularly pollinate different host fig species (Table 7.1). The molecular phylogeny of the neotropical pollinators presented in Fig. 7.2, was reconstructed by using the most common haplotype of each wasp species found in our survey and, therefore, represents the phylogeny of the most common pollinators of the 17 neotropical figs included, based on our genetic sampling. Inclusion of the multiple pollinators in the cophylogenetic analyses does not generate a better fit between the fig and pollinator phylogenies (*Tpi*, $P = 0.33$; *G3pdh*, $P = 0.15$).

These findings reveal previously unsuspected levels of complexity in the mutualism, and they seriously challenge the idea that strict-sense cospeciation should be the dominant pattern of codivergence among figs and their pollinators. The inferred host switches and the observation of different fig hosts sharing identical pollinators suggests the potential for hybridization and genetic introgression between host fig species (see below), a situation that could generate incongruent histories of divergence between the mutualists. Thus, our observations provide a clear mechanism for explaining the incongruent phylogenetic histories of neotropical figs and their pollinators (Fig. 7.2), and they also provide a potential mechanism to explain phylogenetic incongruencies at higher taxonomic levels (Fig. 7.1).

ASSESSING THE EVIDENCE OF HYBRIDIZATION AND INTROGRESSION AMONG NEOTROPICAL FIG SPECIES

To test for evidence of introgression in neotropical figs, we conducted a divergence population genetic study (Kliman *et al.*, 2000; Machado *et al.*, 2002) in three species of neotropical *Urostigma* figs: *Ficus popenoei*, *Ficus bullenei*, and *Ficus near-trigonata*. These three species share one pollinator, and also, each species has one or two additional pollinators that, based on our current sample sizes, are not shared with other fig species (Table 7.1). Thus, we expect to observe some evidence of past or current introgression between these species pairs. We collected DNA polymorphism data for these three species (Table 7.2) by using the two nuclear loci described above (*Tpi* and *G3pdh*) and the locus *C2E19*, an EST isolated from a cDNA library made from leaf tissue of a neotropical *Urostigma* fig species (*Ficus citrifolia*). The locus *C2E19* has no clear homologues in GenBank, and therefore, we consider it here to be an ORFan (an expressed gene with unknown function). PCR products were directly sequenced in both directions by using standard protocols, and polymorphic sequences were cloned to distinguish haplotypes.

The polymorphism data were fitted to a null model of speciation with no gene flow (isolation model) by using the method described by Wang, Wakeley, and Hey (WWH) (1997). In addition to the WWH method, we also determined the fit of the fig nucleotide data to the isolation model by

TABLE 7.2 Polymorphism Statistics for Three Nuclear Loci Sequenced in Three Species of Neotropical Ficus (sect. *Urostigma*)

Locus	Species	n[a]	L[b]	S[c]	$\hat{\theta}$[d]	π[e]	D[f]
Tpi	*F. popenoei*	16	517	2	0.00117	0.00193	1.6871
	F. near-trigonata	10	521	30	0.02034	0.01858	–0.4146
	F. bullenei	10	517	14	0.00957	0.00606	–1.6802
G3pdh	*F. bullenei*	10	719	1	0.00049	0.00049	0.0150
	F. near-trigonata	12	719	3	0.00138	0.00148	0.2219
	F. bullenei	8	720	2	0.00107	0.00070	–1.3101
C2E19[g]	*F. popenoei*	9	704	13	0.00679	0.00458	–1.5570
	F. near-trigonata	8	704	13	0.00712	0.01040	2.3306[h]
	F. bullenei	11	704	3	0.00145	0.00232	2.0454[h]

[a]No. of alleles sequenced.
[b]Average length (bp) of the sequences from each species.
[c]No. of polymorphic sites.
[d]Estimate of 4Nμ per bp using the no. of polymorphic sites (Watterson, 1975).
[e]Estimate of 4Nμ by using the average no. of nucleotide differences per site (Nei, 1987).
[f]Tajima's statistic (Tajima, 1989).
[g]ORFan locus isolated from a cDNA library.
[h]Significantly different from zero ($P < 0.01$).

comparing observed to expected values of different polymorphism types (shared, exclusive, and fixed differences) by using a χ^2 statistic (Kliman *et al.*, 2000). Under the isolation model, two very recently diverged species are expected to share polymorphisms that are present in their ancestor. As the species diverge, shared polymorphisms are lost, exclusive polymorphisms arise, and fixed genetic differences accumulate in each lineage through genetic drift. With observations from a number of loci, one expects to find a consistent negative correlation between fixed differences and shared polymorphisms across loci. However, if the null model of speciation is not correct and gene flow has occurred during species divergence, the divergence of a given locus will be slowed down because gene flow removes and prevents the accumulation of fixed differences at the same time as it introduces shared polymorphisms. Gene flow will elevate the numbers of shared polymorphisms and reduce both the number of exclusive polymorphisms and the number of fixed differences between taxa. More importantly, if gene flow occurs at some loci but not at others, it will elevate the variance among loci in the numbers of shared polymorphisms and fixed differences. These ideas form the basis of the WWH method (Wang *et al.*, 1997). This method is conservative and takes into account the number of exclusive polymorphisms for each species, and the numbers of shared polymorphisms and fixed differences between the species. These polymorphism counts are used to fit the isolation model to the data and estimate the four parameters of the model: the three population mutation rates (one for the ancestral population and one for each descendant), and the time since divergence. By using these four parameters and estimates of the recombination rate for each locus, it is then possible to conduct coalescent simulations to determine the significance of the WWH (Wang *et al.*, 1992) and χ^2 (Kliman *et al.*, 2000) test statistics.

The fit of the fig polymorphism data to the isolation model is poor (Table 7.3), and significantly so for the *F. popenoei/F. bullenei* and *F. popenoei/F. near-trigonata* comparisons. These results suggest the occurrence of gene flow among these two species pairs during their divergence. Based on this small data set, we can infer that gene flow has occurred but has not occurred recently because we see no evidence of shared identical haplotypes across species at any of the three studied loci.

An assumption of these tests is that the DNA sequences should be evolving neutrally. We first determined whether the amounts of polymorphism and divergence across loci are correlated, as expected under neutrality, by using the Hudson–Kreitman–Aguade (HKA) test (Hudson *et al.*, 1987). The significance of the χ^2 statistic from this test was evaluated by using a distribution generated from 1,000 coalescent simulations. HKA tests were applied to each of the three taxa, in each case by using a single sequence from *F. citrifolia* as an outgroup. The data fit the neutral model

TABLE 7.3 Fitting *Ficus* Sequence Data to an Isolation Model of Species Divergence

Species 1	Species 2	θ_1	θ_2
F. popenoei	*F. near-trigonata*	15.91 (4.84–59.57)	5.37 (1.95–10.78)
F. popenoei	*F. bullenei*	5.512 (2.38–2.03)	6.828 (3.25–22.67)
F. near-trigonata	*F. bullenei*	9.377 (0–37.51)	3.560 (0–10.45)

The 95% confidence intervals are shown in parentheses. θ_A is the estimate of the population mutation parameter for the ancestor of species 1 and 2. T is the estimated time of divergence between the two species in $2N_1$ generation units (where N_1 is the estimate of the

in all three species: *F. popenoei*, $\chi^2 = 1.40$ and $P = 0.212$; *F. neartrigonata*, $\chi^2 = 2.23$ and $P = 0.103$; and *F. bullenei*, $\chi^2 = 1.73$ and $P = 0.128$. A second type of neutrality tests evaluates the fit of the data to the neutral model by using DNA polymorphism data from a single locus (Fu and Li, 1993; Tajima, 1989). These tests evaluate whether the frequency spectrum of mutations is significantly different from that expected under neutrality. Tajima's D statistic (Tajima, 1989) is proportional to the difference between two estimates of the population mutation parameter $4N\mu$, the mean pairwise difference between the sampled sequences (π) and Watterson's estimator (θ) (Watterson, 1975). Under a neutral model with constant population size, both estimators have the same expected value, and therefore, the value of Tajima's D under neutrality is zero. The observed values of Tajima's D for the *C2E19* sequences of *F. near-trigonata* ($D = 2.3306$) and *F. bullenei* ($D = 2.0454$) are significantly positive (Table 7.2). Significant positive values of Tajima's D are usually interpreted as evidence of balancing selection (Tajima, 1989). However, in the case of *F. near-trigonata*, that result is likely to be caused by introgression: groups of *F. near-trigonata C2E19* alleles are more closely related to *F. popenoei* alleles than to other conspecific alleles (data not shown), a result consistent with the poor fit of the data from that species pair to a strict isolation model of species divergence. Introgression generates patterns of polymorphism that mimic those produced by natural selection (Sweigart and Willis, 2003; Wright *et al.*, 2003). In the case of *F. bullenei*, balancing selection could be the explanation for the large positive value of Tajima's D. However, we question that inference because an alternative test of neutrality that also uses DNA polymorphism from a single locus, Fu and Li's D (Fu and Li, 1993), is not significant ($D = 1.1284$).

Our findings on the pollinators of these fig species contrast with those of their hosts. None of the available data from microsatellite loci (Molbo *et al.*, 2003, 2004) or mitochondrial sequences (Molbo *et al.*, 2003) show any evidence of genetic introgression across different species of pollinator. Allele size distributions of microsatellites for different cryptic pollinator

θ_A	T	χ^2	$P\chi^2$	WWH	P_{WWH}
15.561 (0–33.34)	0.436 (0.11–1.42)	70.02	<0.001	11	0.079
0 (0–0.002)	2.442 (0.49–2.82)	34.33	0.001	9	0.016
27.354 (0–47.51)	0.397 (0.14–1.82)	36.03	0.088	11	0.333

effective population size of species 1). The *P* values for both the χ^2 and the WWH test statistics are the proportion of simulated values greater than or equal to the observed value.

species are nonoverlapping (Molbo *et al.*, 2003, 2004). Nucleotide sequences of mitochondrial genes (Molbo *et al.*, 2003) from different species are reciprocally monophyletic. These findings suggest that the process of divergence in figs and in the pollinators is very different. Thus, whereas figs may exchange genes during divergence because of nonspecific pollinators, the pollinators do not, and thus, they seem to be reproductively isolated species that satisfy a strict interpretation of the biological species concept.

A second potential reason could partially explain the observed phylogenetic incongruencies among neotropical figs and their pollinators (Fig. 7.2): the large effective population size (*Ne*) that is believed to characterize neotropical figs (Nason *et al.*, 1998). Large *Ne* could interfere with species-level phylogenetic analyses because of the expected large amounts of ancestral polymorphism shared across recently diverged species. The slow lineage sorting of speciesspecific alleles into reciprocally monophyletic clades generates the well known ''gene tree vs. species tree'' problem (Pamilo and Nei, 1988; Tajima, 1983), leading to erroneous phylogenetic inference for recently diverged lineages and to low resolution of species-level phylogenies. However, we believe that this explanation is unlikely for three reasons. First, this issue is effectively considered by the test of the strict isolation model. The test seeks to determine whether the observed level of shared polymorphisms, exclusive polymorphisms, and fixed differences can be explained based on the estimates of the four basic parameters of the model (the population sizes of the two species and their ancestor, and the time of divergence). The lack of fit of the data to the model suggests that shared variation is the result of gene flow and not of ancestral polymorphism. Second, the data from *G3pdh*, used to test cospeciation (Fig. 7.2*B*), shows 7–10 fixed differences between the three species pairs and not a single shared polymorphism. Thus, the significant incongruency among the fig *G3pdh* phylogeny and the pollinator *COI–COII* phylogeny is clearly not due to unsorted ancestral polymorphisms. Third, although the nucleotide variability in neotropical fig species is

expected to be high based on results from protein electrophoresis studies (Nason *et al.*, 1998), Table 7.2 shows that the observed level of DNA polymorphism in these species is not uncommonly high for the three surveyed genes. In fact, DNA polymorphism in these three species is much lower than that observed in other outbred plant species. Average values of θ, Watterson's estimator of the population mutation rate (4Nμ) (Watterson, 1975), for *F. popenoei* (0.00292), *F. bullenei* (0.00403), *F. near-trigonata* (0.00961), are approximately two to nine times lower than averages values at silent sites from the two outcrossing plants (*Zea mays* ssp. *parviglumis* and *Arabidopsis lyrata* ssp. *petraea*) that have the highest levels of nucleotide diversity that have been measured (Wright and Gaut, 2005). However, our survey of variation is based on a small sample of genes and species, and thus, genetic studies that include more loci and species would help to resolve this last point.

CONCLUSIONS AND PROSPECTS: MULTIPLE POLLINATORS AND THE ORIGIN OF FIG DIVERSITY

In the past, it has been difficult to explain the origin of the remarkable global and local diversity of fig species given the assumption of strict-sense cospeciation and one-to-one pollinator species-specificity. As a result, there is no generally accepted view on the mechanisms of speciation in figs and their pollinators (Michaloud *et al.*, 1996; Weiblen, 2002). A few authors have proposed that speciation in these organisms happens by allopatric isolation (Janzen, 1979; Michaloud *et al.*, 1996; Ramirez, 1970a). However, genetic studies using paternity analyses have shown that neotropical figs have the highest documented distances of gene flow of any tropical plant (Nason *et al.*, 1996, 1998). Despite growing at very low densities (1–10 individuals per km^2) and having asynchronous flowering within populations, individual fig trees receive pollinating wasps (and pollen) from a large number of individual trees. Conservative estimates suggest that fig wasps routinely disperse pollen over distances of >10 km and that breeding populations of figs constitute hundreds of individuals spread over areas >100 km^2 (Nason *et al.*, 1998). Therefore, allopatric isolation in monoecious neotropical species is highly unlikely. Other authors have proposed that temporal isolation (allochrony) generated by flower asynchrony could lead to isolation in sympatry (Kiester *et al.*, 1984). However, population genetic studies of the pollinators reveal no evidence of population subdivision (Molbo *et al.*, 2004), an essential condition for the allochronic speciation model (Kiester *et al.*, 1984).

In a short essay published in 1961, Baker proposed pollinator generalization (lack of species specificity) and hybridization among fig species as a potential explanation for the large diversity of recognized species of

Ficus: ''if hybridization is possible . . . it may be that only the natural eagerness of taxonomists to name new (fig) 'species' has prevented the recognition of some of the intermediates between extreme forms as hybrids.'' Given the scarce evidence of breakdowns to the one-to-one rule at that time, Baker's suggestion was not taken as seriously as more recent results now suggest it should be. Our preliminary genetic results, the findings on the large number of exceptions to the one-to-one rule (Compton, 1990; Kerdelhué *et al.*, 1997, 1999; Lopez-Vaamonde *et al.*, 2002; Michaloud *et al.*, 1985, 1996; Molbo *et al.*, 2003; Ramirez, 1970b; Rasplus, 1996; Wiebes, 1994), and the observation of hybrid (or backcross) figs in nature reviewed above, support Baker's suggestion and provide the foundations for developing an appropriate hypothesis for understanding the coevolutionary dynamics of this mutualism.

Here, we propose that hybridization and introgression due to pollinator host switches and pollinator host sharing may be a major factor underlying much of the tremendous diversity of fig species. Hybridization can lead to generation of new genotypic combinations that may then diversify and lead to the evolution of additional specialized pollinators. The process of divergence in the pollinators can be reinforced by their inbred population structure (Askew, 1986), and fine-tuning of the host recognition mechanisms would promote pollinator divergence and speciation. By this view, the coevolutionary history of the mutualism is that of semispecific wasps (that are good biological species) moving back and forth between figs that may not be good biological species. Although this hypothesis does not yet provide a detailed model for explaining speciation in figs and wasps, the recognition that wasps are not strictly host-specific and that interspecific hybridization in figs is not rare over both ecological and evolutionary time is almost certainly crucial for developing such a model. Minimally, breakdowns in specificity appear to explain the frequent observation of fig individuals with intermediate or mosaic morphologies, the existence of species complexes of figs (Berg, 1989), as well as the observed lack of congruence of fig and wasp phylogenies at various taxonomic scales. Also, host switching of pollinators followed by the introgression of complete gene complexes could help to explain several interesting cases of gains and losses of elaborate characters (e.g., passive or active pollination) (Cook *et al.*, 2004; Kjellberg *et al.*, 2001; Machado *et al.*, 2001).

Progress will depend fundamentally on improved population genetic and phylogenetic datasets for resolving the many outstanding questions on the coevolutionary dynamics of the mutualism both at local and regional geographic scales. Furthermore, we need to document in detail the reproductive consequences of pollinator ''mistakes'' (host switches and host sharing) for both partners in the mutualism. Understanding the evo-

lutionary trajectories of ecologically important fig-associated characters that have been shown to influence reproductive success and host recognition in the wasps (flower number, seed size, seed number, and receptive volatiles) (Barker, 1985; Gibernau and Hossaert-McKey, 1998; Grison *et al.*, 1999; Herre, 1989, 1999; Herre and West, 1997; Hossaert-McKey *et al.*, 1994; Van Noort *et al.*, 1989; Ware *et al.*, 1993) could allow us understand what allows some combinations of wasp fig-pollination mistakes or host-switches to occur more often than others. Understanding the consequence of pollination mistakes appears to be critical for understanding the processes that affect gene flow, coadaptation, and cospeciation at a fine taxonomic scale in the fig/fig-wasp mutualism.

Given the information now available to us, it appears that strict-sense cospeciation of one-to-one species specific figs and wasps should not be the default paradigm for formulating hypotheses to explain the extraordinary diversification of fig and wasp species. For cases in which appropriate studies have been conducted, most figs appear to be pollinated by more than one species of wasp, many of these wasps appear to colonize new species of figs, and interspecific hybridization and introgression appear to be widespread among figs. We propose that the best model for understanding evolutionary dynamics in this mutualism is one in which groups of genetically well defined species of wasps coevolve with groups of genetically less well defined (frequently hybridizing) groups of figs.

ACKNOWLEDGMENTS

We thank Adalberto Gomez and Maritza Lopez for help with field work; Whitman Schofield, Erica Hudson, Zuleyka Mainard, and Jeovanna Lowe for help with molecular data collection; G. Michaloud, S. G. Compton, W. Ramirez, C. C. Berg, J. M. Cook, and M. Zavodna for discussion of their observations suggesting hybridization in nature and species complexes in figs; and an anonymous reviewer and E. Leigh, Jr., S. Van Bael, and J. Hey for comments that greatly improved the presentation. This work was supported by the Smithsonian Tropical Research Institute, National Science Foundation Grant DEB-0108475, and start-up research funds from the University of Arizona.

REFERENCES

Askew, R. R. (1968) Considerations on speciation in Chalcidoidea (Hymenoptera). *Evolution (Lawrence, Kans.)* **22,** 642–645.

Baker, H. G. (1961) *Ficus* and *Blastophaga*. *Evolution (Lawrence, Kans.)* **15,** 378–379.

Barker, N. P. (1985) Evidence of a volatile attractant in *Ficus ingens* (Moraceae). *Bothalia* **15,** 607–611.

Berg, C. C. (1989) Classification and distribution of *Ficus*. *Experientia* **45,** 605–611.

Berg, C. C. & Wiebes, J. T. (1992) *African Fig Trees and Fig Wasps* (North-Holland, Amsterdam).
Boucher, D. H. (1985) *The Biology of Mutualism: Ecology and Evolution* (Croom Helm, London).
Bronstein, J. L. (2001) Mutualisms. In *Evolutionary Ecology: Concepts and Case Studies*, eds. Fox, C. W., Roff, D. A. & Fairbairn, D. J. (Oxford Univ. Press, Oxford), pp. 315–330.
Compton, S. G. (1990) A collapse of host specificity in some African fig wasps. *S. Afr. J. Sci.* **86,** 39–40.
Compton, S. G. (1993) One way to be a fig. *Afr. Entomol.* **1,** 151–158.
Cook, J. M. & Rasplus, J.-Y. (2003) Mutualist with attitude: coevolving fig wasps and figs. *Trends Ecol. Evol.* **18,** 241–248.
Cook, J. M., Bean, D., Power, S. A. & Dixon, D. J. (2004) Evolution of a complex coevolved trait: active pollination in a genus of fig wasps. *J. Evol. Biol.* **17,** 238–246.
Corner, E. J. H. (1952) *Wayside Trees of Malaya* (Government Printer Office, Singapore).
Corner, E. J. H. (1958) An introduction to the distribution of *Ficus*. *Reinwardtia* **4,** 325–355.
Corner, E. J. H. (1965) Check-list of *Ficus* in Asia and Australasia, with keys to identification. *Gdns. Bull. Singapore* **21,** 1–186.
Corner, E. J. H. (1985) *Ficus* (Moracea) and *Hymenoptera* (Chalcidoidea): Figs and their pollinators. *Biol. J. Linn. Soc.* **25,** 187–195.
Datwyler, S. L. & Weiblen, G. D. (2004) On the origin of the fig: Phylogenetic relationships of Moraceae from NDHF sequences. *Am. J. Bot.* **91,** 767–777.
Douglas, A. E. (1994) *Symbiotic Interactions* (Oxford Univ. Press, Oxford).
Fu, Y. X. & Li, W. H. (1993) Statistical tests of neutrality of mutations. *Genetics* **133,** 693–709.
Galil, J. (1977) Fig biology. *Endeavour* **1,** 52–56.
Galil, J. & Eisikowitch, D. (1968) On the pollination ecology of *Ficus sycomorus* in East Africa. *Ecology* **49,** 259–269.
Gibernau, M. & Hossaert-McKey, M. (1998) Are olfactory signals sufficient to attract fig pollinators? *Ecoscience* **5,** 306–311.
Grison, L., Edwards, A. A. & Hossaert-McKey, M. (1999) Interspecies variation in floral fragrances emitted by tropical Ficus species. *Phytochemistry* **52,** 1293–1299.
Harrison, R. D. (2005) Figs and the diversity of tropical rain forests. *Bioscience*, in press.
Herre, E. A. (1989) Coevolution of reproductive characteristics in 12 species of new world figs and their pollinator wasps. *Experientia* **45,** 637–647.
Herre, E. A. (1996) An overview of studies on a community of Panamanian figs. *J. Biogeogr.* **23,** 593–607.
Herre, E. A. (1999) Laws governing species interactions? Encouragement and caution from figs and their associates. In *Levels of Selection in Evolution*, ed. Keller, L. (Princeton Univ. Press, Princeton), pp. 209–237.
Herre, E. A. & West, S. A. (1997) Conflict of interest in a mutualism: Documenting the elusive fig wasp-seed trade-off. *Proc. R. Soc. London Ser. B* **264,** 1501–1507.
Herre, E. A., Machado, C. A., Bermingham, E., Nason, J. D., Windsor, D. M., McCafferty, S. S., VanHouten, W. & Bachmann, K. (1996) Molecular phylogenies of figs and their pollinator wasps. *J. Biogeogr.* **23,** 521–530.
Herre, E. A., Knowlton, N., Mueller, U. G. & Rehner, S. A. (1999) The evolution of mutualisms: exploring the paths between conflict and cooperation. *Trends Ecol. Evol.* **14,** 49–53.
Herr, E. A., Machado, C. A. & West., S. A. (2001) In *Adaptationism and Optimality*, eds. Orzack, S. & Sober, E. (Cambridge Univ. Press, New York), pp. 191–218.
Hossaert-McKey, M., Gibernau, M. & Frey, J. E. (1994) Chemosensory attraction of fig wasps to substances produced by receptive figs. *Entomol. Exp. Appl.* **70,** 185–191.
Hudson, R. R., Kreitman, M. & Aguade, M. (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* **116,** 153–159.
Janzen, D. H. (1979) How to be a fig. *Annu. Rev. Ecol. Syst.* **10,** 13–51.

Jousselin, E., Rasplus, J.-Y. & Kjellberg, F. (2003) Convergence and coevolution in a mutualism: Evidence from a molecular phylogeny of *Ficus*. *Evolution (Lawrence, Kans.)* **57,** 1255–1269.

Kalko, E. K. V., Herre, E. A. & Handley, C. O. (1996) Relation of fig characteristics to fruit-eating bats in the New and Old World Tropics. *J. Biogeogr.* **23,** 565–576.

Kato, M., Takimura, A. & Kawakita, A. (2003) An obligate pollination mutualism and reciprocal diversification in the tree genus *Glochidion* (Euphorbiaceae). *Proc. Natl. Acad. Sci. USA* **100,** 5264–5267.

Kerdelhué, C., Hochberg, M. E. & Rasplus, J. Y. (1997) Active pollination of *Ficus sur* by two sympatric fig wasp species in West Africa. *Biotropica* **29,** 69–75.

Kerdelhué, C., Le Clainche, I. & Rasplus, J. Y. (1999) Molecular phylogeny of the *Ceratosolen* species pollinating *Ficus* of the subgenus *Sycomorus sensu stricto*: Biogeographical history and origins of the species-specificity breakdown cases. *Mol. Phylogenet. Evol.* **11,** 401–414.

Kiester, A. R., Lande, R. & Schemske, D. W. (1984) Models of coevolution and speciation in plants and their pollinators. *Am. Nat.* **124,** 220–243.

Kjellberg, F., Jousselin, E., Bronstein, J.L., Patel, A., Yokoyama, J. & Rasplus, J. Y. (2001) Pollination mode in fig wasps: The predictive power of correlated traits. *Proc. R. Soc. London Ser. B* **268,** 1113–1121.

Kliman, R. M., Andolfatto, P., Coyne, J. A., Depaulis, F., Kreitman, M., Berry, A. J., McCarter, J., Wakeley, J. & Hey, J. (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156,** 1913–1931.

Korine, C., Kalko, E. K. V. & Herre, E. A. (2000) Fruit characteristics and factors affecting fruit removal in a Panamanian fig community. *Oecologia* **123,** 560–568.

Lopez-Vaamonde, C., Dixon, D. J., Cook, J. M. & Rasplus, J.-Y. (2002) Revision of the Australian species of Pleistodontes (Hymenoptera: Agaonidae) fig-pollinating wasps and their host-plant associations. *Zool. J. Linn. Soc.* **136,** 637–683.

Machado, C. A., Herre, E. A., McCafferty, S. & Bermingham, E. (1996) Molecular phylogenies of fig pollinating and non-pollinating wasps and the implications for the origin and evolution of the fig-fig wasp mutualism. *J. Biogeogr.* **23,** 531–542.

Machado, C. A., Jousselin, E., Kjellberg, F., Compton, S. G. & Herre, E. A. (2001) Phylogenetic relationships, historical biogeography, and character evolution of fig-pollinating wasps. *Proc. R. Soc. London Ser. B* **268,** 685–694.

Machado, C. A., Kliman, R. M., Markert, J. A. & Hey, J. (2002) Inferring the history of speciation using multilocus DNA sequence data: The case of *Drosophila pseudoobscura* and close relatives. *Mol. Biol. Evol.* **19,** 472–488.

Margulis, L. & Fester, R. (1991) *Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis* (MIT Press, Cambridge, MA).

Maynard Smith, J. & Szathmáry, E. (1995) *The Major Transitions in Evolution* (Freeman Spektrum, Oxford).

McKey, D. (1989) Population biology of figs: Applications for conservation. *Experientia* **45,** 661–673.

Michaloud, G., Michaloud-Pelletier, S., Wiebes, J. T.& Berg, C. C. (1985) The co-occurrence of two pollinating species of fig wasp and one species of fig. *Proc. K Ned. Akad. Wet. C* **88,** 93–119.

Michaloud, G., Carriere, S. & Kobbi, M. (1996) Exceptions to the one:one relationship between African fig trees and their fig wasp pollinators: Possible evolutionary scenarios. *J. Biogeogr.* **23,** 513–520.

Molbo, D., Machado, C. A., Sevenster, J. G., Keller, L. & Herre, E. A. (2003) Cryptic species of fig-pollinating wasps: Implications for the evolution of the fig-wasp mutualism, sex allocation, and precision of adaptation. *Proc. Natl. Acad. Sci. USA* **100,** 5867–5872.

Molbo, D., Machado, C. A., Herre, E. A. & Keller, L. (2004) Inbreeding and population structure in two pairs of cryptic fig wasp species. *Mol. Ecol.* **13,** 1613–1623.

Murray, M. G. (1985) Figs (*Ficus* spp.) and fig wasps (Chalcidoidea, Agaonidae): Hypotheses for an ancient symbiosis. *Biol. J. Linn. Soc.* **26,** 69–81.

Nason, J. D., Herre, E. A. & Hamrick, J. L. (1996) Paternity analysis of the breeding structure of strangler fig populations: Evidence for substantial long-distance wasp dispersal. *J. Biogeogr.* **23,** 501–512.

Nason, J. D., Herre, E. A. & Hamrick, J. L. (1998) The breeding structure of a tropical keystone plant resource. *Nature* **391,** 685–687.

Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).

Page, R. D. M. (1994) Parallel phylogenies—Reconstructing the history of host-parasite assemblages. *Cladistics* **10,** 155–173.

Pamilo, P. & Nei, M. (1988) Relationships between gene trees and species trees. *Mol. Biol. Evol.* **5,** 568–583.

Parrish, T. L., Koelewijn, H. P. & van Dijk, P. J. (2003) Genetic evidence for natural hybridization between species of dioecious *Ficus* on island populations. *Biotropica* **35,** 333–343.

Pellmyr, O. (2003) Yuccas, yucca moths and coevolution: A review. *Ann. Missouri Bot. Gard.* **90,** 35–55.

Ramirez, W. (1970a) Host specificity of fig wasps (Agaonidae). *Evolution (Lawrence, Kans.)* **24,** 681–691.

Ramirez, W. (1970b) Taxonomic and biological studies of neotropical fig wasps (Hymenoptera: Agaonidae). *Univ. Kansas Sci. Bull.* **49,** 1–44.

Ramirez, W. (1974) Coevolution of *Ficus* and Agaonidae. *Ann. Missouri Bot. Gard.* **61,** 770–780.

Ramirez, W. (1991) Evolution of the mandibular appendage in fig wasps (Hymenoptera, Agaonidae). *Rev. Biol. Trop.* **39,** 87–95.

Ramirez, W. (1994) Hybridization of *Ficus religiosa* with *F. septica* and *F. aurea* (Moraceae). *Rev. Biol. Trop.* **42,** 339–342.

Rasplus, J.-Y. (1996) The one-to-one species-specificity of the Ficus-Agaoninae mutualism: How casual? In *The Biodiversity of African Plants*, eds. van der Maesen, L. J. G., van der Burgt, X. M. & van Medenbach de Rooy, J. M. (Kluwer, Wageningen, The Netherlands), pp. 639–649.

Rasplus, J.-Y., Kerdelhué, C., Le Clainche, I. & Mondor, G. (1998) Molecular phylogeny of fig wasps. Agaonidae are not monophyletic. *C. R. Acad. Sci. III* **321,** 517–527.

Shimodaira, H. & Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16,** 1114–1116.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87,** 651–701.

Strand, A. E., Leebens-Mack, J. & Milligan, B. G. (1997) Nuclear DNA-based markers for plant evolutionary biology. *Mol. Ecol.* **6,** 113–118.

Sweigart, A. L. & Willis, J. H. (2003) Patterns of nucleotide diversity in two species of *Mimulus* are affected by mating system and asymmetric introgression. *Evolution (Lawrence, Kans.)* **57,** 2490–2506.

Swofford, D. L. (1998) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer, Sunderland, MA).

Tajima, F. (1983) Evolutionary relationships of DNA sequences in finite populations. *Genetics* **105,** 437–460.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123,** 585–595.

Terborgh, J. (1986) Keystone plant resources in the tropical forest. In *Conservation Biology: The Science of Scarcity and Diversity*, ed. Soulé, M. E. (Sinauer, Sunderland, MA), pp. 330–344.

Thompson, J. N. (1994) *The Coevolutionary Process* (Univ. of Chicago Press, Chicago).

Van Noort, S., Ware, A. B. & Compton, S. G. (1989) Pollinator-specific volatile attractants released from the figs of *Ficus burtt-davyi*. *S. Afr. J. Sci.* **85,** 323–324.

Wang, R. L., Wakeley, J. & Hey, J. (1997) Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* **147,** 1091–1106.

Ware, A. B. & Compton, S. G. (1992) Breakdown of pollinator specificity in an African fig tree. *Biotropica* **24,** 544–549.

Ware, A. B., Kaye, P. T., Compton, S. & Van Noort, S. (1993) Fig volatiles: their role in attracting pollinators and maintaining pollinator specificity. *Plant Syst. Evol.* **186,** 147–156.

Watterson, G. A. (1975) On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7,** 256–276.

Weiblen, G. D. (2000) Phylogenetic relationships of functionally dioecious *Ficus* (Moraceae) based on ribosomal DNA sequences and morphology. *Am. J. Bot.* **87,** 1342–1357.

Weiblen, G. D. (2001) Phylogenetic relationships of fig wasps pollinating functionally dioecious figs based on mitochondrial DNA sequences and morphology. *Syst. Biol.* **50,** 243–267.

Weiblen, G. D. (2002) How to be a fig wasp. *Annu. Rev. Entomol.* **47,** 299–330.

Weiblen, G. D. (2004) Correlated evolution in fig pollination. *Syst. Biol.* **53,** 128–139.

Weiblen, G. D. & Bush, G. L. (2002) Speciation in fig pollinators and parasites. *Mol. Ecol.* **11,** 1573–1578.

Wiebes, J. T. (1963) Taxonomy and host plant preference of Indo-Australian fig wasps of the genus *Ceratosolen* (Agaonidae). *Tijdschr. Entomol.* **106,** 1–112.

Wiebes, J. T. (1979) Co-evolution of figs and their insect pollinators. *Annu. Rev. Ecol. Syst.* **10,** 1–12.

Wiebes, J. T. (1982) The phylogeny of the agaonidae (Hymenoptera, Chalcidoidea). *Neth. J. Zool.* **32,** 395–411.

Wiebes, J. T. (1987) Coevolution as a test of the phylogenetic tree. In *Systematics and Evolution: A Matter of Diversity*, ed. Hovenkamp, P. (Utrecht Univ. Press, Utrecht, The Netherlands), pp. 309–314.

Wiebes, J. T. (1994) *The Indo-Australian Agaoninae: Pollinators of Figs* (North-Holland, Amsterdam).

Wright, S. I. & Gaut, B. S. (2005) Molecular population genetics and the search for adaptive evolution in plants. *Mol. Biol. Evol.* **22,** 506–519.

Wright, S. I., Lauga, B. & Charlesworth, D. (2003) Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. *Mol. Ecol.* **12,** 1247–1263.

Yokoyama, J. (1995) Insect-plant coevolution and speciation. In *Biodiversity and Evolution*, eds. Arai, R., Kato, M. & Doi, Y. (The National Science Museum Foundation, Tokyo), pp. 115–130.

8
Evolutionary Animation: How Do Molecular Phylogenies Compare to Mayr's Reconstruction of Speciation Patterns in the Sea?

STEPHEN R. PALUMBI* AND H. A. LESSIOS†

Ernst Mayr used the geography of closely related species in various stages of increasing divergence to ''animate'' the process of geographic, or allopatric, speciation. This approach was applied to a wide set of taxa, and a seminal paper by Mayr used it to explore speciation patterns in tropical sea urchins. Since then, taxonomic information in several of these genera has been augmented by detailed molecular phylogenies. We compare Mayr's animation with the phylogenies of eight sea urchin genera placed by Mayr into four speciation groups. True to Mayr's predictions, early-stage genera have on average lower species divergence and more polytypic species than genera in later stages. For six of these genera, we also have information about the evolution of the gamete recognition protein bindin, which is critical to reproductive isolation. These comparisons show that later-stage genera with many sympatric species tend to be those with rapid bindin evolution. By contrast, early-stage genera with few sympatric species are not necessarily earlier in the divergence process; they happen to be those with slow rates of bindin evolution. These results show that the rate of speciation in sea urchins does not only depend on the steady accumulation of genome divergence over time, but also on

*Hopkins Marine Station, Department of Biological Sciences, Stanford University, Pacific Grove, CA 93950; and †Smithsonian Tropical Research Institute, Unit 2072, Apartado Postal 0843-03092, Balboa, Panama.

Abbreviation: COI, cytochrome oxidase I.

the rate of evolution of gamete recognition proteins. The animation method used by Mayr is generally supported by molecular phylogenies. However, the existence of multiple rates in the acquisition of reproductive isolation complicates placement of different genera in an evolutionary series.

Ernst Mayr built an argument for the way speciation occurred based on the geographic patterns of variation among closely related species (Mayr, 1942, 1963). He showed that there was a hierarchy of species descriptions that could be ordered in a series of increasing complexity. Some descriptions pertained to recently diverged species, with morphologically identical populations inhabiting a continuous range. Other descriptions were of polytypic species, those with slightly differentiated populations inhabiting different parts of the range. Further along the speciation axis were superspecies, taxa with morphologically distinct, allopatric populations. Still later in the series, Mayr identified groups of related species in which some taxa were sympatric. The trajectory from homogeneous populations to overlapping sympatric species encompassed Mayr's view of the process and pacing of geographic speciation. In addition to describing these separate elements, a major contribution by Mayr was to order these elements in a series. The elements thus served as separate frames in an evolutionary animation that sped up the slow process of speciation so that it could be viewed and studied by biologists.

The geographic distributions of species, subspecies, varieties, and slightly divergent populations constituted the database in Mayr's analyses. He made the implicit assumption that the genetic and evolutionary divergence of these groups increased from population- to species-level distinction and used morphological differentiation as a proxy for evolutionary time. Mayr established sister-species relationships on the basis of morphological similarity and included a tacit phylogenetic framework for his animations based on the best information available at the time.

One difficulty faced by Mayr was that few concrete phylogenetic analyses were available during the development of these ideas. Since that era, molecular phylogenies have made it possible to obtain a statistically robust view of phylogenetic relationships, divergence order, and sister-species status (Hillis *et al.*, 1996). Molecular phylogenies can also give an indication of the timing of divergence events through the application of molecular clock calibrations. Even without precise time calibrations, the record of the order of divergence of taxa permits an examination of the causes of each splitting event. Lastly, phylogenies can provide objective data on divergence levels to test predictions of Mayr's evolutionary reconstructions.

For all these reasons, molecular phylogenies can contribute substantially to an updated view of evolutionary animation. To what extent do modern phylogenies confirm or reject the orderly animations suggested by Mayr's analysis? Mayr's ideas were shaped primarily from his studies of bird systematics, and, indeed, subsequent use of molecular phylogenies showed his insistence on the primacy of the allopatric mode of speciation to be correct for this group [Barraclough and Vogler, 2000; Edwards *et al.*, "Speciation in Birds: Genes, Geography, and Sexual Selection" (Chapter 6)]. Mayr, however, was also interested in applying his view of speciation to all animals, both terrestrial and marine, and his evolutionary animations included more than bird species. Some of the nonavian genera that interested him have been examined extensively for phylogenetic relationships by using molecular tools.

In 1954, Mayr published a paper on the geographic speciation of marine taxa by focusing on the differentiation and geography of species in 20 genera of tropical sea urchins (Mayr, 1954). His goal was to examine the generality of his ideas about species formation by extending them to groups with ecology very different from that of birds. This goal was an important one for Mayr. In his view, even the most comprehensive recounting of speciation mechanisms was wanting if it applied to only a single taxon. Mayr's conclusion at the end of this analysis was that geographic speciation applies equally well to marine species, such as sea urchins, as it does to birds, mammals, and insects.

Mayr took advantage of the publication of a complete monograph of sea urchin taxonomy by Mortensen (1928–1951) to examine the stages of speciation represented by each genus, restricting his attention to genera with species that inhabited the shallow seas in the West Indies. He divided the genera into four groups. Groups 1 and 2 included genera that had strictly allopatric species, the only difference between the groups being that group 2 genera included nontropical species. These two groups represented the earliest step in species formation, with a surfeit of polytypic species, and, presumably, low divergence among allopatric sister species. Group 3 represented the next step in speciation, in which congeneric species had sufficient time to develop genetic divergence, thereby allowing formerly allopatric sister species to invade each other's ranges. Thus, these genera were presumed to show higher levels of divergence and the beginnings of sympatric overlap. Finally, group 4 genera were those in which the species were so old that current geographic ranges had nothing to do with speciation pattern. Whatever record of species formation there may have been among these ancient species was gone.

Although Mayr had the advantage of using the most comprehensive and up-to-date treatment of sea urchins provided by Mortensen's monograph, this treatment was almost strictly taxonomic. Within genera, there

was little or no information of the relationships among species. This lack of systematics prevented Mayr from basing any of his group designations on sister-species relationships. However, in the past 10 years, detailed molecular phylogenetic studies have been completed for six genera in Mayr's original list and two others that do not occur in the West Indies. In this paper, we collate these phylogenies and use them to test general predictions from Ernst Mayr's reconstruction of sea urchin speciation patterns. General predictions across all genera in the study include familiar tenets of allopatric speciation. Testing these predictions with molecular data reveals a considerable concordance with Mayr's animations but also shows some surprises about the way speciation proceeds in sea urchins.

MATERIALS AND METHODS

We compiled molecular data for variation in the mitochondrial cytochrome oxidase I (COI) genes for Mayr's groups 1 and 2 genera *Tripneustes* (Lessios *et al.*, 2003a), *Eucidaris* (Lessios *et al.*, 1999), and *Lytechinus* (Zigler and Lessios, 2004); group 3 genera *Echinometra* (McCartney *et al.*, 2000; Landry *et al.*, 2003) and *Diadema* (Lessios *et al.*, 2003b); and group 4 genera *Strongylocentrotus* (Biermann *et al.*, 2003), *Arbacia* (Metz *et al.*, 1998), and *Heliocidaris* (Zigler *et al.*, 2003). Phylogenetic relationships are taken directly from the original analyses. Data for the gamete recognition molecule bindin are available for six of these eight genera: *Tripneustes* (Zigler and Lessios, 2003a), *Lytechinus* (Zigler and Lessios, 2004), *Echinometra* (Landry *et al.*, 2003; Metz and Palumbi, 1996; McCartney and Lessios, 2004), *Strongylocentrotus* (Biermann, 1998), *Arbacia* (Metz *et al.*, 1998), and *Heliocidaris* (Zigler *et al.*, 2003). In general, phylogenetic relationships at COI and bindin are concordant. Major exceptions are the positions of *Lytechinus williamsi* and *Echinometra insularis* in their respective genera (Landry *et al.*, 2003; Zigler and Lessios, 2004). In *Strongylocentrotus*, we have included the monotypic genera *Allocentrotus* and *Hemicentrotus* because phylogenetic analysis places the species in these genera firmly within the genus *Strongylocentrotus* (Biermann *et al.*, 2003). We used Kimura two-parameter genetic distances based on COI comparisons. In *Diadema*, combined ATPase 8 and 6 and COI sequences were used. Distances were compiled for sister species (two or more species that split at the tip of a branch or a species that forms an outgroup to a clade of closely related species). Allopatric neighbors are defined as allopatric species that are not separated by an obvious, insurmountable geographic barrier, such as a land mass.

We characterize bindin evolution as ''fast'' if the nonsynonymous-to-synonymous substitution ratio in at least an ≈100-bp ''hotspot'' region of the gene is >1 and if there are several codons on which positive selection

has acted. By this definition, bindin is considered to be evolving quickly in the genera *Echinometra*, *Strongylocentrotus*, and *Heliocidaris* (Biermann, 1998; Landry *et al.*, 2003; Metz and Palumbi, 1996; Zigler *et al.*, 2003) and slowly in *Lytechinus*, *Arbacia*, and *Tripneustes* (Metz *et al.*, 1998; Zigler and Lessios, 2003a, 2004). Although positive selection could not be statistically demonstrated for any comparison between species of *Lytechinus*, we consider the bindin of *L. williamsi* and *Lytechinus variegatus variegatus* as rapidly evolving, because its divergence is much higher than expected from comparisons of COI between these species. The COI haplotypes of *L. williamsi* and *L. variegatus* are intermingled, whereas bindin sequences are reciprocally monophyletic, suggesting the possibility of selection on the latter (Zigler and Lessios, 2004).

Estimation of time since species separation from COI divergence assumes a molecular clock, which is calibrated by the complete closure of the seaway between the eastern Pacific and the Atlantic by the isthmus of Panama at ≈3.1 million years ago (Coates and Obando, 1996). In all comparisons involving tropical genera, Atlantic and Pacific species are included. The divergence between clades presumably split by the isthmus provides a rough calibration of a molecular clock for each genus, allowing us to place a general time frame on species divergence patterns.

RESULTS

Groups 1 and 2

Polytypic species and low levels of genetic divergence between allopatric entities within an ocean basin are common in the genera *Tripneustes*, *Lytechinus*, and *Eucidaris* (Fig. 8.1). For example, in the genus *Eucidaris*, all Atlantic nominal species or subspecies, *Eucidaris tribuloides*, *Eucidaris clavata*, and *Eucidaris tribuloides variegatus africana* form one single genetic cluster with no distinction among them (Lessios *et al.*, 1999). Together, this broad polytypic species is sister to the Eastern Pacific pair *Eucidaris galapagensis* and *Eucidaris thouarsi*. Based on a molecular rate estimate provided by the rise of the isthmus of Panama, the latter two sister species diverged ≈2 million years ago. An allopatric neighbor of these species, the widely distributed Pacific–Indian ocean species *Eucidaris metularia* is not a close relative of the eastern Pacific species. Rather, this species is an ancient offshoot in the genus, having diverged 5–8 million years ago.

The genus *Tripneustes* shows a pattern of dissimilar geographic distribution of genetic variation in two major oceanic regions (Lessios *et al.*, 2003a). *Tripneustes depressus* from the eastern Pacific and *Tripneustes gratilla* from the rest of the Indo-Pacific are genetically indistinguishable, forming a single geographically widespread species complex that occupies most of the world's tropical oceans. The Atlantic species *Tripneustes*

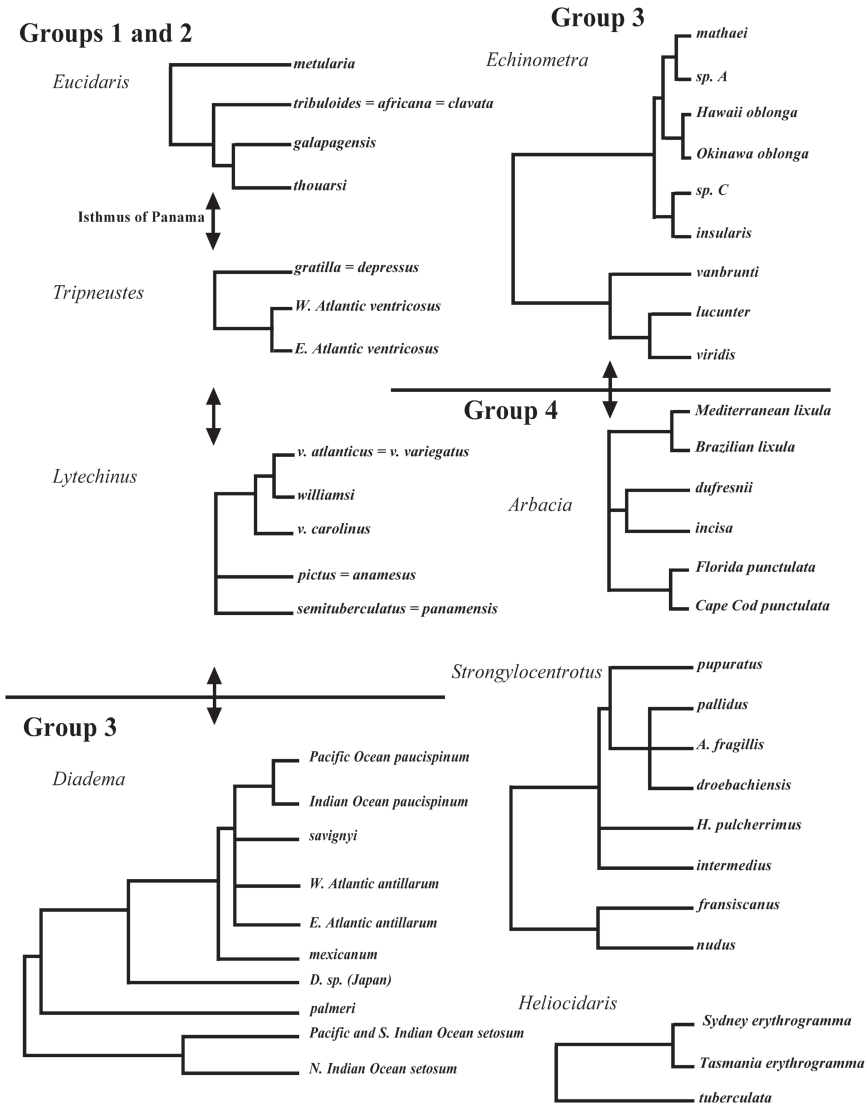


FIGURE 8.1 Molecular phylogenies based on COI (COI plus ATPase8/6 for *Diadema*) from eight sea urchin genera. Data are from Lessios *et al.* (1999, 2003a,b), Zigler and Lessios (2004), McCartney *et al.* (2000), Landry *et al.* (2003), Biermann *et al.* (2003), Metz *et al.* (1998), Zigler *et al.* (2003). The double arrows mark the completion of the Isthmus of Panama at ≈3.1 million years ago. This date is assumed to be approximate for the divergence of Caribbean and eastern Pacific species within the six genera marked by the arrows. Trees are rooted by species from closely related genera [see Biermann *et al.* (2003), Landry *et al.* (2003), Lessios *et al.* (1999, 2003a,b), McCartney *et al.* (2000), Metz *et al.* (1998), Zigler and Lessios (2004), Zigler *et al.* (2003) for details] and are drawn to the same temporal scale. See the text for group definitions.

ventricosus has considerable population structure, indicating a lack of gene flow between the American and African coasts. *T. gratilla* and *T. ventricosus* are assumed to have diverged at the Panamanian closure 3 million years ago.

The last genus in this cluster is more complex. *Lytechinus* has two sets of polytypic species: *Lytechinus anamesus* and *Lytechinus pictus* along the west coast of north America are indistinguishable genetically, as are the eastern Pacific *Lytechinus semituberculatus* and *Lytechinus panamensis* (Zigler and Lessios, 2004). These two Pacific species clusters diverged from each other 3–5 million years ago. In addition, there is a set of Atlantic species with questionable species status. Data from COI show that the subspecies *L. variegatus variegatus* and *Lytechinus variegatus atlanticus* cluster indistinguishably from one another but that *L. williamsi*, partially sympatric with *L. variegatus variegatus*, diverged at ≈500,000 years ago. An outgroup clade to this cluster is *L. variegatus carolinus*, which diverged 2–3 million years ago. The genealogy of bindin shows one discrepancy from that of COI: The phylogenetic positions of *L. williamsi* and *L. variegatus carolinus* are switched (Zigler and Lessios, 2004). In addition, there is evidence for acceleration of bindin evolution in the two sympatric species compared to COI, although maximum likelihood analysis fails to show positive selection, possibly because of low statistical power in these closely related sequences.

The summary of these studies of groups 1 and 2 genera is that molecular phylogenies support Mayr's conclusions that widely distributed polytypic species are commonplace and that allopatric splitting events within ocean basins are sometimes very recent. However, some allopatric neighbors have been in existence for 2–8 million years without evidence that their ranges have begun to overlap.

Group 3

Molecular phylogenies also support Mayr's classification of genera into group 3 because their species are in the initial stages of sympatry; but they also show that they are comprised of species groups with very different evolutionary patterns. In *Diadema*, the widely distributed species *Diadema setosum* and *Diadema savignyi* overlap throughout the western Pacific and Indian oceans (Lessios *et al.*, 2003b). Based on ATPase and COI sequence differences, these two species are highly divergent, having split 7–14 million years ago (Lessios *et al.*, 2003b). The widespread *Diadema savignyi* is also sympatric in Japan and the Marshal islands with an undescribed species, from which it diverged 6.5–13.5 million years ago. Isozymes and mitochondrial DNA have recently uncovered unsuspected cases of sympatry between *Diadema paucispinum*, a species originally

thought to be limited to Hawaii, and the other Indo-West Pacific species (Lessios and Pearse, 1996; Lessios *et al.*, 2003b). The divergence time between the sympatric *D. paucispinum* and *D. savignyi* is <2 million years. By contrast, the eastern Pacific *Diadema mexicanum* has remained allopatric from the Indo-West Pacific species for 3 million years, with only a hint of range overlap with *D. savignyi* at the Clipperton Atoll, the closest point to the central Pacific (Lessios *et al.*, 1996, 2003b). Within this genus, broadly distributed species tend to show divergence of supposedly conspecific allopatric populations. *Diadema antillarum* populations from the eastern and western Atlantic are as different from one another as are accepted species in this genus. Two clades of *D. setosum* in the Indian Ocean probably diverged 5 million years ago. Thus, the species in this genus show a generally higher degree of genetic divergence than genera at earlier stages, with moderately old allopatric populations within a morphospecies. Sympatry occurs between very old species pairs. Allopatric neighbors are old. One exception is the previously unsuspected sympatry of the relatively recently diverged species pair of *D. savignyi* and *D. paucispinum*.

By contrast, the genus *Echinometra* shows a large number of sympatric species with low divergence from one another. In the Pacific there is a cluster of very closely related sympatric species. *Echinometra mathaei*, *Echinometra oblonga* and a currently unnamed species ''*Echinometra* sp. A'' diverged 1–2 million years ago (Landry *et al.*, 2003; Palumbi, 1996). COI data show that this cluster split from *Echinometra* sp. C and the Easter Island endemic *E. insularis* at about the same time. In the Caribbean, the sympatric species *Echinometra lucunter* and *Echinometra viridis* diverged ≈1.5 million years ago (McCartney *et al.*, 2000). One ancient allopatric split persists in this genus: The eastern Pacific *Echinometra vanbrunti* differs from other Pacific species by ≈13% nucleotide differences in COI, corresponding to separation of ≈3.5 million years. However, the allopatry of *E. vanbrunti* from the other Pacific species may be in the process of being erased through infrequent larval influx from the central into the eastern Pacific: *E. oblonga,* although rare, is now present in the outer islands of the eastern Pacific (Lessios *et al.*, 1996; McCartney *et al.*, 2000). There is also one very recent allopatric split: *E. oblonga* appears to be at least two species (*E. oblonga* Okinawa and *E. oblonga* Hawaii). Distinguishable by sperm morphology and genetics, these species are allopatric and have diverged at most 250,000 to 500,000 years ago (Landry *et al.*, 2003).

Data from the bindin locus show rapid evolution in this genus, and generally support the COI phylogeny (Landry *et al.*, 2003; McCartney and Lessios, 2004; Metz and Palumbi, 1996). One major exception is that bindin alleles in *E*. sp. C differ dramatically from one region to another. In this genus, major differences in bindin gene sequence are associated with strong reproductive isolation among closely related species (Landry *et al.*,

2003; Metz and Palumbi, 1996; Palumbi and Metz, 1991). Where *E*. sp. C is in sympatry with *E. oblonga* (Okinawa) bindin alleles are highly divergent; in contrast, where they are allopatric, *E*. sp. C and *E. oblonga* (Hawaii) have very similar alleles (Geyer and Palumbi, 2003). In addition, bindin sequences in *E. insularis* are distinct and monophyletic and do not suggest a close relationship with *E*. sp C (Landry *et al*., 2003). Thus, the genus *Echinometra* has species with extensive sympatry, whereas the genus *Diadema* shows sympatry of just a few species pairs. However, contrary to predictions of Mayr's animation, overall genetic divergence between species of *Echinometra* is smaller than between those of *Diadema* (Fig. 1). In particular, sympatric species of *Echinometra* show much less genetic divergence than sympatric species of *Diadema*.

Group 4

The genus *Arbacia* consists of purely allopatric species, but Mayr included it in group 4 because he doubted the validity of specific rank, even for species that were found in different oceans. He may actually have been right for the wrong reasons, because COI and bindin show that the original species designations correspond to divergent molecular clades and should not be considered as conspecific. However, neighboring allopatric clades are old and thus qualify the genus for inclusion in group 4. Based on COI sequences, *Arbacia punctulata* along the east coast of North America and *Arbacia lixula* from the eastern Atlantic and from Brazil are ≈3–5 million years old (Metz *et al*., 1998). The most recently derived species pair, *Arbacia dufresnei* and *Arbacia incisa*, are 2–4 million years old. Species also tend to be widespread: *A. lixula* occurs from the coast of Brazil to the Mediterranean, with an ≈500,000-year divergence between these genetically distinct populations. The western Atlantic species *A. punctulata* ranges from Cape Cod to Curacao, Trinidad and Tobago, and shows a 2% sequence divergence in COI between Florida and the northern end of its range.

Although Mayr mentioned studies of morphological variation in *Strongylocentrotus droebachiensis* and *Strongylocentrotus pallidus*, he did not place this genus in any of his groups, because it lacks tropical representatives. However, because molecular and morphological variation in *Strongylocentrotus* have been well studied, it can receive the same consideration as the other genera. There are many sympatric species in this genus. A cluster of species sympatric in the Northeast Pacific diverged from one another at ≈3–5 million years ago based on COI and fossil evidence (Biermann *et al*., 2003). *Strongylocentrotus purpuratus*, *Strongylocentrotus droebachiensis*, *S. pallidus*, and *Allocentrotus fragilis* all diverged from one another at about the same time. Although broadly overlapping

in geographic distribution, these species tend to live at different depths. A similar sympatric cluster in Japan is composed of several species that diverged slightly earlier. The two species clusters, one along North America and one along the coast of Japan, show no geographic overlap. A deep split in the genus at ≈10–15 million years ago gave rise to two allopatric sister species: *Strongylocentrotus franciscanus* in the east and *Strongylocentrotus nudus* in the west. The divergence time of this pair of species was ≈5 million years ago. Bindin evolution is rapid in this genus (Biermann, 1998).

An additional genus, also not included by Mayr because it does not occur in the West Indies, has received ample attention with regards to its molecular evolution, in part because of the remarkable divergence in mode of development between its two species. The genus *Heliocidaris* is restricted to Australia and consists of *Heliocidaris tuberculata* and *Heliocidaris erythrogramma*. These two species overlap in range along the southeast coast of Australia, and diverged at ≈5 million years ago (McMillan *et al.*, 1992; Zigler *et al.*, 2003). *H. erythrogramma*, which has an extremely large egg and has evolved direct development, has subpopulations in western and eastern Australia. There are no published sequence data from the western Australian subspecies, but populations from Sydney and Tasmania differ by ≈2% in COI. Bindin evolution has been rapid along the lineage leading to the direct developing *H. erythrogramma*.

General Tests of Predictions

The prediction that genetic divergence of sister species increases from stage to stage of speciation is borne out by comparison of phylogenies and genetic distances among the eight genera of sea urchins (Fig. 8.2). Median COI genetic distance among 10 comparisons in groups 1 and 2 genera (the first stage of allopatric speciation) is ≈1%, whereas divergence is 3% and >10% for groups 3 and 4, respectively. Sister species in groups 1 and 2 tend to have adjoining ranges. In addition, polytypic species are most common in groups 1 and 2 genera. Five of 10 species comparisons in group 1 genera show no discernable genetic divergence, indicating genetic exchange among allopatric populations rather than species-level differentiation. By contrast, all sister-species comparisons in groups 3 and 4 genera are >2%, and there are no species with zero genetic divergence. Instead, in these genera, widely distributed species often contain allopatric populations with genetic divergences as high as that of many species pairs.

Simple predictions begin to falter, however, when comparisons are divided into those between sympatric species and those between allopatric species. Sympatric species are not generally more divergent than allo-

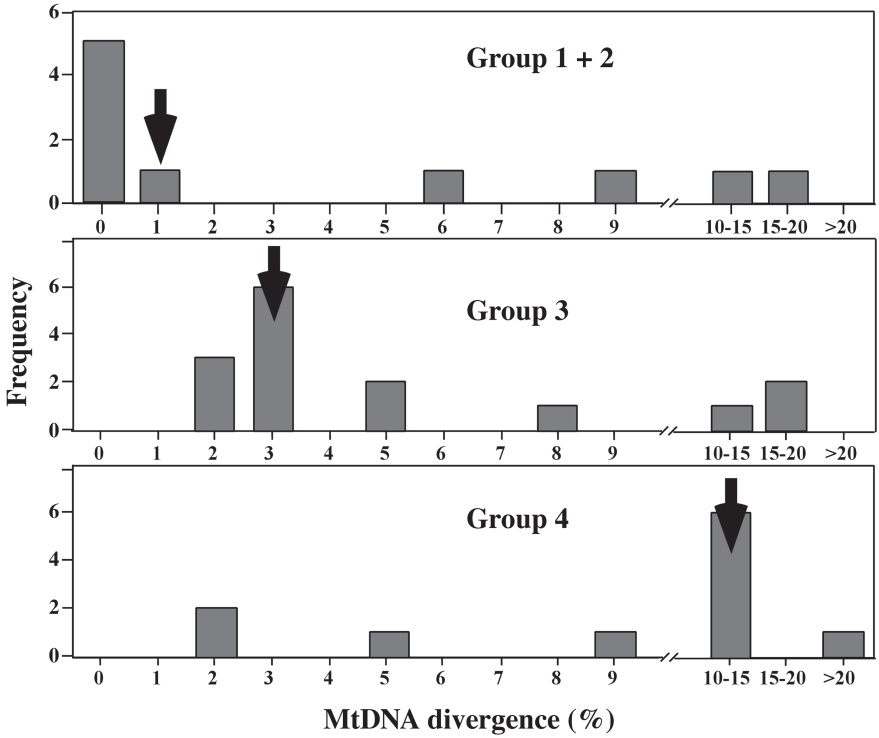


FIGURE 8.2 Genetic divergence among sister species in eight sea urchin genera separated into four groups, representing different stages of allopatric speciation. See the text for the definition of each group. Mitochondrial DNA divergence is based on Kimura two-parameter distances at the COI gene (COI plus ATPase8/6 for *Diadema*). Data are based on Fig. 8.1 and Lessios *et al.* (1999, 2003a,b), Zigler and Lessios (2004), McCartney *et al.* (2000), Landry *et al.* (2003), Biermann *et al.* (2003), Metz *et al.* (1998), Zigler *et al.* (2003). Medians are marked by arrows.

patric species. In fact, in groups 1 and 2 genera, allopatric species are more divergent than sympatric species (Fig. 8.3). This pattern is largely due to the existence of a large fraction of ancient allopatric neighbors, species that have diverged from sister taxa long ago and are separated by no insurmountable major geographic barrier, such as a land mass, even in early-stage genera. Dividing the species into allopatric and sympatric comparisons shows that divergence between sympatric species increases stage by stage but that divergence between allopatric species does not (Fig. 8.3).

Further comparisons show that most sympatric species are found in genera in which the sperm recognition protein bindin is evolving quickly

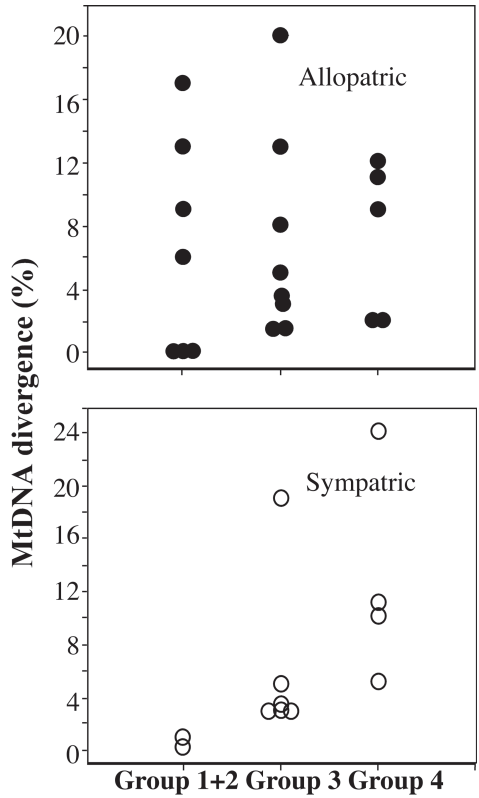


FIGURE 8.3 Genetic divergence of allopatric versus sympatric sister species of sea urchins in groups representing different stages of speciation. See the text for group definitions. Data are as in Figs. 8.1 and 8.2.

(Fig. 8.4). There is significant association between the frequency distribution of sympatric and allopatric species and the rate of their bindin evolution (Fisher's exact test, $P = 0.0033$). Because the species are not phylogenetically independent, the statistical significance should not be interpreted to mean that the two quantities are related directly but rather that they tend to cooccur in the same genera. In the genera *Echinometra* and *Strongylocentrotus*, sympatric species are common, whereas in genera with slow rates of bindin evolution, sympatric species are rare. Among genera with rapid bindin evolution, 10 of 16 sister species comparisons are between sympatric species. In other genera, only two of 19 comparisons are between sympatric species (Fig. 8.4). The exception to this pattern is in the group 2 genus *Lytechinus*. Bindin evolution is generally slow in this ge-

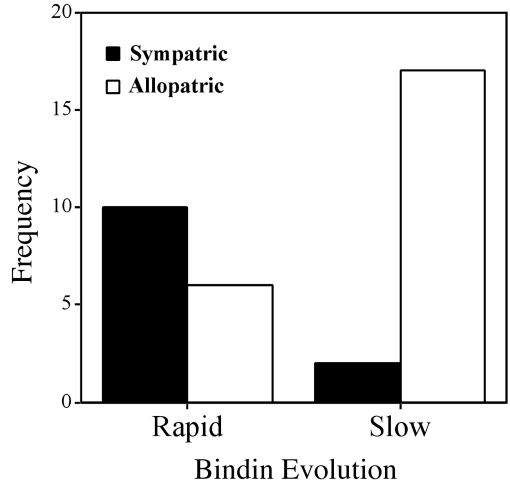


FIGURE 8.4 Frequency of sympatric and allopatric species in sea urchin genera with different rates of bindin evolution.

nus, and most species are allopatric. However, one sympatric species pair (*L. williamsi* and *L. variegatus variegatus*) can be found in the Caribbean. Divergence in bindin between these two species appears to be accelerated relative to COI divergence, although there is no statistically significant signal of positive selection.

DISCUSSION

Evolutionary series based on geography and taxonomy can be independently investigated by comparing molecular phylogenies of species groups at different stages of diversification. Phylogenies based on mitochondrial COI for eight genera of sea urchins, including 45 species, show general agreement with the evolutionary animation proposed by Mayr (1954). Groups 1 and 2 genera have lower genetic divergence and a higher incidence of polytypic species than genera at later stages. Many sister species pairs tend to be allopatric and closely related at the groups 1 and 2 stages, but, by groups 3 and 4, sister species are largely sympatric or are old allopatric neighbors.

However, ancient allopatric neighbors also occur in genera at early stages of speciation. Many of these ancient allopatric species exist across a deep water stretch of the Pacific Ocean termed the East Pacific Barrier (Ekman, 1953). Mayr (1954) writes ''of the existing barriers, by far the

most potent has been that between Polynesia and the American coast.'' Molecular phylogenies have revealed this statement to hold true for *Eucidaris* (Lessios *et al.*, 1999), *Diadema* (Lessios *et al.*, 2003b), and *Echinometra* (McCartney *et al.*, 2000), although in the latter two genera there are indications that larvae are occasionally able to breach the barrier, which may represent the early stages leading toward secondary sympatry. *Tripneustes* in the Indo-Pacific, on the other hand, shows no evidence that its gene flow is in any way impeded by the 5,000 km of deep open water between Clipperton and the Marquesas (Lessios *et al.*, 2003a), and a similar situation exists for *Echinothrix*, an Indo-Pacific genus of sea urchins that appears to have recently colonized the eastern Pacific (Lessios *et al.*, 1996, 1998). It is unclear why some species are able to traverse the barrier so easily when the majority cannot, because there are no pronounced differences in the length of the competent larval stage of the genera.

Lessios and coworkers (Lessios and Cunningham, 1990; McCartney *et al.*, 2000) suggest that rare immigrants into the range of an allopatric neighbor will most likely fail to reproduce or will hybridize with the resident species. Maintenance of a rare species within the range of a more numerous one demands some mechanism of reproductive isolation or assortative mating. Both mechanisms operate at the surfaces of gametes during sea urchin spawning (Palumbi, 1992). Sperm attachment and fusion is facilitated by interaction of bindin with a large protein receptor on the egg surface (Kamei and Glabe, 2003; Palumbi, 1999). Rapid evolution of bindin in sea urchins generates assortative mating (Palumbi, 1999) and egg–sperm incompatibility (Biermann, 1998; McCartney and Lessios, 2004; Metz and Palumbi, 1996). Other than the timing of gamete release (Lessios, 1984), few behavioral mechanisms of mate choice operate in these free spawning invertebrates; therefore, interactions of gametes assume a greater role in reproductive isolation. Positive selection in the bindin gene is associated with functional divergence of gamete recognition within and between species (McCartney and Lessios, 2004; Palumbi, 1999). In turn, this divergence may allow sympatry of species at an earlier stage of divergence. Genera with rapid bindin evolution include many closely related sympatric species. By contrast, in genera without rapid change in bindin, closely related or moderately related species are nearly exclusively allopatric (Fig. 8.4).

This difference among genera enhances our understanding of species formation and helps us interpret Mayr's animation in more mechanistic ways. In Mayr's original formulation, groups 1 and 2 genera eventually evolve into groups 3 and 4 genera. The major difference between these categories is the amount of time that species have had to diverge genetically: Reproductive isolation was thought to evolve as a consequence of

overall genetic divergence or adaptive differentiation (Mayr, 1942, 1963). The critical importance of a few gamete recognition loci to reproductive isolation can disrupt the steady pace of evolution of isolation (Dobzhansky, 1937, 1970; Orr, 1991; Wyckoff *et al.*, 2001). In these cases, the way a few loci evolve may be more important than overall genetic divergence or adaptive differentiation (Dobzhansky, 1937, 1970).

Our phylogenetic summary shows that the distinction between sea urchin genera at different speciation stages is related to bindin evolutionary rate. Groups 1 and 2 genera have slow bindin evolution, have largely allopatric species, and are therefore earlier frames in the evolutionary animation. Groups 3 and 4 genera, with many more sympatric species, are classified later in the evolutionary series, but their species are not necessarily older. Instead, these can be the genera in which rapid bindin change generates reproductive isolation among even closely related species.

A similar hierarchy of speciation rate might be present in Mayr's early work on tropical Pacific birds. Genera with bright male plumage and strong sexual selection were textbook cases of geographic variation within sister-species complexes (Mayr, 1942). Such genera might be akin to the sea urchin genera with fast bindin evolutionary rates. A key difference, however, is that evolution of plumage generates morphological diversity across a species that allows it to be divided into taxonomic units based on morphology (Mayr, 1942). By contrast, rapid evolution of bindin does not in and of itself generate strong morphological variation, and in genera with rapid bindin evolution, species designations had to await the ability of molecular tools to assay genetic differences. The most diverse set of closely related sea urchin species known, the five or six Indo-West Pacific species of *Echinometra*, had been classified as a single, large polymorphic species by Mortensen (1928–1951). Only after reproductive barriers and genetic differences became clear were subtle morphological and ecological distinctions discovered (Matsuoka and Hatanaka, 1991; Palumbi and Metz, 1991; Uehara and Shingaki, 1985).

Another potential difference between bindin and plumage evolution is the driver producing different evolutionary rates. Sexual selection is thought to drive the divergence of male coloration in birds through a runaway process based on female preference (Kirkpatrick, 1982; Lande, 1981). Because female preference and male traits coevolve differently in separate isolated areas, different populations can attain novel trait and preference combinations (Ryan and Wilczynski, 1988). For bindin and its receptor genes, an excess of amino acid replacement substitutions is a signal that evolution is driven by selection, but the source of this selection remains unclear. It is possible that selection is for increasingly better fertilization systems, with the bindin genes evolving to produce more optimal fertilization phenotypes. Positive selection in this case would be

driven by a change in fertilization environment, such as water motion or distance from conspecifics, and would not be an ongoing process of male–female coevolution.

Alternatively, several coevolutionary scenarios have been proposed for rapid evolution of gamete recognition. Selection for divergent bindin alleles in sympatric species by a process of reinforcement has been observed in the most recently diverged *Echinometra*. The Okinawa and Hawaiian populations of *E. oblonga* diverged 250,000–500,000 years ago, yet the Okinawa population is highly distinct in bindin sequence and sperm morphology (Landry *et al*., 2003). Rapid evolution in the Okinawa population is related to the presence of sympatric populations of E. sp. C (Geyer and Palumbi, 2003). and is probably responsible for high levels of conspecific sperm precedence and low hybridization (Geyer and Palumbi, 2005). McCartney and Lessios (2004) suggested that rapid evolution of bindin in Caribbean *Echinometra* was due to divergence of the fertilization systems of two sympatric species. In this case, rapid bindin evolution in one species was associated with evolution of egg specificity in the same species (Lessios and Cunningham, 1990; McCartney and Lessios, 2002). A third case of rapid bindin evolution associated with gamete morphological change has been reported by Zigler *et al*. (2003). *Heliocidaris erythrogramma* has a giant egg and has evolved strong developmental differences from typical sea urchins with planktotrphic larvae, including its congener *H. tuberculata*. Bindin has evolved rapidly along the branch leading to the developmentally novel species but not along the branch leading to *H. tuberculata*.

Yet reinforcement or rapid gamete differentiation between species are not the sole evolutionary pressures acting on gamete recognition. Sea urchin species with bindin that is rapidly diverging between species are typically also highly polymorphic intraspecifically, and males with different bindin alleles produce functionally different sperm (Palumbi, 1999). Recent experiments with sperm mixtures also show that heterozygote males are superior fertilizers but only for certain allele combinations (S.R.P., unpublished data). Thus, the maintenance of polymorphism within species is probably due, at least in part, to balancing selection among alleles.

Another plausible mechanism of within- and between-species evolution of bindin is a coevolutionary antagonism between males and females. Males are selected to produce sperm that fertilize quickly and indiscriminately, whereas females are selected to control which single sperm is to be used for fertilizing a single egg (Palumbi, 1998; Rice, 2000; Swanson and Vacquier, 2002a,b). If so, then we would expect the egg receptor gene to be highly polymorphic and evolution toward more and more choosy eggs to be an ongoing process. To date there are no data on the egg receptor to test these predictions.

Why do some sea urchin genera have higher rates of bindin evolution? Bindin evolves by a combination of amino acid substitution and insertion or deletion of repeated amino acid motifs (Zigler and Lessios, 2003b). Zigler and Lessios (2003b) compared the primary sequence of the bindin protein between genera showing different evolutionary rates in this molecule. They found no obvious differences among genera with fast versus slow rates. Levitan (2002) argued that fertilization rates are determined by population densities of conspecific sea urchins, but there is no clear relationship between ecological density and the rate of bindin evolution among the genera.

Overall, the clear relationship between bindin evolution and young, sympatric species remains a strong signal that gamete recognition and species formation are tightly linked. But the underlying mechanisms driving this pattern remain poorly explained. Understanding the coevolution of bindin and the recently described sperm-receptor gene (Kamei and Glabe, 2003) may help make these mechanisms clearer.

CONCLUSIONS

Ernst Mayr's evolutionary animation is supported by recent molecular phylogenies of eight sea urchin genera at different stages of diversification. The details of these phylogenies reveal a tight association between sister-species status and geography at early stages. Molecular phylogenies also provide strong evidence for departures from Mayr's strict series. In particular, genera in which gamete recognition loci are more quickly evolving tend to fall at later stages in Mayr's evolutionary series. These genera are not necessarily comprised of older species. Instead, they can be composed of young species that have more rapidly evolved reproductive isolation and assortative mating. This result indicates that the assumption that reproductive isolation accumulates gradually with time does not hold when such isolation arises from changes in a single locus, instead of being the product of small changes in many loci. Nevertheless, the end result is basically identical to Mayr's assertion that sympatric species tend to be found in genera with greater reproductive isolation. By placing genera at different stages of diversification in a temporal series, Mayr animated the process of speciation and made its mechanisms clearer. The same basic approach remains valuable today and is all the more accurate when phylogenetic data permit the determination of the order in which species were separated from each other.

REFERENCES

Barraclough, T. G. & Vogler, A. P. (2000) Detecting the geographical pattern of speciation from species-level phylogenies. *Am. Nat.* **155,** 419–434.

Biermann, C. (1998) The molecular evolution of sperm bindin in six species of sea urchins (Echinoida: Strongylocentrotidae). *Mol. Biol. Evol.* **15,** 1761–1771.

Biermann, C. H., Kessing, K. & Palumbi., S. R. (2003) Phylogeny and development of marine model species: Strongylocentrotid sea urchins. *Evol. Dev.* **5,** 360–371.

Coates, A. G. & Obando, J. A. (1996) Geological evolution of the Central American isthmus. In *Evolution and Environment in Tropical America*, eds. Jackson, J. B. C., Coates, A. G. & Budd, A. (Univ. of Chicago Press, Chicago), pp. 21–56.

Dobzhansky, T. (1937) *Genetics and the Origin of Species* (Columbia Univ. Press, New York).

Dobzhansky, T. (1970) *Genetics of the Evolutionary Process* (Columbia Univ. Press, New York).

Ekman, S. (1953) *Zoogeography of the Sea* (Sidgwick & Jackson, London).

Geyer, L. B. & Palumbi, S. R. (2003) Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. *Evolution* **57,** 1049–1060.

Geyer, L. B. & Palumbi, S. R. (2005) Conspecific sperm precedence in two species of tropical sea urchins. *Evolution* **59,** 97–105.

Hillis, D. M., Moritz, C. & Mable, B. K. (1996) *Molecular Systematics* (Sinauer, Sunderland, MA), 2nd Ed.

Kamei, N. & Glabe, C. G. (2003) The species-specific egg receptor for sea urchin sperm adhesion is EBR1, a novel ADAMTS protein. *Genes Dev.* **17,** 2505–2507.

Kirkpatrick, M. (1982) Sexual selection and the evolution of female choice. *Evolution* **36,** 1–12.

Lande, R. (1981) Models of speciation by sexual selection on polygenic traits. *Proc. Natl. Acad. Sci. USA* **78,** 3721–3725.

Landry, C., Geyer, L. B., Arakaki, Y., Uehara, T. & Palumbi, S. R. (2003) Recent speciation in the Indo-West Pacific: rapid evolution of gamete recognition and sperm morphology in cryptic species of sea urchin. *Proc. R. Soc. London Ser. B* **270,** 1839–1847.

Lessios, H. A. (1984) Possible prezygotic reproductive isolation in sea urchins separated by the Isthmus of Panama. *Evolution* **38,** 1144–1148.

Lessios, H. & Cunningham, C. W. (1990) Gametic incompatibility between species of the sea urchin *Echinometra* on the two sides of the Isthmus of Panama. *Evolution* **44,** 933–941.

Lessios, H. A. & Pearse, J. S. (1996) Hybridization and introgression between Indo-Pacific species of *Diadema. Mar. Biol. (Berlin)* **126,** 715–723.

Lessios, H. A., Kessing, B. D., Wellington, G. M. & Graybeal, A. (1996) *Coral Reefs* **15,** 133–142.

Lessios, H. A., Kessing, B. D. & Robertson, D. R. (1998) Massive gene flow across the world's most potent marine biogeographic barrier. *Proc. R. Soc. London Ser. B* **265,** 583–588.

Lessios, H. A., Kessing, B., Robertson, D. R. & Paulay, G. (1999) Population structure and speciation in tropical seas: Global phylogeography of the sea urchin *Diadema*. *Evolution* **53,** 806–817.

Lessios, H. A., Kane, J. & Robertson, D. R. (2003a) Phylogeography of the pantropical sea urchin *Tripneustes*: Contrasting patterns of population structure between oceans. *Evolution* **57,** 2026–2036.

Lessios, H. A., Kessing, B. & Pearse, J. (2003b) Population structure and speciation in tropical seas: Global phylogeography of the sea urchin *Diadema*. *Evolution* **55,** 955–975.

Levitan, D. R. (2002) The relationship between conspecific fertilization success and reproductive isolation among three congeneric sea urchins. *Evolution* **56,** 1599–1609.

Matsuoka, N. & Hatanaka, T. (1991) Molecular evidence for the existence of four sibling species within the sea-urchin, *Echinometra mathaei* in Japanese waters and their evolutionary relationships. *Zool. Sci.* **8,** 121–133.

Mayr, E. (1942) *Systematics and the Origin of Species* (Columbia Univ. Press, New York).

Mayr, E. (1954) Geographic speciation in tropical echinoids. *Evolution* **8,** 1–18.

Mayr, E. (1963) *Animal Species and Evolution* (Harvard Univ. Press, Cambridge, U.K.).

McCartney, M. A. & Lessios, H. A. (2002) Quantitative analysis of gametic incompatibility between closely related species of neotropical sea urchins. *Biol. Bull. (Woods Hole, Mass.)* **202,** 166–181.

McCartney, M. A. & Lessios, H. A. (2004) Adaptive evolution of sperm bindin tracks egg incompatibility in neotropical sea urchins of the genus *Echinometra*. *Mol. Biol. Evol.* **21,** 732–745.

McCartney, M. A., Keller, G. & Lessios, H. A. (2000) Dispersal barriers in tropical oceans and speciation in Atlantic and Pacific sea urchins of the genus *Echinometra*. *Mol. Ecol.* **9,** 1391–1400.

McMillan, W. O., Raff, R. A. & Palumbi, S. R. (1992) Population genetic consequences of developmental evolution and reduced dispersal in sea urchins (genus *Heliocidaris*). *Evolution* **46,** 1299–1312.

Metz, E. C. & Palumbi, S. R. (1996) Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol. Biol. Evol.* **13,** 397–406.

Metz, E. C., Gomez, G. G. & Vacquier, V. D. (1998) Mitochondrial DNA and bindin gene sequence evolution among allopatric species of the sea urchin genus *Arbacia*. *Mol. Biol. Evol.* **15,** 185–195.

Mortensen, T. (1928–1951) *A Monograph of the Echinoidea* (C. A. Reitzel, Copenhagen).

Orr, H. A. (1991) Is single-gene speciation possible? *Evolution* **45,** 764–769.

Palumbi, S. R. (1992) Marine speciation on a small planet. *Trends Ecol. Evol.* **7,** 114–118.

Palumbi, S. R. (1996) What can molecular genetics contribute to marine biogeography? An urchin's tale. *J. Exp. Mar. Biol. Ecol.* **203,** 75–92.

Palumbi, S. R. (1998) Species formation and the evolution of gamete recognition loci. In *Endless Forms: Species and Speciation*, eds. Howard, D. & Berlocher, S. (Oxford Univ. Press, New York), pp. 271–278.

Palumbi, S. R. (1999) All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proc. Nat. Acad. Sci. USA* **96,** 12632–12637.

Palumbi, S. R. & Metz, E. (1991) Strong reproductive isolation between closely related tropical sea urchins (genus *Echinometra*). *Mol. Biol. Evol.* **8,** 227–239.

Rice, W. R. (2000) Dangerous liaisons. *Proc. Natl. Acad. Sci. USA* **97,** 12953–12955.

Ryan, M. J. & Wilczynski, W. (1988) Coevolution of sender and reciever: Effect on local mate preference in cricket frogs. *Science* **240,** 1786–1789.

Swanson, W. J. and Vacquier, V. D. (2002a) Rapid evolution of reproductive proteins. *Nat. Rev. Genet.* **3,** 137–144.

Swanson, W. J. and Vacquier, V. D. (2002b) Reproductive protein evolution. *Ann. Rev. Ecol. Syst.* **33,** 161–179.

Uehara, T. & Shingaki, M. (1985) Taxonomic studies in the four types of the sea urchin, *Echinometra mathaei*, from Okinawa, Japan. *Zool. Sci.* **2,** 1009.

Wyckoff, G. J., Wang, W. & Wu, C.-I. (2001) Rapid evolution of male reproductive genes in the descent of man. *Nature* **403,** 304–309.

Zigler, K. S. & Lessios, H. A. (2003a) Evolution of bindin in the pantropical sea urchin *Tripneustes*: Comparisons to bindin of other genera. *Mol. Biol. Evol.* **20,** 220–231.

Zigler, K. S. & Lessios, H. A. (2003b) 250 million years of bindin evolution. *Biol. Bull. (Woods Hole, Mass.)* **205,** 8–15.

Zigler, K. S. & Lessios, H. A. (2004) Speciation in the coasts of the New World: Phylogenography and evolution of the bindin in the sea urchin gneus *Lytechinus*. *Evolution* **58,** 1225–1241.

Zigler, K. S., Raff, E. C., Popodi, E., Raff, R. A. & Lessios, H. A. (2003) Adaptive evolution of bindin in the genus *Heliocidaris* is correlated with the shift to direct development. *Evolution* **57,** 2293–2302.

9

Mayr, Dobzhansky, and Bush and the Complexities of Sympatric Speciation in *Rhagoletis*

JEFFREY L. FEDER,[*] XIANFA XIE,[*] JUAN RULL,[†] SEBASTIAN VELEZ,[*] ANDREW FORBES,[*] BRIAN LEUNG,[*‡] HATTIE DAMBROSKI,[*§] KENNETH E. FILCHAK,[*] AND MARTIN ALUJA[†]

The *Rhagoletis pomonella* sibling species complex is a model for sympatric speciation by means of host plant shifting. However, genetic variation aiding the sympatric radiation of the group in the United States may have geographic roots. Inversions on chromosomes 1–3 affecting diapause traits adapting flies to differences in host fruiting phenology appear to exist in the United States because of a series of secondary introgression events from Mexico. Here, we investigate whether these inverted regions of the genome may have subsequently evolved to become more recalcitrant to introgression relative to collinear regions, consistent with new models for chromosomal speciation. As predicted by the models, gene trees for six nuclear loci mapping to chromosomes other than 1–3 tended to have shallower node

*Department of Biological Sciences, University of Notre Dame, Galvin Life Science Center, P.O. Box 369, Notre Dame, IN 46556-0369; [†]Instituto de Ecología, Asociación Civil, Kilometro 2.5 Antigua Carretera a Coatepec 351, 91070 Xalapa, Veracruz, Mexico; [‡]Department of Biology, McGill School of Environment, McGill University, Montreal, QB, Canada H3A 1B1; and [§]Cereal Disease Laboratory, United States Department of Agriculture, Agricultural Research Service, Midwest Area Office, University of Minnesota, 1551 Lindig Avenue, St. Paul, MN 55104.

Abbreviations: Mya, million years ago; ML, maximum likelihood; MP, maximum parsimony; RND, relative node depth.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AY152477–AY152526 and AY930466–AY931013).

depths separating Mexican and U.S. haplotypes relative to an outgroup sequence than nine genes residing on chromosomes 1–3. We discuss the implications of secondary contact and differential introgression with respect to sympatric host race formation and speciation in *Rhagoletis*, reconciling some of the seemingly dichotomous views of Mayr, Dobzhansky, and Bush concerning modes of divergence.

Ernst Mayr helped to transform speciation into a holistic science. With his influential book *Systematics and the Origins of Species*, Mayr (1942) integrated and synthesized information from genetics, natural history, biogeography, and phylogenetics into a coherent concept of a biological species and a theory for allopatric speciation. Mayr stressed the critical importance of biogeography and systematics as cornerstones for understanding speciation. Divorced from time, space, and phylogenetic relationship, the analysis of reproductive isolation (the defining characteristic of biological species) loses evolutionary context and meaning. The proper chronological ordering of taxa at various stages of divergence also becomes untenable, prohibiting evaluation of the type, sequence, and importance of ecological, demographic, and genetic factors leading to speciation. Therefore, Mayr (1942, 1963) presented a cogent strategy for studying speciation, clarifying the nature of the question and the critical parameters for investigating the process.

Despite widespread acceptance of Mayr's general framework for studying speciation, several seemingly dichotomous views and personalities nevertheless have shaped and still greatly influence our understanding of the process. Mayr (1942, 1963) made a forceful argument that geographic isolation (allopatry) is a requisite first step for facilitating divergence in animals. He stressed the coadapted nature of the genome and gene pools, as well as the need for allopatry to break the cohesive chains of gene flow to permit populations to diverge independently. In contrast, Guy Bush (1966, 1969) championed the importance of ecological adaptation in speciation. This view was epitomized in his arguments that certain phytophagous insect specialists speciate sympatrically in the process of shifting and adapting to new host plants. Theodosius Dobzhansky (1937, 1981) pioneered the genetic study of speciation, mapping genetic factors responsible for hybrid sterility and inviability and surveying natural populations to assess levels of genetic variation. He crystallized the view that speciation represents the transformation of within-population variation into between-taxa differences through the evolution of inherent reproductive isolating barriers. Dobzhansky (1981) also was a strong advocate of genetic coadaptation, especially with regard to balanced

polymorphisms in the form of chromosomal inversions. All of these themes are subsumed within the cladistic framework of Hennig and the paleontological perspective of Simpson; speciation represents population bifurcations in which an ancestral population is split into two distinct, descendent daughter lineages on separate evolutionary paths (Wiley, 1978).

However, many of the dichotomies that we envision concerning modes of divergence, the cladistic splitting of taxa, and systematic categories of organisms may blur during species formation. For example, it may not always be the case that geographic isolation is absolute or complete during active stages of species formation. Populations at different stages of divergence may experience periods of spatial isolation interspersed with episodes of contact and differential introgression. Sometimes the new genetic variation introduced by introgression may even open novel environments for populations, facilitating local adaptive divergence (Arnold, 1992). Other times, gene flow will homogenize much of the genome, leaving behind a core of coadapted genes differentially evolved between populations. When these gene complexes persist (because of strong selection and their likely linkage in regions of reduced recombination, for example, in inversions) they form a nucleus from which further divergence can build in sympatry by means of reinforcement or ecological specialization (Rieseberg, 2001), as well as in allopatry if isolation reoccurs. Thus, there has been an increasing realization over the last several decades that fields of recombination often extend beyond taxonomic boundaries (Carson, 1975; Hey, 2001; Templeton, 1989) and introgression can occur between hybridizing species in parts of their genomes but not others (Beltran *et al.*, 2002; Brown *et al.*, 2004; Butlin, 1998; Jiang *et al.*, 2000; Machado *et al.*, 2002; Noor *et al.*, 2001a; Rieseberg *et al.*, 1999; Wu, 2001). Moreover, although bifurcating phylogenies may be constructed for taxonomic groups viewed from the perspective of deep evolutionary time, forcing such a pattern on population divergence may misrepresent a more dynamic and reticulate speciation process (Beltran *et al.*, 2002; Brown *et al.*, 2004; Hewitt, 1989; Machado and Hey, 2003) and, hence, the evolutionary relationships of taxa. Rather than the analogy of a ''tree of life,'' a ''delta of life'' composed of many intertangled banks may be more appropriate in several instances.

Here, we investigate the biogeography of the *Rhagoletis pomonella* sibling species complex, a group of tephritid fruit fly specialists with a potentially reticulate genetics and history (Feder *et al.*, 2003a). Detailed study of the biogeography of *R. pomonella* flies may seem paradoxical. The four described and several undescribed sibling species constituting the complex are a model for ecological divergence without geographic isolation by sympatric host plant shifts (Bush, 1966, 1969). Moreover, the recent

shift of the species *R. pomonella* from its ancestral host hawthorn (*Crataegus* spp.) to introduced, domesticated apple (*Malus pumila*) within the last 150 years in the eastern United States is often cited as an example of host race formation in action, the hypothesized initial stage of sympatric speciation (Bush, 1966, 1969). However, we have discovered a surprising geographic source of genetic variation contributing to sympatric host shifts (Feder *et al.*, 2003a). Based on gene trees constructed for three anonymous nuclear loci mapping to separate rearrangements on chromosomes 1–3 of the *R. pomonella* genome, as well as mtDNA, we inferred that an ancestral, hawthorn-infesting fly population became geographically subdivided into Mexican and ''Northern'' (United States) isolates ≈1.57 million years ago (Mya). Episodes of gene flow from the Altiplano highland fly population in Mexico subsequently infused the Northern population with inversion polymorphism affecting key diapause traits, forming adaptive clines. Later, diapause variation in the latitudinal inversion clines appears to have aided flies in the United States in shifting and adapting to various new plants with different fruiting times. These shifts were mediated by population-level changes in allele (inversion) frequencies, generating premating and postmating reproductive isolation in the process and helping to spawn several new host-specific taxa, including the recently formed apple race. We stress that we are not contending that the *R. pomonella* complex in the United States evolved in allopatry. Rather, certain raw genetic material contributing to the adaptive radiation of *R. pomonella* in the United States originated in a different time and place than the proximate ecological host shifts triggering sympatric divergence.

The evidence for past introgression and its contribution to sympatric host shifts could be interpreted as indicating that inversions preferentially flowed from the Mexican Altiplano into the Northern fly population after secondary contact. However, the persistence of latitudinal clines in the United States suggests that environmental factors may have constrained the spread (prevented the fixation) of the inversions relative to other genes. Hawthorns tend to fruit later in southern latitudes (H.D. and J.L.F., unpublished data). Hawthorn-fly populations in the United States track this geographic variation in host phenology, possessing inversion genotypes for chromosomes 1–3 in the ''South'' that cause them to eclose later in the season (Feder and Filchak, 1999; Feder *et al.*, 2003a; Filchak *et al.*, 2000). The pattern continues into Mexico. In the Altiplano, flies infest their primary hawthorn hosts, *Crataegus mexicana* and *Crataegus rosei* var. *rosei*, from mid-October to late December (J.R., J.L.F., S. Berlocher, and M.A., unpublished data). However, in the United States, *R. pomonella* infests various different hawthorn species, mainly from mid-August to late October. Mexican flies take significantly longer to eclose than U.S. flies, even those from Texas (H.D., J.L.F., J.R., S. Berlocher, and M.A., unpub-

lished data). Consequently, the positions of the inversion clines represent a balance between diapause selection and migration. In contrast to the inversions, loci mapping to other chromosomal regions generally do not differ in allele frequency between the host races, vary clinally, correlate with the timing of eclosion, nor display high levels of linkage disequilibrium in nature (Berlocher, 2000; Berlocher and McPheron, 1996; Feder *et al.*, 1990, 1993, 2003b). Therefore, these apparently collinear regions of the genome may have introgressed more readily at times in the past between Mexico and the North, homogenizing in frequency because of a lack of differential selection combined with recombination.

The contrasting pattern of genetic differentiation seen for chromosomes 1–3 vs. other genomic regions is consistent with new models of chromosomal speciation (Navarro and Barton, 2003; Noor *et al.*, 2001b; Rieseberg, 2001). In these models, reduced recombination associated with rearrangements facilitates the retention of linked genes conferring adaptation or reproductive isolation between hybridizing taxa. However, collinear portions of the genome tend to introgress because recombination results in weak or no linkage of most genes in these regions to loci causing reproductive isolation. Studies in sunflowers (Rieseberg *et al.*, 1999), the *Drosophila pseudoobscura* subgroup (Brown *et al.*, 2004; Machado *et al.*, 2002; Noor *et al.*, 2001a), and *Anopheles* mosquitoes (Besansky *et al.*, 2003; dellaTorre *et al.*, 1997) have found evidence for greater introgression in collinear segments of the genome than inverted segments. If differential introgression is true also for *Rhagoletis*, then the prediction is that loci mapping outside the inversion carrying chromosomes 1–3 should generally show less genetic divergence between Altiplano and U.S. flies compared with genes within the rearranged chromosomes. Coalescence times for noninverted regions should primarily date to the most recent period of contact and gene flow; rearrangements should display deeper divergence times congruent with the initial separation of Mexican and Northern populations. Thus, the chromosome model is predicated on *Rhagoletis* inversions having partially introgressed at a distant time in the past. During subsequent periods of geographic isolation between Mexican and Northern populations, these inverted regions accumulated additional host-related, as well as possibly non-host-related, genetic changes. Some of the changes, because of their linkage in rearrangements differing between the populations, reduced the potential for the inversions to introgress between Mexican and U.S. flies.

Here, we examine the applicability of the ''rearrangement'' model to *R. pomonella* by means of an expanded DNA sequence analysis of loci encompassing both inverted and likely collinear regions of the genome of the fly. We report a pattern of genetic differentiation that is consistent with the rearrangement hypothesis for differential gene flow; gene trees

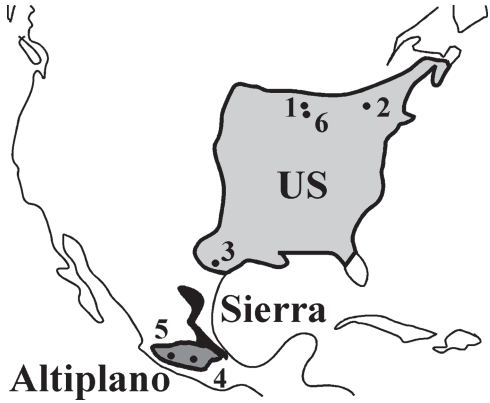


FIGURE 9.1 The current range of *R. pomonella* in North America. Estimated distributions for the hawthorn-infesting U.S. (light gray) and Mexican Altiplano (dark gray) populations of flies, as well as the recently discovered Sierra Madre Oriental population (black; see text for discussion of ''Sierra'' flies), are shown. The fly is also distributed patchily in the western United States. Further work is needed to clarify the distribution and origins of these western populations, because they may represent recent introductions. Numbers indicate sampling sites in study (see the Fig. 9.2 legend for site descriptions).

for six nuclear loci mapping to chromosomes other than 1–3 tended to have shallower relative node depths (RNDs) separating Mexican and U.S. sequences than nine genes residing on chromosomes 1–3. We discuss the implications of secondary contact and differential gene flow with respect to sympatric host race formation and speciation in *Rhagoletis*.

MATERIALS AND METHODS

Fly Populations

Taxa, host plants, collecting sites, and sampling dates for flies are given in Figs. 9.1 and 9.2. Flies were collected as larvae in infested fruit and either (*i*) dissected from the fruit and frozen for later genetic analysis or (*ii*) reared to adulthood in the laboratory.

DNA Sequencing

Sequence data were generated for 16 nuclear loci isolated from an *R. pomonella* EST library, in addition to the three nuclear genes (P220, P2956, and P7) and mtDNA (3′ portion of COI, tRNA-Leu, and COII) analyzed in Feder

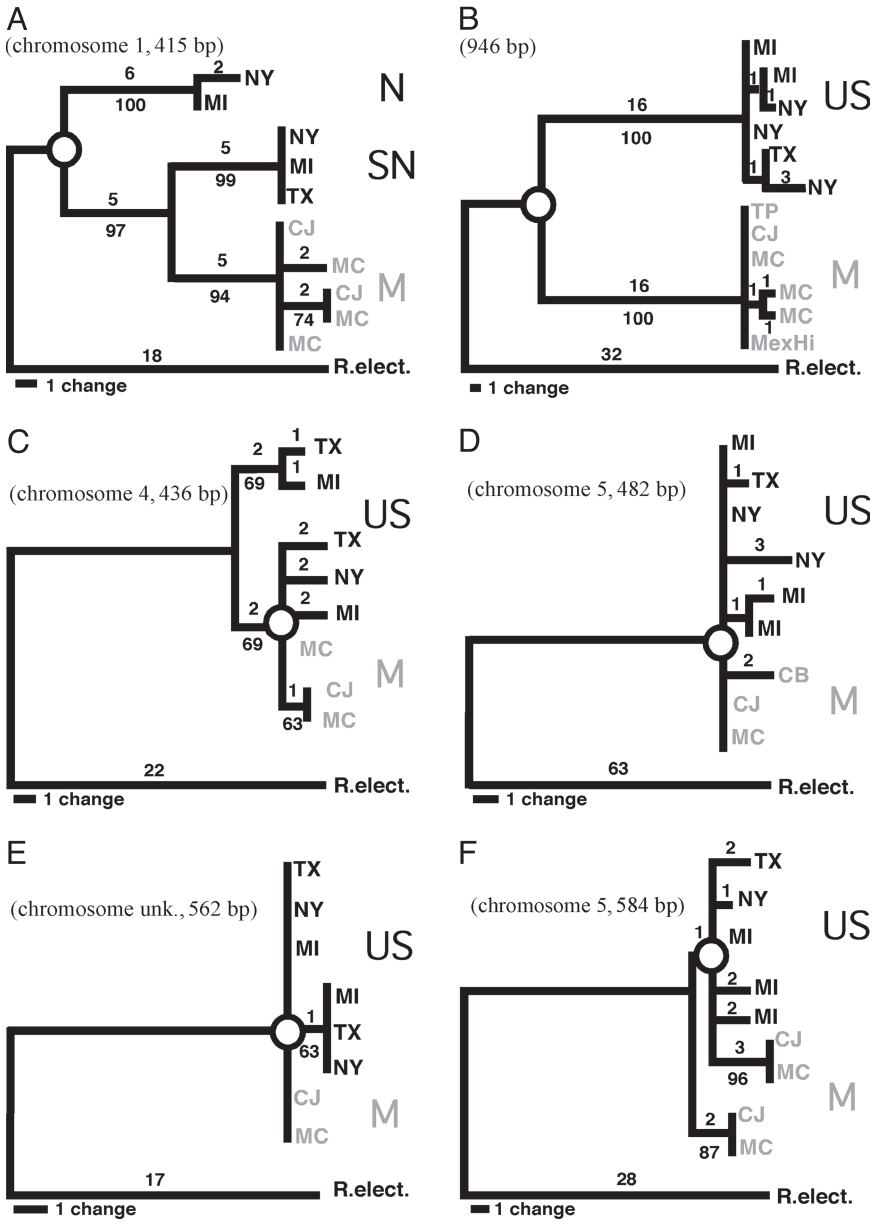


FIGURE 9.2 MP gene trees for P220 (*A*), mtDNA (*B*), P661 (*C*), P309 (*D*), P3060 (*E*), and P2620 (*F*). Trees are scaled so that the longest distance from an allele to the outgroup *R. electromorpha* (*R. elect.*) are relatively the same across loci. Chromosome position for loci, bootstrap support for nodes (10,000 replicates),

et al. (2003a). Nine of the new loci map to the inversion containing chromosomes 1–3, whereas seven genes reside elsewhere in regions that genetic data suggest are not associated with rearrangements (Berlocher, 2000; Berlocher and McPheron, 1996; Feder *et al.*, 1990, 1993, 2003b; Roethele *et al.*, 2001) (Table 9.1). Genomic DNA were PCR amplified for 35 cycles (94°C for 30 sec, 52°C for 1 min, and 72°C for 1.5 min) by using locus-specific primers (Roethele *et al.*, 2001). Products were TA-cloned into pCR II vectors (Invitrogen). PCR amplification products initially were cloned separately for two to three flies from each study site, with from four to six clones sequenced per locus per fly in the 5′ and 3′ directions on an ALF sequencer (Amersham Pharmacia Biotech). To increase sample sizes for certain sites, we also separately amplified genomic DNA for eight flies from the site and TA-cloned the pooled amplification products for sequencing. To avoid analysis of identical alleles from the same individual, sequences generated from the pooled cloning were not included unless they differed from each other.

Gene-Tree Construction

Maximum parsimony (MP) and maximum likelihood (ML) gene trees were constructed by using PAUP* (Version 4.0, Beta 10; Swofford, 2002). For the MP analysis, deletions were treated as a fifth base pair, with indels of identical length and position recoded to count as single mutational steps. *Rhagoletis electromorpha*, belonging to the sister species group (*Rhagoletis tabellaria*) to *R. pomonella* (Bush, 1966), was used as an outgroup.

---

sequence lengths (in bp), and branch lengths (no. of steps) are given. Exact location for P3060 is not known, but the locus does not map to chromosomes 1–3. For locus 220 in *A*, haplotypes are abbreviated as follows: N, North United States; SN, South/North United States; and M, Mexican (gray). For the other four nuclear loci and mtDNA, U.S. haplotypes are black and Mexican haplotypes are gray. Open circles indicate the contrast between deep and shallow RNDs shown for chromosome 1–3 loci and mtDNA (*A* and *B*) vs. genes that do not reside in rearrangements (*C*–*F*). Taxa, hosts, location, and collecting dates (month/day/year) are as follows: MI, *R. pomonella* [apple (*M. pumila*) and hawthorn (*C. mollis*)], Grant, MI, 8/15/95; NY, *R. pomonella* [hawthorn (*C. mollis*)], Geneva, NY, 9/16/00; TX, *R. pomonella* [hawthorn (*C. mollis*)], Brazos Bend, TX, 10/6/89; CJ, *R. pomonella* n.s. [hawthorn (*C. mexicana*)], Coajomulco, Morelos, Mexico, 11/12/02; MC, *R. pomonella* n.s. [hawthorn (*C. mexicana* and *C. rosei rosei*)], Tancitaro, Michoacan, Mexico, 11/15/02; and *R. electromorpha* [gray dogwood (*Cornus racemosa*)], Dowagiac, MI, 9/12/99. Gene trees do not include all of the sequenced alleles for each locus, but subsets that encapsulate the general topological structure for trees. Also, networks incorporating recombinant alleles are not shown for P661 and P3060. Additional alleles and networks are given in supporting information.

MP and ML gene trees were very similar, and, thus, only the MP results are presented. Intragenic recombination was tested by using the method of Hudson and Kaplan (1985). Putative recombinant alleles and gene regions were identified, and the alleles were excluded from initial MP gene-tree construction. Recombinant alleles were then added to the trees by hand to generate sequence networks. The molecular clock was tested for each locus for *R. pomonella* and *R. electromorpha* sequences by comparing log likelihood scores enforcing vs. relaxing the clock hypothesis for the best supported DNA substitution model identified by using MODELTEST (Posada and Crandall, 1998). To quantify gene-tree topology and genetic divergence, RNDs were calculated between major haplotype classes of alleles segregating in Mexican and U.S. fly populations by dividing the number of substitution differences between a given pair of Mexican and U.S. alleles by the mean number of substitutions between each of these alleles and the *R. electromorpha* outgroup sequence. Assuming a molecular clock (which none of the nuclear loci or mtDNA violated; Table 9.1), the mean RND for all pairs of Mexican and U.S. alleles between two haplotype classes estimates the age of separation of the haplotypes relative to the divergence of *R. electromorpha*, given a low to moderate effective size for the ancestral *R. pomonella/R. electromorpha* population.

RESULTS AND DISCUSSION

Nuclear and mtDNA Gene Trees

Of the 19 total nuclear loci analyzed in the study, 4 were determined to be duplicated loci and excluded from further analysis (P341, P2480, P70, and P2919, mapping to chromosomes 1, 3, 3, and 6, respectively). MP gene trees for the remaining 15 nuclear loci and mtDNA are shown in Fig. 9.2 and supporting information, which is published on the PNAS web site). None of the sequenced genes deviated significantly from a molecular clock (Table 9.1). Nine of the 15 nuclear loci displayed evidence for possible recombination by the method of Hudson and Kaplan (Table 9.1). However, exchange was limited to alleles within identified haplotype classes (i.e., M, S/N, or N) or within geographic populations (Altiplano or United States). The only exceptions were the loci P667, P1700, and P2473, where recombination occurred between major haplotypes within the U.S. population (see supporting information). As would be expected for a nonrecombinant molecule, there was no evidence for exchange among mtDNA sequences.

TABLE 9.1 Loci Sequenced in This Study

Locus	*c*	*P*	Model	*r*	RND
P181	1	0.89	F81	0	0.375, 0.710
P220	1	0.62	TrN + I	0	0.402, 0.677
P3072	1	0.32	TrN + I	1	0.298
P2473	2	0.12	TrN + I	1	0.378, 0.651
P2956	2	0.64	TrN + I	0	0.669, 0.755
P667	2	0.78	TrN	2	0.287, 0.391
P8	2	0.70	HKY	1	0.656
P22	3	0.48	HKY	0	0.281, 0.439
P7	3	0.32	HKY	1	0.281, 0.639
P2963	4	0.30	F81	3	0.323
P661	4	0.40	TrN + I	1	0.134, 0.267
P1700	5	0.24	HKY	4	0.313, 0.361
P2620	5	0.73	F81	0	0.144, 0.155
P309	5	0.73	HKY	0	0.078
P3060	?	0.75	HKY	1	0.158
mtDNA	—	0.43	TrN + I	0	0.691

NOTE: The chromosomem appositions (*c*), the probability level (*P*) for whether loci conform to a molecular clock, and the ML substitution model (Model) as determined by MODELTEST (Akaike information criterion estimate; Posada and Crandall, 1998) are given. The exact map position of P3060 is not known, but the gene does not reside on chromosomes 1–3. Also, it is shown whether genes display evidence for recombination (*r,* minimum no. of recombination events as estimated by the method of Hudson and Kaplan, 1985), and RND values for loci between major haplotype classes segregating in Mexican and U.S. populations. Cases in which two haplotypes are segregating in the U.S. population (or Altiplano for P1700) have two RND values (one value for each haplotype).

Gene Tree Topologies and RND

Gene tree topologies differed significantly between loci mapping to chromosomes 1–3 and those residing elsewhere in the genome (Fig. 9.2). A summary of the differences is shown in Fig. 9.3, where RNDs are plotted between major haplotype classes segregating at loci in Altiplano vs. U.S. flies. RNDs clustered into three groups, corresponding to deep, intermediate, and shallow divergence between Mexican and U.S. haplotypes. Loci tended to fall into different RND categories based on their chromosomal location. Loci mapping to chromosomes 1–3 had significantly greater RNDs than genes on other chromosomes. Six of the nine loci on chromosomes 1–3, as well as mtDNA, had RNDs >0.63 between at least one pair of haplotypes segregating in the United States and Altiplano (Table 9.1). Two of the three loci not displaying deep RNDs (P3072 and P667) showed low levels of disequilibrium, with linked allozymes differentiating the apple and hawthorn host races (standardized disequilib-

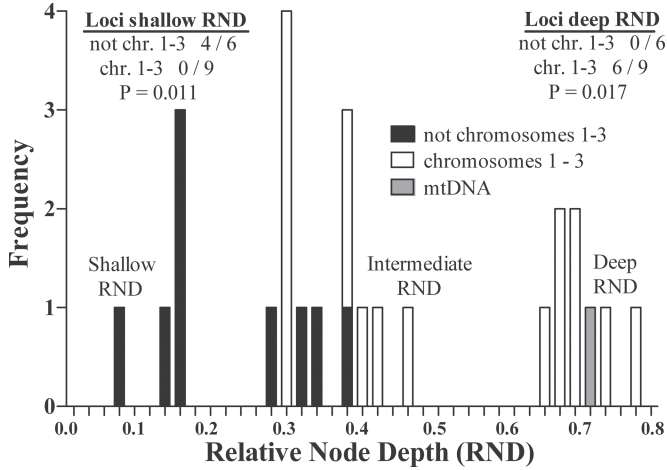


FIGURE 9.3 Distribution of RNDs for nuclear loci not residing on chromosomes 1–3 (black bars), for genes located on chromosomes 1–3 (white bars), and for mtDNA (gray bar). The list on the left gives the number of loci displaying shallow RNDs (<0.16) for the six sequenced genes not mapping to chromosomes 1–3 vs. the nine genes that do. The list on the right gives the number of loci displaying deep RNDs (>0.63). RND values for each locus are given in Table 9.1. *P* values were determined by two-tailed Fisher's exact tests.

rium between P3072 and *Aat-2*, 0.077; $P = 0.504$; $n = 75$ scored chromosomes; *r* value P667/*Me*, 0.117; $P = 0.259$; $n = 92$), suggesting possible weaker associations of these genes with inversions or targets of selection on chromosomes 1 and 2, respectively. In contrast, none of the six loci residing on chromosomes other than 1–3 possessed a deep RND (Table 9.1). Indeed, the deepest RND for any of these six loci was 0.361 (P1700), which was shallower than P22, the third locus on chromosome 3 not possessing a deep RND. Four of the six loci not on chromosomes 1–3 also displayed shallow RNDs of <0.16, not appreciably greater than values found segregating within haplotype classes for these loci within Altiplano and U.S. populations (Fig. 9.2). No locus residing on chromosomes 1–3 possessed a shallow RND (Table 9.1).

Implications of the Gene Trees: Isolation, Contact, and Differential Gene Flow

The tripartite distribution of RNDs for nuclear and mtDNA gene trees is consistent with a hypothesis that Mexican and U.S. fly populations have undergone two cycles of geographic isolation and differential intro-

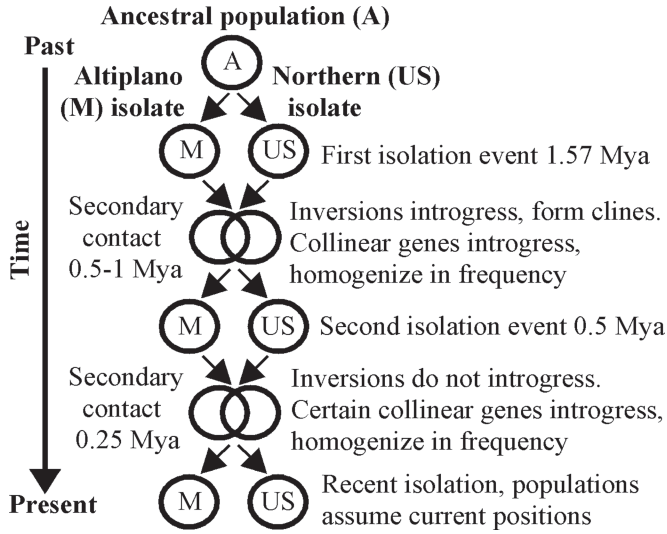


FIGURE 9.4 Biogeographic model depicting two cycles of isolation and differential introgression between Mexican Altiplano and Northern (United States) populations of *R. pomonella*.

gression (Fig. 9.4). The deep and congruent RNDs for six of the loci on chromosomes 1–3 and mtDNA suggest an initial population subdivision of a Mexican/U.S. common ancestor/1.57 Mya based on an insect mtDNA clock (1.15×10^{-8} substitutions per bp per year) (Brower, 1994). We propose that this initial isolation event was followed by a period of contact from 0.5–1 Mya, during which time gene flow was considerable. Extensive population mixing accounts for the large number of loci displaying intermediate RNDs, as well as for the establishment of adaptive clines for inversions on chromosome 1–3. We do not know the location or extent of the contact zone or clines when they first formed. However, we presume that ecological factors related to host phenology affected the clines in the past in a similar manner as they do currently. Loci residing in other regions of the genome not under selection moved readily between Mexican and Northern populations and recombined, accounting for the lack of deep RNDs for chromosome 4 and 5 loci. In contrast, mtDNA did not introgress during this or any subsequent period of contact.

We hypothesize that the initial period of contact was followed by a second cycle of isolation and introgression (Fig. 9.4). Gene flow was differential during the most recent contact period. Loci residing on chromosomes 4 and 5 tended to move readily between populations, accounting for the shallow RNDs observed for most (4/6; 67%) of these genes (Fig. 9.3

and Table 9.1). In comparison, loci on chromosomes 1–3 did not introgress, resulting in a lack of shallow RNDs. The pattern of gene flow suggests that genetic differences accumulated on chromosomes 1–3 during the second isolation period. To the extent that these changes are defined by inversions (a supposition supported by genetic cross data and population-level linkage disequilibrium values within U.S. populations), they concur with rearrangement models of chromosomal speciation (Rieseberg, 2001; Noor *et al.*, 2001b). Also, the accumulation of additional inversion changes after the hypothesized time when clines were first established suggests that not all diapause-related differences among U.S. flies trace to Mexican origins.

Alternative Hypotheses for the Gene Trees

The pattern of differentiation seen for nuclear loci could potentially also be explained by incomplete linage sorting of balanced inversion polymorphisms present in the ancestral Mexican/U.S. population. In this scenario, Mexican and Northern isolates diverged recently from a common ancestor of modest population size, accounting for the shallow RNDs for loci mapping to chromosomes 4 and 5. In contrast, rearranged regions on chromosomes 1–3 tend to have deeper RNDs because of (*i*) limited recombination between inversion karyotypes (Mexican and U.S. haplotypes on alternate inversions may often be restricted from coalescing until before the origin of the chromosomal rearrangement separating them in the common ancestor) and (*ii*) the increased retention time of rearrangements in the ancestral population due to overdominance. At the time of population subdivision, the inversions may have been arrayed in the form of primary clines. Inversions prominent in the South consequently sorted into the Mexican fly population, while a large portion of the polymorphism was retained in the North. As a result, SN haplotypes (alleles that now vary clinally and are found in increasing frequency in southern U.S. fly populations) are genetically more closely related to M haplotypes in Mexico than to alternate N haplotypes segregating in the same host populations (Fig. 9.2*A* and supporting information).

However, in the absence of a mechanism that coordinately generates inversions throughout the genome, the incomplete lineage-sorting hypothesis has difficulty explaining the clustered distribution of RND values for chromosome 1–3 loci into intermediate and deep categories (Fig. 9.3). Correlated RND values may be expected among loci residing in the same inverted region of a chromosome but not among rearranged regions on different chromosomes, as noted. Moreover, incomplete lineage sorting cannot readily account for the deep RND seen for mtDNA and its congruence with many chromosome 1–3 loci (Figs. 9.2 and 9.3 and Table

9.1). Given a recent time of separation and modest effective size for the ancestral population, mtDNA should have coalesced quickly and should display minimal differentiation between Mexican and U.S. flies. Last, although inverted regions can be biased toward containing haplotypes with deeper RNDs, unless population splitting was precise, one would still expect to see a subset of inversions shared in common between Mexican and U.S. flies. Haplotypes in the shared inversions should show shallow RNDs, similar to loci on chromosomes 4 and 5. Consequently, the observed gene trees are more consistent with the hypothesis of repeated isolation and secondary contact, with inversions on chromosomes 1–3 becoming increasing more recalcitrant to introgression through time relative to collinear regions of the genome.

Our data could also be explained by a series of gene duplication and deletion events within *R. pomonella* and the outgroup species *R. electromorpha* such that many of the haplotype comparisons made in the study were between paralogous rather than orthologous sequences. Four of the original 19 loci amplified in the study were found to be duplicate loci. If similar duplications were accompanied by deletions for many of the other 15 loci, then these duplications/deletions could confound our biogeographic interpretation of the gene trees. However, the deletion scenario, considered alone, suffers the same difficulties as the lineage-sorting hypotheses in explaining the tripartite distribution and deep congruence of chromosome 1–3 nuclear and mtDNA RND values. But it is possible that a composite biogeography/deletion model could account for the pattern. Under this scenario, Mexican and Northern isolates formed ≈1.57 Mya. A period of secondary contact and gene flow followed from 0.5–1 Mya. After this time, Altiplano and Northern populations have remained disjunct. The shallow RNDS observed for loci not on chromosomes 1–3 would be due reciprocal deletions of paralogous genes in *R. pomonella* and *R. electromorpha*, resulting in improper comparisons of orthologous Mexican and U.S. haplotypes within *R. pomonella* to a highly diverged paralogous outgroup sequence for *R. electromorpha*.

The deletion hypothesis would not negate the contributory roles of allopatry and secondary introgression in facilitating the sympatric radiation of the *R. pomonella* group by means of host shifting. However, it would call into question whether gene flow was differential for inverted vs. collinear regions of the genome. In essence, there would not have been a second period of recent contact when such a pattern could have been fully generated. Genetic crosses of flies imply that U.S. haplotypes represent allelic variation segregating at single loci (Feder *et al.*, 2003b; Roethele *et al.*, 2001). However, it is difficult to completely rule out the possibility that deletions at very tightly linked duplicated loci generated the observed segregation patterns. Moreover, test cross results for *R. pomonella*

are not directly germane to resolving the status of *R. electromorpha* sequences. However, sequence data available for the more distantly related *R. cingulata* and *R. suavis* for P661, P309, P2620, and P3060 (loci with shallow RNDs) place *R. electromorpha* between these two species and *R. pomonella*. The lack of interspersed clades of sequences containing all or subsets of the four species implies that variation at P661, P309, P2620, and P3060 is allelic and not paralogous.

A Second Mexican Population

Recently, we have discovered a second population of *R. pomonella*-like flies that infest hawthorns in the Sierra Madre Oriental Mountains of Mexico (Fig. 9.1). The genetics, biogeography, and phenology of the Sierra Oriental population suggest that it may have been a conduit for gene flow between the Altiplano and the North in the past (J.R., J.L.F., X.X., S. Berlocher, and M.A., unpublished data). DNA sequence analysis indicates that Sierra flies are differentiated but, overall, appear to be most closely related to southern U.S. populations (X.X. and J.L.F., unpublished data). The Sierra population abuts the Altiplano population through parts of the states of Veracruz, Puebla, and Hidalgo, Mexico (Fig. 9.1) (J.R., J.L.F., X.X., S. Berlocher, and M.A., unpublished data). We do not know whether the Sierra population contacts U.S. flies. However, if it does, this contact zone is spotty and ephemeral. Hawthorns are rare through the border region but are present in isolated patches in southern New Mexico and, possibly, the Davis Mountains of Texas. The primary hawthorn host for Sierra flies is *C. rosei* var. *parrayana*, which is infested from September to early October (J.R., J.L.F., S. Berlocher, and M.A., unpublished data). As is the case for Altiplano and U.S. flies, the diapause characteristics of the Sierra population match host phenology. Sierra flies eclose significantly earlier than Altiplano flies, resulting in potentially substantial allochronic isolation (J.R., J.L.F., X.X., S. Berlocher, and M.A., unpublished data). However, host specificity is not absolute in Mexico. In the transition region between the Altiplano and Sierra, *C. mexicana* and *C. rosei rosei* cooccur with *C. rosei parrayana* and can be found infested by genetically Sierra populations of flies. Here, *C. mexicana* and *C. rosei rosei* fruit earlier than they do on the Altiplano and are infested from late September to early November. Thus, host specificity is not as critical a factor isolating Mexican flies as it is for the *R. pomonella* complex in the United States. However, the spatial and temporal overlap of hawthorns in the transition zone provides a potential bridge for past introgression between the Altiplano and North via the Sierra population.

A Golden Braid

The views of Mayr, Dobzhansky, and Bush may not be as trichotomous as they seem with respect to *Rhagoletis*. Geographic isolation appears to have established an initial kernel of genetic differentiation that was later expanded on and contributed to sympatric host shifts and new fly taxa. Thus, although geographic context is critical for understanding speciation, allopatry and sympatry should not always be considered as diametrically opposed modes of divergence along an axis of spatial isolation. Differentiation and processes occurring in isolation and contact can interact and compliment each other to accentuate species formation, arguing for a more pluralistic view of modes of speciation (Mallet, 2005). In the case of *R. pomonella*, the relationship involves a likely sequence of geographic isolation, life-history adaptation, secondary contact, differential introgression, inversion clines, and sympatric host shifts. The evolution of reinforcement can be viewed in an analogous manner, involving non-host-related traits affecting prezygotic isolation rather than ecological adaptation *per se*. Also, there is no reason to presume that host-related differences that originated in sympatry cannot be solidified by periods of geographic isolation between host-associated populations, although such allopatry is not required to complete the speciation process. Thus, during the time course of differentiation, populations can assume characteristics of both allopatric and sympatric modes of divergence, with phenotypic and genetic elements interacting to further the speciation process.

The connectivity of speciation mode is perhaps best epitomized for *R. pomonella* if one views the phylogeography of the fly as reflecting sequential adaptation to spatially more finely packaged phenological host niches. At the coarsest level, Altiplano, Sierra Oriental, and Northern *R. pomonella* populations initially became differentially adapted to temporal and spatial disjunctions in hawthorn fruiting time through a ''modular'' genetics associated with inversions affecting diapause. After secondary introgression from Mexico, the modular gene blocks became arrayed in the formof broad inversion clines in the North in response to latitudinal variation in hawthorn fruiting time. Last, life-history variation inherent in the clines was extracted on a microgeographic scale [primarily by shifts in allele (inversion) frequencies] to facilitate sympatric shifts and specialization of *R. pomonella* in the United States to a number of cooccurring host plant species with differing fruiting times. However, host specificity does not appear to be a factor reproductively isolating Altiplano and Sierra flies. Here, geography may act as habitat fidelity does in sympatry, limiting migration and facilitating divergence.

The differences between Altiplano, Sierra, and U.S. populations raise a number of questions. For example, the apparently reduced potential for

rearrangements to introgress implies that these regions of the genome have accumulated additional genetic changes, causing reproductive isolation between Mexican and U.S. flies after their initial establishment in secondary inversion clines in the North. Not all of these changes necessarily reflect host-associated or ecological adaptations. There is no reason that non-host-related differences resulting in prezygotic isolation and hybrid inviability and sterility should not also have accumulated between Mexican and Northern demes during periods of allopatry. Given secondary contact and differential gene flow, the chromosomal speciation models predict that these differences should be concentrated in inversions (Arnold, 1992; Noor *et al.*, 2001b). Therefore, the extent to which intrinsic genomic incompatibilities map to inversion differences between Mexican and U.S. flies needs to be examined. If true, then the inversions would be simultaneously affecting speciation across both allopatric and sympatric scales in *R. pomonella*. It would be particularly intriguing if any derived differences in the inversions in the U.S. related to latitudinal variation in hawthorn fruiting phenology or interactions between sympatric *R. pomonella* taxa using different hosts feed back to restrict gene flow between U.S. and Mexican flies, closing the speciation mode braid.

Also, we stress that not all of the host-related changes contributing to sympatric host shifts are diapause-related. Differences in host discrimination (habitat-specific mating) also played a key role in generating the *R. pomonella* complex. Recently, we demonstrated that host fruit-odor discrimination is an important element of habitat choice for *R. pomonella* (Linn *et al.*, 2003, 2004). Host choice is important in *Rhagoletis* because the fly mates only on or near the fruit of its respective host plants (Prokopy *et al.*, 1971, 1972). Thus, variation in host choice translates directly into differences in mate choice and prezygotic isolation. The genetics of fruit-odor discrimination appear to involve loci affecting both preference and avoidance for volatile compounds emitted from the surface of natal and nonnatal fruit (H.D. and J.L.F., unpublished data). Also, F1 hybrids appear to have a reduced ability to orient to host fruit odor in flight-tunnel tests, signifying potentially reduced fitness in the field (Linn *et al.*, 2004). Therefore, the genetics and evolutionary history of host discrimination may prove to be different from diapause traits and not associated with periods of geographic isolation. Because hawthorns are fundamentally similar in Mexico and the United States, there is no reason to expect hawthorn discrimination to be under differential selection or to display a cline.

In conclusion, our study highlights the reticulate nature of speciation at both the population and genomic levels. Students of plant speciation have long embraced this perspective (Anderson, 1949; Arnold, 1992; Grant, 1971; Stebbins, 1959), whereas workers in animal systems are gaining an increased appreciation for the importance of hybridization in meta-

zoan diversity (Arnold, 1992; Coyne and Orr, 2004; Dowling and Secor, 1997; Grant and Grant, 2002; Vollmer and Palumbi, 2002). Many questions remain. Genetic crosses are needed between Altiplano, Sierra, and U.S. flies to assess their taxonomic status and more accurately define the extent of inversion differences separating the populations to strengthen tests for differential introgression. Polytene chromosome spreads are of such poor quality in *Rhagoletis* that these questions cannot be answered cytologically, but it is nevertheless important to determine whether, for example, SN and M haplotypes now reside on the same or different sets of inversions in U.S. and Mexican populations. Preliminary mating studies indicate that Mexican and U.S. flies are interfertile. However, the relative sterility and viability of F1 and second-generation hybrids remain to be quantified. Moreover, our current understanding of the biogeography of Mexico must be refined, especially in the potential contact zone between Altiplano and Sierra populations, as well as Sierra and U.S. flies, to test for active gene flow. The cause for the lack of mtDNA introgression must also be resolved. Two possibilities are male-mediated gene flow and cytonuclear gene interactions affecting host choice. Last, the paleobiology of Mexico and the Southwest must be further investigated to determine whether the distributions of cooccurring fauna and flora, as well as environmental conditions, are consistent with our historical hypothesis for differential gene flow in *R. pomonella*. Nevertheless, our results underscore Mayr's (1942, 1963) emphasis of the critical importance for a fully resolved biogeography and systematics for understanding speciation, even in cases of sympatric divergence in which attention usually is focused on documenting spatial overlap during differentiation. Knowledge of historical information on the biogeography and phylogeography of *R. pomonella* has helped clarified our understanding of the mechanism of sympatric speciation in these flies by adding a contributory, secondary role for allopatrically evolved inversions in the process.

ACKNOWLEDGMENTS

We thank S. Berlocher, N. Lobo, M. Pale, U. Stolz, and the University of Notre Dame sequencing facility. This work was supported by grants from the National Science Foundation, the United States Department of Agriculture, and the State of Indiana 21st Century Fund (to J.L.F.), and the Mexican Campaña Nacional Contra Moscas de la Fruta and the Instituto de Ecología, Asociación Civil (to J.R. and M.A.).

REFERENCES

Anderson, E. (1949) *Introgressive Hybridization* (Wiley, New York).

Arnold, M. L. (1992) Natural hybridization as an evolutionary process. *Annu. Rev. Ecol. Syst.* **23,** 237–261.

Beltran, M., Jiggins, C. D., Bull, V., Linares, M., Mallet, J., McMillan, W. O. & Bermingham, E. (2002) Phylogenetic discordance at the species boundary: Comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Mol. Biol. Evol.* **19,** 2176–2190.

Berlocher, S. H. (2000) Radiation and divergence in the *Rhagoletis pomonella* species group: inferences from allozymes. *Evolution (Lawrence, Kans.)* **54,** 543–557.

Berlocher, S. H. & McPheron, B. A. (1996) Population structure of *Rhagoletis pomonella*, the apple maggot fly. *Heredity* **77,** 83–99.

Besansky, N. J., Krzywinski, J., Lehmann, T., Simard, F., Kern, M., Mukabayire, D., Fontenille, D., Toure, Y. & Sagnon, N.F. (2003) Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: Evidence from multilocus DNA sequence variation. *Proc. Natl. Acad. Sci. USA* **100,** 10818–10823.

Brower, A. V. Z. (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proc. Natl. Acad. Sci. USA* **91,** 6491–6495.

Brown, K. M., Burk, L. M., Henagan, L. M. & Noor, M. A. F. (2004) A test of the chromosomal rearrangement model of speciation in *Drosophila pseudoobscura*. *Evolution (Lawrence, Kans.)* **58,** 1856–1860.

Bush, G. L. (1966) *The Taxonomy, Cytology, and Evolution of the Genus Rhagoletis in North America* (Mus. Comp. Zool., Cambridge, MA).

Bush, G. L. (1969) Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera, Tephritidae). *Evolution (Lawrence, Kans.)* **23,** 237–251.

Butlin, R. (1998) What do hybrid zones in general, and the *Chorthippus parallelus* zone in particular, tell us about speciation? In *Endless Forms: Species and Speciation*, eds. Howard, D. J. & Berlocher, S. H. (Oxford Univ. Press, New York), pp. 367–389.

Carson, H. L. (1975) The genetics of speciation at the diploid level. *Am. Nat.* **109,** 83–92.

Coyne, J. A. & Orr, H. A. (2004) *Speciation* (Sinauer, Sunderland, MA).

dellaTorre, A., Merzagora, L., Powell, J. R. & Coluzzi, M. (1997) Selective introgression of paracentric inversions between two sibling species of the *Anopheles gambiae* complex. *Genetics* **146,** 239–244.

Dobzhansky, T. (1937) *Genetics and the Origins of Species* (Columbia Univ. Press, New York).

Dobzhansky, T. (1981) *Dobzhansky's Genetics of Natural Populations I–XLIII*, eds. Lewontin, R. C., Moore, J. A., Provine, W. B. & Wallace, B. (Columbia Univ. Press, New York).

Dowling, T. E. & Secor, C. L. (1997) The role of hybridization and introgression in the diversification of animals. *Annu. Rev. Ecol. Syst.* **28,** 593–618.

Feder, J. L. & Filchak, K. E. (1999) It's about time: The evidence for host plant-mediated selection in the apple maggot fly, *Rhagoletis pomonella*, and its implications for fitness trade-offs in phytophagous insects. *Ent. Exp. Appl.* **91,** 211–225.

Feder, J. L., Chilcote, C. A. & Bush, G. L. (1990) The geographic pattern of genetic differentiation between host associated populations of *Rhagoletis pomonella* (Diptera: Tephritidae) in the eastern United States and Canada**.** *Evolution (Lawrence, Kans.)* **44,** 570–594.

Feder, J. L., Hunt, T. A. & Bush, G. L. (1993) The effects of climate, host plant phenology and host fidelity on the genetics of apple and hawthorn infesting races of *Rhagoletis pomonella*. *Entomol. Exp. App.* **69,** 117–135.

Feder, J. L., Berlocher, S. H. Roethele, J. B. Smith, J. J., Perry, W. L., Gavrilovic, V., Filchak, K. E. & Aluja, M. (2003a) Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proc. Natl. Acad. Sci. USA* **100,** 10314–10319.

Feder, J. L., Roethele, J. B, Filchak, K., Niedbalski, J. & Romero-Severson, J. (2003b) Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics* **163,** 939–953.

Filchak, K. E., Roethele, J. B. & Feder, J. L. (2000) Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*. *Nature* **407,** 739–742.

Grant, P. R. & Grant, B. R. (2002) Unpredictable evolution in a 30-Year study of Darwin's finches. *Science* **296,** 707–711.

Grant, V. (1971) *Plant Speciation* (Columbia Univ. Press, New York).

Hewitt, G. M. (1989) The subdivision of species by hybrid zones. In *Speciation and Its Consequences*, eds. Otte D. & Endler, J.A. (Sinauer, Sunderland, MA), pp. 85–110.

Hey, J. (2001) *Genes, Categories, and Species: The Evolutionary and Cognitive Causes of the Species Problem* (Oxford Univ. Press, Oxford).

Hudson, R. R. & Kaplan, N. L. (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111,** 147–164.

Jiang, C.-X., Chee, P. W., Draye, X., Morrell, P. L., Smith, C. W. & Paterson, A. H. (2000) Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). *Evolution (Lawrence, Kans.)* **54,** 798–814.

Linn, C., Jr., Feder, J. L., Nojima, S., Dambroski, H. R., Berlocher, S. H. & Roelofs, W. (2003) Fruit odor discrimination and sympatric host race formation in *Rhagoletis*. *Proc. Natl. Acad. Sci. USA* **100,** 11490–11493.

Linn, C., Jr., Dambroski, H. R., Feder, J. L., Berlocher, S. H., Nojima, S. & Roelofs, W. (2004) Host specific mating, sympatric speciation and reduced response of hybrid *Rhagoletis* flies to fruit volatiles. *Proc. Natl. Acad. Sci. USA* **101,** 17753–17758.

Machado, C. A. & Hey, J. (2003) The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proc. R. Soc. London Ser. B* **270,** 1193–1202.

Machado, C. A., Kliman, R. M., Markert, J. A. & Hey, J. (2002) Inferring the history of speciation from multilocus DNA sequence data: The case of *Drosophila pseudoobscura* and close relatives. *Mol. Biol. Evol.* **19,** 472–488.

Mallet, J. (2005) Speciation in the 21st century. *Heredity*, in press.

Mayr, E. (1942) *Systematics and the Origins of Species* (Columbia Univ. Press, New York).

Mayr, E. (1963) *Animal Species and Evolution* (Belknap, Cambridge, MA).

Navarro, A. & Barton, N. H. (2003) Accumulating postzygotic isolation genes in parapatry: A new twist on chromosomal speciation. *Evolution (Lawrence, Kans.)* **57,** 447–459.

Noor, M. A. F., Grams, K. L., Bertucci, L. A. & Reiland, J. (2001a) The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. *Evolution (Lawrence, Kans.)* **55,** 512–521.

Noor, M. A. F., Grams, K. L., Bertucci, L. A. & Reiland, J. (2001b) Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci. USA* **98,** 12084–12088.

Posada, D. & Crandall, K. A. (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **14,** 817–818.

Prokopy, R. J., Bennett, E. W. & Bush, G. L. (1971) Mating behavior in *Rhagoletis pomonella* (Diptera: Tephritidae). I. Site of assembly. *Can. Entomol.* **103,** 1405–1409.

Prokopy, R. J., Bennett, E. W. & Bush, G. L. (1972) Mating behavior in *Rhagoletis pomonella*. II. Temporal organization. *Can. Entomol.* **104,** 97–104.

Rieseberg, L. H. (2001) Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16,** 351–358.

Rieseberg, L. H., Whitton, J. & Gardner, K. (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152,** 713–727.

10

On the Origin of Lake Malawi Cichlid Species: A Population Genetic Analysis of Divergence

YONG-JIN WON,*† ARJUN SIVASUNDAR,*‡ YONG WANG,* AND JODY HEY*

The cichlid fishes of Lake Malawi are famously diverse. However, phylogenetic and population genetic studies of their history have been difficult because of the great amount of genetic variation that is shared between species. We apply a recently developed method for fitting the ''isolation with migration'' divergence model to a data set of specially designed compound loci to develop portraits of cichlid species divergence. Outgroup sequences from a cichlid from Lake Tanganyika permit model parameter estimates in units of years and effective population sizes. Estimated speciation times range from 1,000 to 17,000 years for species in the genus *Tropheops*. These exceptionally recent dates suggest that Malawi cichlids as a group experience a very active and dynamic diversification process. Current effective population size estimates range form 2,000 to near 40,000, and to >120,000 for estimates of ancestral population sizes. It appears that very recent speciation and gene flow are among the reasons why it has been difficult to discern the phylogenetic history of Malawi cichlids.

*Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ 08854.

†Present address: Department of Life Sciences, Ewha Womans University, Seoul 120-750, Korea.

‡Present address: Hopkins Marine Station, Stanford University, Pacific Grove, CA 93950.

Abbreviations: STR, short-tandem repeat; IM, isolation with migration.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AY997249–AY997295).

The extraordinary number of species of cichlid fishes (Teleostei: Cichlidae) of the African great lakes Malawi, Tanganyika and Victoria are a classic evolutionary mystery, and biologists have long wondered how so many species could have evolved over short time periods. In the case of Lake Malawi, the estimated geological age of the lake is 4–5 million years, but the lake probably dried out at times, perhaps as recently as 570,000 years ago (Delvaux, 1996).

A complicating factor for phylogenetic and population genetic investigations of the Lake Malawi cichlids is that species tend to share much of their genetic variation, which has been seen with allozymes (Kornfield, 1978; McKaye *et al.*, 1982, 1984), mitochondrial haplotype data (Moran and Kornfield, 1993; Parker and Kornfield, 1997), microsatellite or short-tandem repeat (STR) loci (Kornfield and Parker, 1997), and nuclear DNA sequences (Hey *et al.*, 2004). The fact of shared variation means that neither allelic nor haplotypic data from individual loci (or from a small number of loci) can provide phylogenetic resolution (Kornfield and Parker, 1997; Moran and Kornfield, 1993; Parker and Kornfield, 1997), and in recent years investigators have had to turn to using very large numbers of amplified fragment-length polymorphism markers to estimate phylogenies (Albertson *et al.*, 1999; Allender *et al.*, 2003).

Shared genetic variation also raises important, albeit difficult, population genetic questions. The extensive sharing of genetic variation by closely related cichlid species has traditionally been attributed to the simple persistence of variation that was present in ancestral species (Albertson *et al.*, 1999; Moran and Kornfield, 1993; Parker and Kornfield, 1997). However, shared variation and low levels of divergence between cichlid species have also been interpreted as evidence of ongoing low levels of gene flow (Danley and Kocher, 2001). Direct evidence of interspecies gene flow comes from hybrids and hybrid populations (Smith *et al.*, 2003; Stauffer *et al.*, 1996; Streelman *et al.*, 2004). If cichlid species are diverging in the presence of gene flow, then it is also necessary to consider the role that natural selection plays, either in driving divergence and/or limiting gene flow.

Hey *et al*. (2004) developed the use of compound loci that have a low-mutation rate component and a high-mutation rate component and then analyzed the data by using a recently developed parameter-rich model of population divergence (Fig. 10.1). Here we extend this approach to a larger set of loci and species. In addition, we include dated outgroup sequences that allow us to estimate the actual times and effective population sizes associated with speciation events.

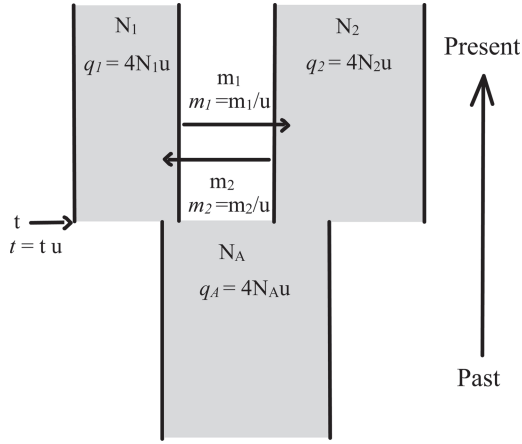


FIGURE 10.1 The IM model is depicted with two parameter sets. The basic demographic parameters are constant effective population sizes (N_1, N_2, and N_A), gene flow rates per gene copy per generation (m_1 and m_2), and the time of population splitting at t generations in the past. The parameters in the second set (in italics) are all scaled by the neutral mutation rate u, and these parameters are actually used in the model fitting.

METHODS

Species and Sample Collection

Species of the genus *Tropheops* are members of the large group of rock-dwelling (mbuna) cichlids (Trewavas, 1984). Ten individuals from each of three species of *Tropheops* were sampled from two locations: Otter Point, on the northwestern part of the Nankumba peninsula in the southern end of the lake, and Harbor Island, a small island in the mouth of Monkey Bay on the eastern side of the Nankumba peninsula. The two sites are separated by ≈18 km along the shore. The protocols for sample collection are given by Hey *et al.* (2004).

HapSTR Loci

We developed a set of compound loci, each including a STR or microsatellite and the flanking unique sequence that may include multiple polymorphic sites. Inspired by a similar approach in which loci that have a SNP and an adjacent STR are referred to as SNPSTRs (Mountain *et al.*, 2002), we call the these loci *Hap*STRs (Hey *et al.*, 2004). A given *Hap*STR haplotype contains a sequence and the number of repeats in the linked STR allele. Six *Hap*STR loci were developed, and haplotypes were deter-

mined from both chromosomes of each of the sampled individuals by following the methods of Hey *et al*. (2004). Oligonucleotide primer information is given in Tables 3 and 4, which are published as supporting information on the PNAS web site.

Outgroup Sequencing

To estimate the times of species formation, an outgroup with a known common ancestry time is required. Phylogenetic studies suggest that the Malawi and Victorian cichlids derive from the tribe Haplochromini, which arose in Lake Tanganyika (Salzburger *et al*., 2002a). Unlike the relatively shallow radiations of Lakes Malawi and Victoria, the much older Lake Tanganyika has cichlids of 12 tribes (including eight endemic tribes) (Nishida, 1991; Poll, 1986). Given that Lake Tanganyika has a long history of cichlid diversification and that it is the likely source for the radiations in Lakes Malawi and Tanganyika, we elected to use as an outgroup a representative of tribe Eretmodini, the oldest monophyletic, endemic clade of Tanganyikan cichlids (Kocher *et al*., 1995; Salzburger *et al*., 2002b). The oldest parts of Lake Tanganyika have been estimated to be 9–12 million years old (Cohen *et al*., 1993), and we used a date of 7 million years for the common ancestor of the outgroup and *Tropheops* (Salzburger and Meyer, 2004). A representative of *Eretmodus cyanostictus* was obtained from a Lake Tanganyika fish importer, and DNA sequences were obtained from those regions corresponding to the sequence portions of the *Hap*STR loci. A nested PCR method was used to increase accuracy in PCR amplification. The product of the initial round of PCR, which includes the flanking sequence and the STR segment, was used as template for a secondary PCR that amplified only the flanking DNA sequences. The primer pairs for these nested PCR are given in Tables 3 and 4. Final PCR products were used directly as templates for bidirectional sequencing on a Li-Cor (Lincoln, NE) 4200 sequencer using dye-labeled M13 forward and reverse primers. PCR amplification and DNA sequencing for one locus, PZMSAT2, were not successful for the outgroup.

Divergence Model and Parameter Estimation

The data for six loci, for pairs of species and populations, were analyzed by using a computer program that estimates the posterior probability density for parameters in the ''isolation with migration'' (IM) model. This version of the IM model has six demographic parameters (Nielsen and Wakeley, 2001), each scaled by the overall neutral mutation rate (Fig. 10.1). For the current multilocus study, we implemented migration in a new way, with each locus having its own pair of migration rate param-

eters. The new migration method is a straightforward extension of the original procedure.

For each of the six *Hap*STR loci, there are two mutation-rate scalar parameters (one for the sequence portion and one for the STR portion) included in the model (Hey and Nielsen, 2004). The procedure, as implemented in a computer program, is to run a Markov chain simulation with appropriate Metropolis–Hastings update criteria as specified under the model (Hey and Nielsen, 2004; Nielsen and Wakeley, 2001). Each simulation is based on a user-specified uniform prior distribution of parameter ranges. The settings for the prior distributions were empirically obtained after preliminary runs using higher upper bounds on parameter distributions (Won and Hey, 2005). Ideally, the posterior distribution that is obtained should fall completely within the prior distribution. However, for some parameters, it was often found that posterior distributions included a peak at a low or intermediate location in the distribution with a flat, or nearly flat, tail over an extended range of higher parameter values. In these cases, it was necessary to choose a prior upper bound that did not include the flat tail of the distribution.

Over the course of the run, for each of the model parameters, a marginal density was recorded as a histogram with 1,000 equally sized bins. The distributions were smoothed by averaging over adjacent points, and the peaks of the resulting distributions were taken as estimates of the parameters (Nielsen and Wakeley, 2001). Depending on the data, the duration of the simulation needed to ensure that the marginal density estimates are based on a good sample of effectively independent values can be very long. In the case of the six-locus *Hap*STR data sets, the autocorrelation of parameter values over the course of individual runs of the computer program proved to be quite high, indicating that it would be difficult to achieve large samples of effectively independent observations. To improve mixing of the Markov chain and shorten the time needed for simulation, runs were done by using multiple Markov chains under the Metropolis coupling protocol (Geyer, 1991; Hey and Nielsen, 2004; Won and Hey, 2005). As many as 110 coupled Markov chains were used for some species pairs, with each multichain simulation lasting several days or weeks. Runs were monitored by using estimates of the effective sample size based on the measured autocorrelation of parameter values over the course of the run. Each analysis was repeated three or more times to ensure that similar density estimates were obtained.

Parameter Scale Conversion

Estimates of the mutation rates for those loci for which outgroup sequence is available can be used to convert model parameter estimates,

which are scaled by mutation rate, to more interpretable scales. Each *Hap*STR locus requires two mutation rate scalar parameters, one for the STR and one for the flanking sequence, for which the probability density is estimated along with the demographic parameters under the IM model (Hey *et al.*, 2004; Hey and Nielsen, 2004). The procedure for converting parameter estimates, for the case when outgroup data are available for only a subset of the loci that are in the IM analysis, requires two values: a quantity, X, which is the geometric mean of the mutation rate scalar estimates that are generated by the IM analysis, for just those loci for which mutation rate estimates are available based on the outgroup; and a quantity, U, which is the geometric mean of the mutation rate per year, for those same loci based on the outgroup divergence and known time of common ancestry. With these values, the scale of the estimate of the divergence time parameter, which is in units of mutations (i.e., t = tu), can be converted to one of years by $\hat{t} = tX/U$. If G is an estimate of the number of generations per year, then an estimate of the number of generations since speciation is $\hat{t}/G$. Similarly, if $\hat{\theta}$ is an estimate of 4Nu for one of the populations, then an estimate of the effective population size, N, can be obtained as $\hat{N}/\hat{\theta}X/(4UG)$. In captivity, mbuna cichlids can reach reproductive age in less than a year; however, the generation time of mbuna cichlids in the wild is not known. Different authors have used times of 1, 2, or 3 years (Parker and Kornfield, 1997; Streelman *et al.*, 2004; Van Oppen *et al.*, 1997b), and here we have used 2 years.

For most population genetic purposes, the relevant scale for migration is the population migration rate, 2Nm, or the effective number of migrants per generation. To show the probability density of the migration parameters on this scale, the migration rate parameters were rescaled by multiplying by the corresponding estimate of 2Nu from the same analysis (e.g., for the $m1$ (= m_1/u) parameter, multiply the estimated values by values by $2\hat{N}_1u = \hat{\theta}_1/2$).

RESULTS

Outgroup Divergence

A summary of polymorphism among the *Tropheops* species and between these and the outgroup *E. cyanostictus* is given in Table 10.1. Each locus revealed a large number of STR alleles and at least one polymorphic site in the flanking sequence. At each *Hap*STR locus (including sequences together with the linked STR alleles) we observed a large amount of haplotype sharing among species, as expected (Hey *et al.*, 2004). Assuming that the time of common ancestry with the outgroup was 7 million years ago, the mean substitution rate among these loci (not weighted by se-

TABLE 10.1 Polymorphism Summaries

Locus	Flanking Sequence	No. of STR Alleles	No. of SNPs	No. of Haplotypes	Divergence	Source
UNH001	278	29	1	2	3.4	Kellogg *et al.* (1995)
U66815	192	31	2	3	1.5	Booton *et al.* (1996)
U66814	634	19	2	3	9	Booton *et al.* (1996)
U14396	453	26	4	6	14.5	Parker and Kornfield (1996)
DXTUCA3	417	23	1	2	4.5	U94850[a]
PZMSAT2	648	29	4	4	—	Van Oppen *et al.* (1997a)

NOTE: Shown is the length of the flanking sequence in base pairs, the number of distinct STR alleles observed across three species of *Tropheops*, the number of SNPs observed in the flanking sequence in the entire sample, the number of distinct sequence (not including STRs) haplotypes across the entire sample, and the average number of sequence differences (divergence) between the outgroup, *E. cyanostictus*, and *Tropheops* over the full length of the flanking sequence. —, not determined.
[a]GenBank accession number.

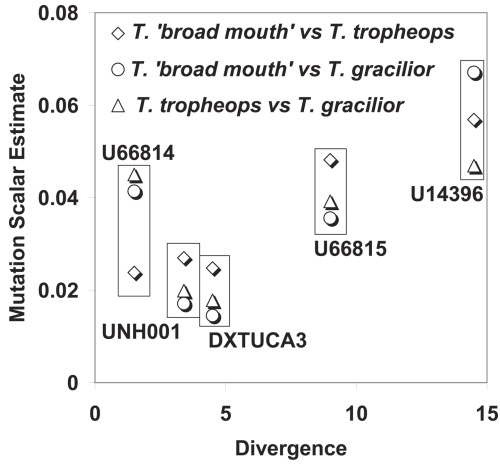


FIGURE 10.2 Mutation rate scalars, estimated in separate analyses for each species pair, are plotted against the average amount of sequence divergence between *Tropheops* and *Eretmodus*. The scalars are the estimated mutation rates relative to other loci (including STRs that are not shown) obtained by fitting the data to the IM model. Points are grouped for each of the five loci for which divergence could be measured.

quence length) is 1.1×10^{-9} substitutions per site per year, roughly one-third of the estimated rates for noncoding nuclear DNA in mammals and birds (Axelsson *et al.*, 2004; Li, 1997).

To check whether levels of polymorphism are consistent with a neutral model, we compared the estimated mutation rate scalars for the sequence portions of the six loci with the amount of divergence observed between *Tropheops* and *Eretmodus*, the outgroup. Although only five loci could be included, there is a general positive relationship between the two independent assessments of the mutation rate (Fig. 10.2) as expected under the neutral model. It is important to note that these scalars have no units and that their magnitude is constrained under the implementation of the model such that the product of all 12 mutation rate scalars (including those for six loci, each with sequenced portions and STR portions) in a given run of the program have a product of one (Hey and Nielsen, 2004). The scalar values are all <1 because they were estimated in runs in which the scalars for the STR regions were also estimated, and these scalar values all have values >1 (i.e., the STR regions are estimated to have high mutation rates, as expected, relative to the flanking sequence).

IM MODEL ANALYSIS

We have sampled six populations (two each of three species); however, our method of analysis can only accommodate pairs of populations. Therefore, the first step is to consider the pairs of populations of each species and to ask how recently they have diverged. Fig. 10.3 shows the posterior probability density estimates of the time of divergence between the two populations for each species pair. For *Tropheops* ''broad mouth,'' the estimated time of splitting is at or near zero, suggesting that the two populations are recently derived from a single population. For *Tropheops tropheops*, the shape of the distribution is fairly flat, with considerable density near zero. In contrast, *Tropheops gracilior* shows a clear peak at ≈2,200 years and an estimated density at zero years that is itself zero. On the basis of these results, we have pooled the two populations of *T. tropheops* and the two populations of *T.* broad mouth and kept the two populations of *T. gracilior* separate for the remainder of the analyses.

Fig. 10.4 shows the marginal probability densities for the population size and migration rate parameters in the contrast between the two populations of *T. gracilior*. The population size parameter scales are numbers of individuals. The scale for migration is different from the other parameters because it is set by an estimate obtained from the analysis of one of those other parameters (see *Methods*). For migration, that scale is in units of

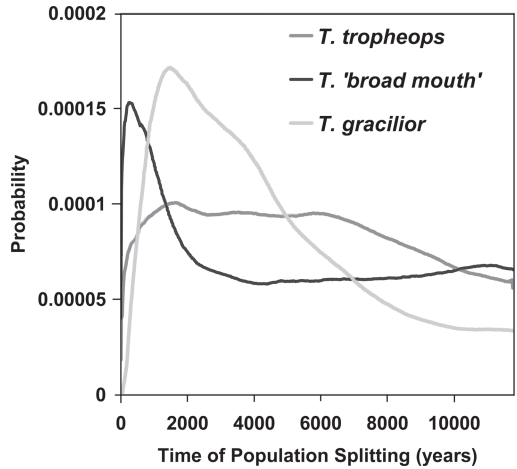


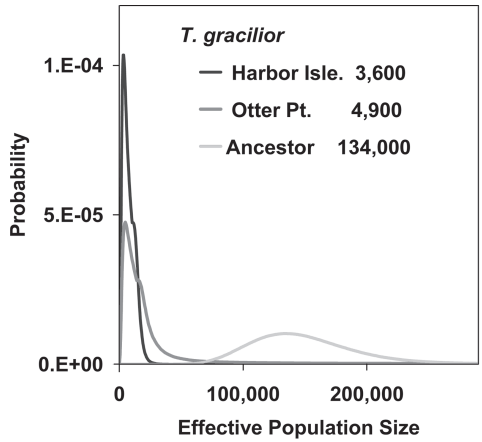
FIGURE 10.3 Probability estimates for the time of population splitting for two populations of each species. The time scale as been converted to years, based on mutation rate scalar estimates and outgroup divergence, as described in *Methods*.

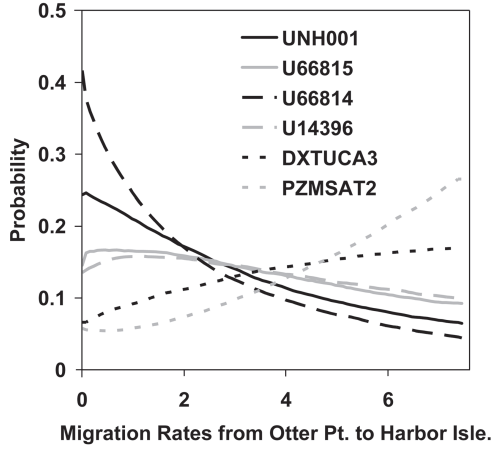
2Nm, the population migration rate, where N is the effective size of the receiving population and m is the probability of migration per gene copy per generation. The population size parameters are the most clearly resolved, with posterior distributions that have a clear single peak and bounds that fall within the prior distribution. It is noteworthy that the ancestral population is estimated to have been much larger than either of the present day populations. The densities for the migration rates vary among loci; however, only one locus has a curve that might be interpreted as clear evidence of nonzero gene flow (Fig. 10.4 *Middle*, PZMSAT2). The densities for the migration rate parameters are fairly flat, indicating that the data contain little information on migration within the framework of the IM model. Because the model assumes a constant rate of gene flow after the population separation, it is expected that migration rates between populations that have recently split, as appears to be the case, will be hard to estimate.

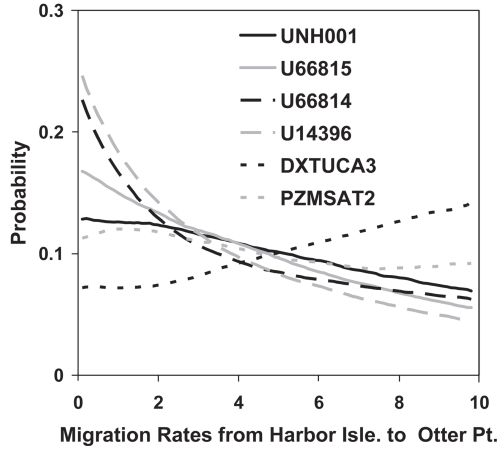
Findings similar to those for the two populations of *T. gracilior* are also found in the other analyses between these populations and the pooled samples of *T. tropheops* and *T.* broad mouth. The parameter estimates for these pairs of populations are given in Table 10.2, and, because each population occurs in multiple contrasts, several recurrent patterns emerge (Delvaux, 1996). The effective population sizes for the two *T. gracilior* populations are smaller, with values ranging from 1,500 to 4,900, than those for *T. tropheops* and *T.* broad mouth, which have values ranging form 15,400 to 19,000 (Kornfield, 1978). The size of the estimated ancestral populations are considerably larger than the current populations, with estimates in the range of 120,900–128,200 (McKaye *et al.*, 1982). The estimates of population splitting time are all very recent and range from 1,000 to 2,300 years.

For the migration parameters, all of the analyses that include *T. gracilior* generated curves that are like those in Fig. 10.4, in which most curves lack a clear peak and loci vary in whether or not they suggest a history of gene flow. Table 10.2 lists those loci that showed a lower probability of zero gene flow relative to the probability at the high end of the distribution. Depending on the contrast, evidence of gene flow was found primarily in the direction of gene flow into the *T. gracilior* population, as opposed to the reverse.

The analyses of *T.* broad mouth and *T. tropheops* yielded a different picture (Fig. 10.5), with an estimated divergence time of 17,700 years and estimated effective population sizes several fold larger (21,300 for *T.* broad mouth and 47,800 for *T. tropheops)* than estimated in the contrasts with *T. gracilior*. The size of the ancestral population in this case (74,000), although larger than for the descendant species, was not as large as the estimates from the contrasts with *T. gracilior*. The migration rate density estimates







were nearly all very flat in this analysis (data not shown), so we do not have a clear picture of how much gene flow may have been occuring in this case.

DISCUSSION

Because of low levels of divergence and widespread shared variation among Malawi cichlids, questions about their phylogenetic history have been difficult. The difficulty has generally been described as a consequence of very recent divergence (Kornfield and Smith, 2000; Parker and Kornfield, 1997; Seehausen *et al.*, 1999). However, in the absence of phylogenetic assessments, it is difficult to know to what degree recent speciation is the cause of shared variation and the lack of phylogenetic resolution. Another possible cause for the lack of phylogenetic resolution in genetic data, in addition to recent speciation, is that genetic variation is shared because of gene exchange. Disentangling the relative contributions of variation shared since ancestry and shared via gene flow is necessary for estimating the time since speciation and for assessing speciation models. Recently, with the use of very large numbers of amplified fragment-length polymorphisms (Albertson *et al.*, 1999; Allender *et al.*, 2003), it has become possible to estimate phylogenetic trees. However, it is difficult to model the substitution process for these markers and therefore it is difficult to use amplified fragment-length polymorphisms to estimate the time since speciation events.

The protocol used here, in which compound loci with highand low-mutation-rate components are analyzed under a parameter-rich model of population divergence, was designed to address questions about cichlid speciation (Hey *et al.*, 2004). In brief, it appears that very recent speciation and gene flow contribute to the shared variation and, therefore, to the difficulty of assessing phylogenetic history.

To answer the first question (How long ago did Malawi cichlid species undergo speciation?), our estimate for *Tropheops* species varies from a low range of 1,000–2,300 years in the case of *T. gracilior* and an estimate of 17,200 years for the divergence of *T.* broad mouth and *T. tropheops*. If such recent dates apply to Malawi cichlids in general, then they suggest that

FIGURE 10.4 Results for two populations of *T. gracilior*. Probability density estimates are shown for effective population sizes (*Top*) and migration rates from Otter Point to Harbor Isle (*Middle*) and from Harbor Isle to Otter Point (*Bottom*). The estimates for effective population size are given in the key in *Top*. The scales for migration rates are set by using these estimates of the effective population size (see *Methods*).

TABLE 10.2 Model Parameter Estimates in Pairs of Populations

Population 1	Population 2	N_1	N_2	N_A	t	$2N_1m_1>0$	$2N_2m_2>0$
T. gracilior (HI)	*T. gracilior* (OP)	3,600	4,900	134,000	2,300	PZMSAT2	—
T. tropheops	*T. gracilior* (HI)	18,800	2,000	120,900	1,600	—	U14396, PZMSAT2, DXTUCA3
T. tropheops	*T. gracilior* (OP)	19,100	2,500	121,800	1,100	—	U66815, U14396, PZMSAT2
T. broad mouth	*T. gracilior* (HI)	15,400	1,500	128,200	1,000	—	U14396, DXTUCA3, PZMSAT2,
T. broad mouth	*T. gracilior* (OP)	21,500	3,200	122,100	1,800	—	—
T. broad mouth	*T. tropheops*	21,300	47,800	74,000	17,700	—	—

NOTE: Shown are the species population pairs, with the location of the *T. gracilior* population shown in parentheses: HI, Harbor Isle; OP, Otter Point. The loci that showed a migration parameter curve indicating a low probability of zero migration and a high probability of non-zero migration are listed under $2N_1m_1 > 0$ and $2N_2m_2 > 0$. —, No loci suggested a high probability of non-zero migration.

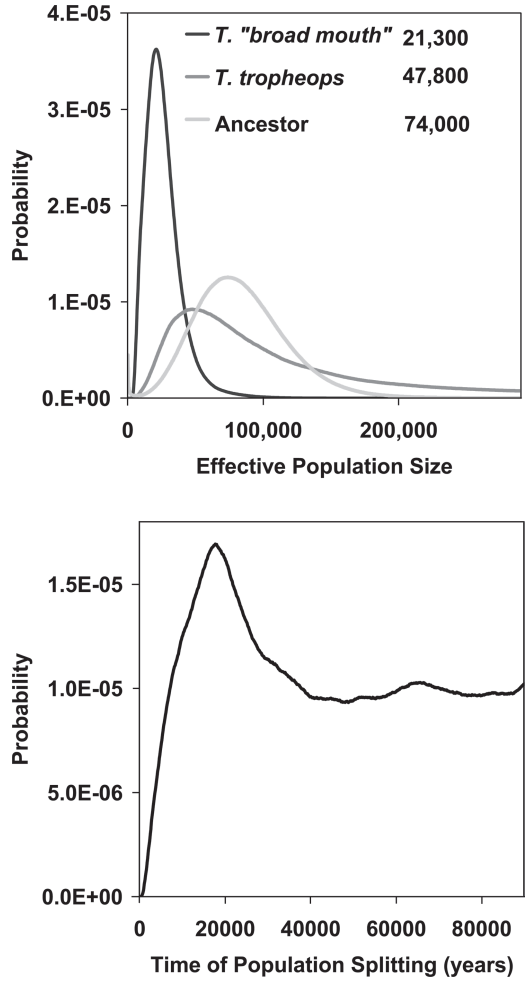


FIGURE 10.5 Results for *T.* broad mouth and *T. tropheops*. Probability density estimates are shown for effective population sizes (*Upper*) and estimated time of divergence (*Lower*). The estimates for effective population size are given in the key.

the Malawi cichlid flock is extraordinarily evolutionarily dynamic. To get a feel for the implications, consider that if we take 10,000 years as the typical time between speciation events and apply it to all of the ≈230 formally described species of mbuna, then a new mbuna species arises every 43 years (i.e., 10,000 years/230 species). If the radiation of mbuna began 1,000,000 years ago with such high rates of speciation, then there may have been >23,000 different species over the years, assuming a steady

state process of speciation and extinction. This estimate does not count the >270 other described, non-mbuna, species of Malawi cichlids. These calculations are simplistic and do not take into account the very high level of species uncertainty that is associated with various aspects of research on Malawi cichlids; however, they do suggest an exceptional rate of diversification.

Could these estimated dates be quite wrong, such that the true values actually lie outside the range of the peaks of the estimated probability densities? Because of the many parameters and the complexity of the method, this question is difficult address. There are at least two general sets of assumptions of which to be aware. First, we have imposed a model of population splitting that assumes that an ancestral population that persisted long before the moment of splitting. Complex population dynamics within the ancestral population are not accounted for and neither is gene exchange with other populations not included within a particular analysis. It is likely that the very large estimated sizes of the ancestral populations (Table 10.2) are the result of gene exchange between ancestral populations. This kind of gene exchange cannot be estimated by the method, although it will elevate the amount of variation in ancestral populations and lead to inflated effective population size estimates. However, it is not clear that such processes would lead to biased estimates of the divergence time of populations.

When considering the appropriateness of the model, the analyses of *T.* broad mouth and *T. tropheops* raised some interesting problems. All of the individual population pair analyses that involved *T. gracilior* required long runs of the computer program and large numbers of coupled Markov chains; however, in the case of *T.* broad mouth and *T. tropheops* it was necessary to run 110 chains, each distinguished by very slight differences in heating level (Geyer, 1991). This high number of chains were required to break up the very strong autocorrelation of parameter values (primarily t) that arose in the course of the simulation and to obtain enough effectively independent measurements so as to have some confidence in the final distribution. Furthermore, the final distribution for t, although showing a peak at 17,700 years, is very broad and appears to plateau to the right of the peak. It is possible that part of the reason for this broad distribution is that each species actually included samples from two separate populations, although the separation of these individual pairs of populations appeared to have been quite recent, as suggested by Fig. 10.3.

A second possible source of bias that needs to be considered are the STR loci. The IM analysis assumes a stepwise mutation model for these loci, and it is possible that a failure of this assumption may bias the results. Care was taken to use loci with STR regions that have simple repeats. In the IM analyses the estimated mutation rate scalars for the STR

portions of the compound loci were typically ≈300,000 times higher than the mutation rate per site in the flanking sequence. If the mutation rate in the flanking sequence is 2×10^{-9} per site per generation (i.e., the per site per year estimate based on the outgroup divergence × 2 years per generation), then our estimate of the STR mutation rate would be 300,000 times higher than this, or $\approx 6 \times 10^{-4}$ per generation, which is a fairly typical rate (Goldstein and Schlötterer, 1999). The IM analyses suggest that populations and species have been exchanging genes, particularly from *T. tropheops* and *T.* broad mouth into populations of *T. gracilior*. However the estimated densities of migration parameters are mostly flat, and there is little resolution of migration rates in most cases. In recent years several authors have argued that some gene flow between species is probably occurring (Markert *et al.*, 2001; Seehausen, 2004; Smith and Kornfield, 2002; Smith *et al.*, 2003), and in particular Kocher and colleagues (Danley and Kocher, 2001; Danley *et al.*, 2000) have argued in support of the ''divergence with gene flow'' model of speciation (Endler, 1973; Rice and Hostert, 1993) for Malawi cichlids. In such models, two populations may diverge in parapatry or sympatry because of selective forces, even in the presence of gene flow. These models differ fundamentally from strictly allopatric models of speciation in that they directly entail a role for divergent natural selection as a cause of species diversity (Rice and Hostert, 1993).

Although the use of dated outgroup sequences and a parameter-rich model of divergence allows us to address difficult questions about the divergence of Malawi cichlids, there are clear limitations to these interpretations. Necessarily, the divergence process has been viewed through the lens of the IM model, and it is not yet clear how the picture would change if we were able to consider more than two populations simultaneously or could better assess the impact of assuming the stepwise mutation model for the STR portions of loci. The consistently very large estimates for ancestral population sizes do suggest that our samples contain variation that arose not just in single ancestral populations but in a wider array of partly intermingled populations. This interpretation is consistent with the evidence for recent gene exchange among populations and species.

ACKNOWLEDGMENTS

We thank Jeff Markert and Matt Arnegard for help collecting samples; Richard Zatha, Daniel Phiri, and David Mwafulirwe for providing valuable field assistance; and Aggrey Ambali (Molecular Biology and Ecology Research Unit, University of Malawi, Zomba, Malawi) for helping to facilitate the collecting trip and for the use of the Molecular Biology and

Ecology Research Unit field station. This work was supported by National Science Foundation Grant DEB 0129724.

REFERENCES

Albertson, R. C., Markert, J. A., Danley, P. D. & Kocher, T. D. (1999) Phylogeny of a rapidly evolving clade: The cichlid fishes of Lake Malawi, East Africa. *Proc. Natl. Acad. Sci. USA* **96,** 5107–5110.

Allender, C. J., Seehausen, O., Knight, M. E., Turner, G. F. & Maclean, N. (2003) Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in nuptial coloration. *Proc. Natl. Acad. Sci. USA* **100,** 14074–14079.

Axelsson, E., Smith, N. G., Sundstrom, H., Berlin, S. & Ellegren, H. (2004) Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. *Mol. Biol. Evol.* **21,** 1538–1547.

Booton, G. C., Kaufman, L., Chandler, M. & Fuerst, P. A. (1996) Use of DNA microsatellite loci to identify populations and species of Lake Victoria haplochromine cichlids. In *Symposium Proceedings, International Congress on the Biology of Fishes*, eds. Donaldson, E. M. & MacKinlay, D. D. (Am. Fisheries Soc., Physiol. Sect., Vancouver, CA), Vol. 9, pp. 105–113.

Cohen, A. S., Soreghan, M. J. & Scholz, C. A. (1993) Estimating the age of formation of lakes: An example from Lake Tanganyika. *Geology* **21,** 511–514.

Danley, P. D. & Kocher, T. D. (2001) *Mol. Ecol.* **10,** 1075–1086.

Danley, P., Markert, J., Arnegard, M. & Kocher, T. (2000) Divergence with gene flow in the rock-dwelling cichlids of Lake Malawi. *Evolution* **54,** 1725–1737.

Delvaux, D. (1996) Age of Lake Malawi (Nyasa) and water level fluctuations. *Mus. R. Afr. Cent. Tervuren, Belg., Rapp. Annu. Dept. Geol. Mineral.* **1995–1996,** 99–108.

Endler, J. A. (1973) Gene flow and population differentiation. *Science* **179,** 243–250.

Geyer, C. J. (1991) Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics, Proceedings of the 23rd Symposium on the Interface*, ed. Keramidas, E. M. (Interface Found. N. Am., Seattle, WA), pp. 156–163.

Goldstein, D. B. & Schlötterer, C. (1999) Microsatellites: Evolution and applications. *Microsatellites: Evolution and Applications* (Oxford Univ. Press, Oxford, U.K.).

Hey, J. & Nielsen, R. (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167,** 747–760.

Hey, J., Won, Y.-J., Sivasundar, A., Nielsen, R. & Markert, J. A. (2004) Using nuclear haplotypes with microsatellites to study gene flow between recently separated Cichlid species. *Mol. Ecol.* **13,** 909–919.

Kellogg, K. A., Markert, J. A., Jr., Stauffer, J. & Kocher, T. D. (1995) Microsatellite variation demonstrates multiple paternity in lekking cichlid fishes from Lake Malawi, Africa. *Proc. R. Soc. London Ser. B* **260,** 79–84.

Kocher, T. D., Conroy, J. A., McKaye, K. R., Stauffer, J. R. & Lockwood, S. F. (1995) Evolution of NADH dehydrogenase subunit 2 in East African cichlid fish. *Mol. Phylogenet. Evol.* **4,** 420–432.

Kornfield, I. (1978) Evidence for rapid speciation in African cichlid fishes. *Experientia* **34,** 335–336.

Kornfield, I. & Parker, A. (1997) Molecular systematics of a rapidly evolving species flock: The mbuna of Lake Malawi and the search for phylogenetic signal. In *Molecular Systematics of Fishes*, eds. Kocher, T. D. & Stepien, C. A. (Academic, New York), pp. 25–37.

Kornfield, I. & Smith, P. F. (2000) African cichlid fishes: Model systems for evolutionary biology. *Annu. Rev. Ecol. Syst.* **31,** 163–196.

Li, W. H. (1997) Molecular evolution. *Molecular Evolution* (Sinauer, Sunderland, MA).

Markert, J. A., Danley, P. D. & Arnegard, M. E. (2001) New markers for new species: Microsatellite loci and the East African cichlids. *Trends Ecol. Evol.* **16,** 100–109.

McKaye, E. R., Kocher, T., Reinthal, P. & Kornfield, I. (1982) A sympatric species complex of Petrotilapia Trewavas from Lake Malawi analysed by enzyme electrophoresis (Pisces, Cichlidae). *Zool. J. Linnean Soc.* **76,** 91–96.

McKaye, K. R., Kocher, T., Reinthal, P., Harrison, R. & Kornfield, I. (1984) Genetic evidence of allopatric and sympatric differentiation among color morphs of a Lake Malawi cichlid fish. *Evolution* **38,** 215–219.

Moran, P. & Kornfield, I. (1993) Retention of an ancestral polymorphism in the *Mbuna* species flock (Teleostei: Cichlidae) of Lake Malawi. *Mol. Biol. Evol.* **10,** 1015–1029.

Mountain, J. L., Knight, A., Jobin, M., Gignoux, C., Miller, A., Lin, A. A. & Underhill, P. A. (2002) SNPSTRs: Empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Res.* **12,** 1766–1772.

Nielsen, R. & Wakeley, J. (2001) Distinguishing migration from isolation. A Markov chain Monte Carlo approach. *Genetics* **158,** 885–896.

Nishida, M. (1991) Lake Tanganyika as an evolutionary reservoir of old lineages of East African cichlid fishes: Inferences from allozyme data. *Experientia* **47,** 974–979.

Parker, A. & Kornfield, I. (1996) Polygynandry in *Pseudotropheus zebra*, a cichlid fish from Lake Malawi. *Environ. Biol. Fishes* **47,** 345–352.

Parker, A. & Kornfield, I. (1997) Evolution of the mitochondrial DNA control region in the mbuna (Cichlidae) species flock of Lake Malawi. *J. Mol. Evol.* **45,** 70–83.

Poll, M. (1986) Classification des cichlidae du lac Tanganika tribus, genres et especes. *Mem. Cl. Sci., Acad. R. Belg., Collect. (Ser. 2)* **45,** 5–163.

Rice, W. R. & Hostert, E. F. (1993) Laboratory experiments on speciation: What have we learned in 40 years. *Evolution* **47,** 1637–1653.

Salzburger, W. & Meyer, A. (2004) The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. *Naturwissenschaften* **91,** 227–290.

Salzburger, W., Baric, S. & Sturmbauer, C. (2002a) Speciation via introgressive hybridization in East African cichlids? *Mol. Ecol.* **11,** 619–625.

Salzburger, W., Meyer, A., Baric, S., Verheyen, E. & Sturmbauer, C. (2002b) Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas. *Syst. Biol.* **51,** 113–135.

Seehausen, O. (2004) Hybridization and adaptive radiation. *Trends Ecol. Evol.* **19,** 198–207.

Seehausen, O., Mayhew, P. J. & Van Alphen, J. J. M. (1999) Evolution of colour patterns in East African cichlid fish. *J. Evol. Biol.* **12,** 514–534.

Smith, P. F. & Kornfield, I. (2002) Phylogeography of Lake Malawi cichlids of the genus *Pseudotropheus*: Significance of allopatric colour variation. *Proc. R. Soc. London Ser. B* **269,** 2495–2502.

Smith, P. F., Konings, A. & Kornfield, I. (2003) Hybrid origin of a cichlid population in Lake Malawi: Implications for genetic variation and species diversity. *Mol. Ecol.* **12,** 2497–2504.

Stauffer, J. R., Bowers, N. J., Kocher, T. D. & McKaye, K. R. (1996) Evidence of hybridization between *Cynotilapia afra* and *Pseudotropheus zebra* (Teleostei: Cichlidae) following an intralacustrine translocation in Lake Malawi. *Copeia* **1996,** 203–208.

Streelman, J. T., Gmyrek, S. L., Kidd, M. R., Kidd, C., Robinson, R. L., Hert, E., Ambali, A. J. & Kocher, T. D. (2004) Hybridization and contemporary evolution in an introduced cichlid fish from Lake Malawi National Park. *Mol. Ecol.* **13,** 2471–2479.

Trewavas, E. (1984) Nouvel examen des genres et sous-genres du complexe Pseudotropheus-Melanochromis du lac Malawi (Pisces, Perciformes, Cichlidae). *Rev. Fr. Aquariol. Herpetol.* **10,** 97–106.

Van Oppen, M. J. H., Rico, C., Deutsch, J. C., Turner, G. F. & Hewitt, G. M. (1997a) Isolation and characterization of microsatellite loci in the cichlid fish *Pseudotropheus zebra*. *Mol. Ecol.* **6,** 387–388.

Van Oppen, M. J. H., Turner, G. F., Rico, C., Deutsch, J. C., Ibrahim, K. M., Robinson, R. L. & Hewitt, G. M. (1997b) Unusually fine-scale genetic structuring found in rapidly speciating Malawi cichlid fishes. *Proc. R. Soc. London Ser. B* **264,** 1803–1812.

Won, Y. J. & Hey, J. (2005) Divergence population genetics of chimpanzees. *Mol. Biol. Evol.* **22,** 297–307.

Part III

THE NATURE OF SPECIES AND THE MEANING OF "SPECIES"

When Mayr outlined several species concepts, including the biological species concept in his 1942 book, he started a new era in species-problem debate. From that point on, Mayr was the major figure in both the biological and philosophical components of the debate (Mayr, 1957, 1987). In this volume, we have three papers that address, from widely different perspectives, the very nature of species. The many biologists who, like Mayr, take a primarily zoological perspective, will appreciate the case studies presented by Anne Yoder *et al.* in "A Multidimensional Approach for Detecting Species Patterns in Malagasy Vertebrates" (Chapter 11) for vertebrate species complexes that are endemic to Madagascar. The authors describe, with examples, a protocol that begins with field collections and existing taxonomic resources and proceed to devise hypotheses of species boundaries and priorities for additional collecting and experimental work.

But, unlike animals, and indeed most eukaryotes, prokaryotes have always presented special species-related challenges because of the absence of regular gene exchange. Yet is it possible that lateral gene transfer (LGT) that does occur among bacteria, often across wide taxonomic chasms, can provide perspective on the species question in bacteria? This topic is addressed by Howard Ochman, Emmanuelle Lerat, and Vincent Daubin in "Examining Bacterial Species under the Specter of Gene Transfer and Exchange" (Chapter 12), who take a whole-genome approach to ask about the historical and phylogenetic distribution of LGT. They find that although LGT generally can obscure older phylogenetic histories, the

subset of LGT that leads to homologous recombination is largely limited to closely related bacteria. This finding supports a view of bacterial species that resembles, in some respects, the biological species concept.

On the matter of the multiplicity of species concepts, Kevin de Queiroz, in "Ernst Mayr and the Modern Concept of Species" (Chapter 13), has contributed an article that directly targets one of the main sources of confusion that arises in species concept debate. That confusion lies between species criteria, as articulated in various species concepts, which are actually contingent properties of species, and the necessary properties of species as they are understood in the general sense of being evolutionary lineages. Biologists who disagree about which contingent properties of species are the most useful for identification and classification should be able to find common ground by recognizing the contingent, as opposed to necessary, aspect of the features they prefer to study.

REFERENCES

Mayr, E., ed. (1957) *The Species Problem* (Am. Assoc. Adv. Sci., Washington, DC).

Mayr, E. (1987) The ontological status of species: Scientific progress and philosophical terminology. *Biol. Philos.* **2,** 145–166.

11

A Multidimensional Approach for Detecting Species Patterns in Malagasy Vertebrates

ANNE D. YODER,* LINK E. OLSON,†¶
CAROL HANLEY,* KELLIE L. HECKMAN,*
RODIN RASOLOARISON,‡ AMY L. RUSSELL,*
JULIE RANIVO,‡§ VOAHANGY SOARIMALALA,‡§
K. PRAVEEN KARANTH,* ACHILLE P. RASELIMANANA,§
AND STEVEN M. GOODMAN‡§¶

The biodiversity of Madagascar is extraordinarily distinctive, diverse, and endangered. It is therefore urgent that steps be taken to document, describe, interpret, and protect this exceptional biota. As a collaborative group of field and laboratory biologists, we employ a suite of methodological and analytical tools to investigate the vertebrate portion of Madagascar's fauna. Given that species are the fundamental unit of evolution, where micro- and macroevolutionary forces converge to generate biological diversity, a thorough understanding of species distribution and abundance is critical for understanding the evolutionary, ecological, and biogeographic forces that have shaped Malagasy vertebrate diversity. We illustrate the means by which we apply Mayr's "three basic tasks" of the systematist [Mayr, E. (1942) *Systematics and the Origin of Species from the Viewpoint of a Zoologist*

*Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208105, New Haven, CT 06520; †University of Alaska Museum of the North, 907 Yukon Drive, Fairbanks, AK 99775; ‡Département de Biologie Animale, Université d'Antananarivo, BP 906, Antananarivo (101), Madagascar; §WWF Madagascar, BP 738, Antananarivo (101), Madagascar; and ¶Department of Zoology, Field Museum of Natural History, 1400 South Lake Shore Drive, Chicago, IL 60605.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. DQ003345–DQ003479, DQ004403–DQ004461, and DQ005718–DQ005850).

(Harvard Univ. Press, Cambridge, MA)] to identify, classify, and study the organisms that together constitute Madagascar's vertebrate community. Using field inventory methods, specimen-based studies, and morphological and molecular analyses, we formulate hypotheses of species identity that then serve as the foundation for subsequent studies of biology and history. Our experience, as well as that of other investigators, has shown that much of the vertebrate species diversity in Madagascar is ''cryptic'' for both biological and practical reasons. Beyond issues of cryptic biological diversity, the resolution of species identity in Madagascar has been hampered because of a lack of vouchered comparative material at the population level. Through our activities, we are attempting to remedy these limitations while simultaneously enhancing research capacity in Madagascar.

The actual demarcation of species taxa uses morphological, geographical, ecological, behavioral, and molecular information to infer the rank of isolated populations.

Ernst Mayr (1995, p. 276)

Biologists disagree, often vehemently, over the question of what constitutes a species. Virtually all agree, however, that species are a fundamental unit of evolution where micro- and macroevolutionary forces converge to generate biological diversity. Thus, the theoretical and practical issues relating to species identification are essential for the purposes of documenting, describing, and preserving biodiversity. Mayr's fundamental contribution, with the formulation of the biological species concept (BSC), was to express the question as a biological rather than a typological problem: ''The most important aspect of the biological species definition is that it uses no artificial criteria, but decides each case on the basis of whether certain organisms behave as if they were conspecific or not'' (Mayr, 1942, pp. 119–120). He makes clear the importance that morphological and genetic information serve as clues to biological distinctiveness, especially in those cases where reproductive isolation cannot be determined for reasons of geographic separation. Thus, although the BSC by no means provides a universally applicable recipe whereby the working biologist can diagnose a species, and thus define its evolutionary significance, the concept provides a practicable toolkit for embarking upon the enterprise.

As a collaborative group of field and lab biologists, we are motivated by an interest in documenting, describing, understanding, and preserving the endangered vertebrate biota of Madagascar. The exceptional floral

and faunal diversity of this island are well known (Goodman and Benstead, 2003). Madagascar lies ≈300 miles to the east of Africa at the narrowest point of the Mozambique Channel, where it has been isolated from the African continent for nearly 160 million years and from all other significant landmasses for the past 88 million years (Coffin and Rabinowitz, 1992; Storey *et al.*, 1995). Its status as one of the world's top 12 ''megadiversity'' countries is without question due to the remarkable levels of taxonomic endemism found there (Myers *et al.*, 2000). For example, 95% of the reptile species, 99% of amphibian species, and 100% of the island's primate species occur nowhere else on Earth. Certain faunas are either poorly represented (e.g., only four orders of terrestrial eutherians are currently represented) or are completely absent (e.g., there are no salamanders, vipers, or varanid lizards), whereas other groups show unrivaled diversity (e.g., chameleons). Madagascar thus generates intense interest among evolutionary biologists who wish to understand the extent to which geographic and environmental constraints influence organismal evolution. The investigation of Madagascar's biodiversity has the potential to offer insight into key concepts of ecosystem formation and function, biogeographic mechanisms such as vicariance and dispersal, and the sustainability of biodiversity (de Wit, 2003).

Fundamental to the investigation of these key concepts is basic knowledge concerning the presence, distribution, and diversity of Madagascar's remarkable biota. In turn, this basic knowledge depends upon the identification and geographic delineation of biological species. In our efforts to understand the evolutionary forces that govern the distribution and abundance of Malagasy vertebrates, we apply a suite of empirical and analytical tools that permit us first to formulate hypotheses of species distinction, and then progress by means of genetic and morphological analysis to questions of the geographic and temporal context of the species' history. Our methods closely mirror what Mayr describes as the three basic tasks of the systematist: (*i*) identification (analytical stage), (*ii*) classification (synthetic stage), and (*iii*) the study of species formation and of factors of evolution (Mayr, 1942, pp. 9–10).

New vertebrate species have been described from Madagascar at a vigorous pace over the past few years (e.g., see Andreone and Greer, 2002; Goodman and Cardiff, 2004; Goodman and Soarimalala, 2004a; Goodman *et al.*, 1997; Jenkins and Goodman, 1996; Olson *et al.*, 2004; Vences and Glaw, 2004; Vences *et al.*, 2002), and the rate of discovery continues unabated. The distinct majority of these discoveries have been based on fieldwork and the collection of new material, rather than on reassessment of specimens already held in museum collections, although this latter material is paramount for points of taxonomic reference. Virtually any surveyed region of Madagascar with remaining natural habitat

has been found to harbor new species of vertebrates, even for well studied groups such as mammals. Moreover, there are considerable areas of Madagascar that have received little to no inventory activity within the past century, particularly in the West, leading us to conclude that there are untold numbers of species, across the phylogenetic spectrum, that await discovery.

The process required to identify and document these species is far from an academic exercise, and indeed, is urgent. Madagascar has been designated as one of the most critical geographic priorities for conservation action (Andreone and Luiselli, 2003; Myers *et al.*, 2000; Sechrest *et al.*, 2002), retaining <10% of the natural habitats that existed before human colonization (Dufils, 2003; Green and Sussman, 1990; Horning, 2000). The coming few years offer an unprecedented opportunity for working with the Malagasy Government to establish conservation priorities, and may possibly represent the last chance to make large-scale progress in the designation of protected areas. There is keen interest among Malagasy officials to prioritize regions of the country in need of protection, and these priorities will be largely based upon basic biological knowledge relating to species diversity and distribution. Given this urgency, we as biologists do not have the luxury to contemplate and deliberate the meaning of species, without simultaneously taking the necessary action required to formulate biological hypotheses of species distribution and abundance. In other words, the urgency of the problem does not coincide well with leisurely reflection. Numerous investigators have chosen to mitigate the uncertainties of species identification by relying upon the concept of the evolutionary significant unit (ESU) (Crandall *et al.*, 2000; Karl and Bowen, 1999; Moritz, 1994, 1995; Ryder, 1986). In this conceptual framework, the investigator is concerned with recognizing the evolutionary heritage and potential of a given population, typically by using genetic tools (Moritz, 1995), with a subsequent focus on the long-term conservation of that population. This approach has significant appeal for conservation biologists, although it does not entirely avoid the theoretical and operational issues inherent to the species problem (Moritz, 1994).

In this article, we describe a series of empirical and analytical steps that we undertake to accomplish the three goals of identification, classification, and evolutionary study that Mayr set out for the systematist bent upon species discovery. We do so by presenting four case studies, of different vertebrate groups, that are currently in various stages of development. The methodological steps that we employ in such investigations are taken in no predefined order (beyond the fact that field inventory is primary) and are certainly not unique to our enterprise. They are, however, mutually illuminating, and, when conducted as a coordinated effort among biologists of complementary basic, analytic, and organismal ex-

pertise, can be both efficient and powerful for identifying species units and for analyzing their evolutionary context.

A GENERAL APPROACH FOR RECOGNIZING, DESCRIBING, AND UNDERSTANDING SPECIES DIVERSITY IN AN UNDEREXPLORED ENVIRONMENT: CASE STUDIES FROM MADAGASCAR

The accumulated experience of our group, as well as that of other investigators (Carleton and Goodman, 1996; Glaw and Vences, 1994; Hafen *et al.*, 1998; Jenkins, 1992; Vences and Glaw, 2002), has shown that much of the vertebrate species diversity in Madagascar is ''cryptic.'' The proximal causes of obscure species diversity relate to a variety of issues, both biological and practical. It is often true that there are few if any externally visible diagnostic features associated with species identity, or, if they are present, the variation is often so subtle as to be detectable only by a highly trained specialist in that particular vertebrate group. In such cases, species can be said to be cryptic in the definitive sense of the word. Mayr originally described such cryptic variation among closely related species as ''sibling species'' (Mayr, 1942). As he noted, cryptic variation can be especially problematic for poorly analyzed groups, with the assumption that subtle diagnostic characters exist for those species, but have not yet been discovered. Mayr also states in more recent work (1995) that there are also a great number of ''good biological species'' that do not differ phenotypically at all, or only so slightly as to be easily confounded with intraspecific variation. All of the biological and practical variables raised by Mayr apply in the case of species discovery in Madagascar. To confound the issues of cryptic biological diversity, the resolution of species identity has been hampered due to a lack of specimen material and genetic sampling at the population level necessary for understanding patterns of variation. Thus, it is our task to first assemble the necessary specimen data within and among populations, and across their geographic distributions, before we can even attempt to penetrate the biological complexities of true cryptic variation.

As stated above, the essential first step in formulating our hypotheses of species diversity begins with field inventory activities. A field team of researchers based at WWF Madagascar, in the context of a project known as the Ecology Training Program (ETP), has a field inventory program associated with documenting the biota of the island. This team has developed a field methodology that allows for rapid yet thorough biological inventory. The information gained from these surveys is the critical first step toward establishing the sound biological data needed to support the designation of future protected areas. In determining the suitability and

need for protection, it is essential first to determine the density and diversity of species contained within that area. This information provides the essential groundwork for understanding the evolution and ecosystem dynamics that uniquely define a given habitat. The data gathered are vital for understanding patterns of species turnover along different types of ecological and geographic gradients, and for understanding their relationship to a series of biotic and abiotic parameters. After years of working with a variety of terrestrial vertebrates, the survey team members have accumulated a largely unsurpassed knowledge of the relevant groups. Most team members have an exceptional ability to detect subtle differences in coloration, morphology, and other systematic characters associated with their respective study groups. It is this subtle power of detection of novel biological diversity that usually triggers a more detailed study of the morphology, genetics, phylogenetics, and patterns of geographic distribution of newly discovered organisms.

Upon completion of field collection, our first task typically involves reconciling the observed diversity with existing taxonomy, using both specimen data and molecular phylogenetic methods. Current classifications are often complex and misleading, either because certain organismal specialists have overemphasized slight morphological differences in erecting their taxonomy (i.e., oversplitting), or the converse, where others may have overlooked important biological clues, at least in part due to insufficient number of specimens across the pertinent geographical range. Thus, our first exercise in the lab is typically to generate a molecular phylogeny as a first approximation of the fit of current taxonomy to patterns of historical diversification among species. From there, we generally focus our attention on morphological and genetic covariation. In certain cases, it is the observation of subtle morphological variation that instigates the analysis of genetic patterns (Rasoloarison *et al.*, 2000), whereas in other cases it has been the opposite situation (Olson *et al.*, 2004). In all cases, we find that the proper estimation of species identity is essential for understanding the historical underpinnings of the species distributions and interactions (Olson *et al.*, 2004; Rasoloarison *et al.*, 2000; K.L.H., C.H., R.R., S.M.G., and A.D.Y., unpublished work; and A.L.R., J.R., E. Palkovacs, S.M.G., and A.D.Y., unpublished work).

Here, we present four case studies, progressively ordered by their degree of development, to illustrate our approach. Each study focuses on a species complex endemic to Madagascar, and each, we hope, illustrates the importance of comprehensive biological inventory and the careful examination of associated specimens for analyzing vertebrate species in the context of historical, geographic, phenotypic, and genotypic context. It is this general synthetic approach that allows us to formulate our hypotheses of species identity, which then serve as the fundamental frame-

work to be offered to other biologists for further testing. As Mayr has so forcefully yet eloquently argued, a veritable symphony of isolating mechanisms can differentially be at play in any given species complex. Thus, the biologist is well served to investigate any and all ecological, behavioral, and life history factors that may subtly define the margins of reproductive isolation among species.

Case 1: Sorting Out Taxonomy and Generating Species Hypotheses for Malagasy Plated Lizards

The plated lizards (family Cordylidae, subfamily Zonosaurinae) of Madagascar are an ecologically and taxonomically diverse group that consists of >18 species. These animals occupy a nearly comprehensive range of habitat types in Madagascar and can be found in virtually all regions of the island. Current systematic treatments recognize two genera, *Zonosaurus*, which is quite speciose, containing at least 16 species, and *Tracheloptychus*, which contains only two species. Although field inventory has been extensive for this group, our molecular phylogenetic investigations are in their earliest stages. Accordingly, the present focus is simply on mapping the current taxonomy onto the phylogeny to identify problematic areas in need of further investigation. At the same time, we are beginning the process of examining species history in the context of species ranges and habitat preferences. It is such consideration that will ultimately allow us to formulate hypotheses of speciation mechanisms in this group, as well as to identify areas of high diversity and endemism.

Thus far, data from the mitochondrial cytochrome *b* gene (Fig. 11.1) support the reciprocal monophyly of the two named genera. Additionally, these data support the monophyly of most of the species contained in the genus *Zonosaurus*. In this genus, the five most widely distributed species (*Zonosaurus aeneus*, *Zonosaurus karsteni*, *Zonosaurus laticaudatus*, *Zonosaurus madagascariensis*, and *Zonosaurus ornatus*) exhibit very different phylogenetic structures. Whereas *Z. karsteni*, *Z. laticaudatus*, and *Z. ornatus* exhibit species monophyly, the widely distributed *Z. madagascariensis* is paraphyletic with respect to *Zonosaurus haraldmeieri*, which, in turn, has a very restricted distribution. Given that *Z. haraldmeieri* is morphologically very similar to and genetically nested within *Z. madagascariensis*, it will probably be advisable to sink *Z. haraldmeieri* into *Z. madagascariensis* if no fixed, diagnostic characters are found to support the continued recognition of the former. In fact, previous distinctions between these two species have been based predominantly on their geographic distributions and habitat preferences. *Z. madagascariensis* is an evergreen rainforest species that is widely distributed across the island, except in the extreme southeast and in the extreme north, whereas *Z.*

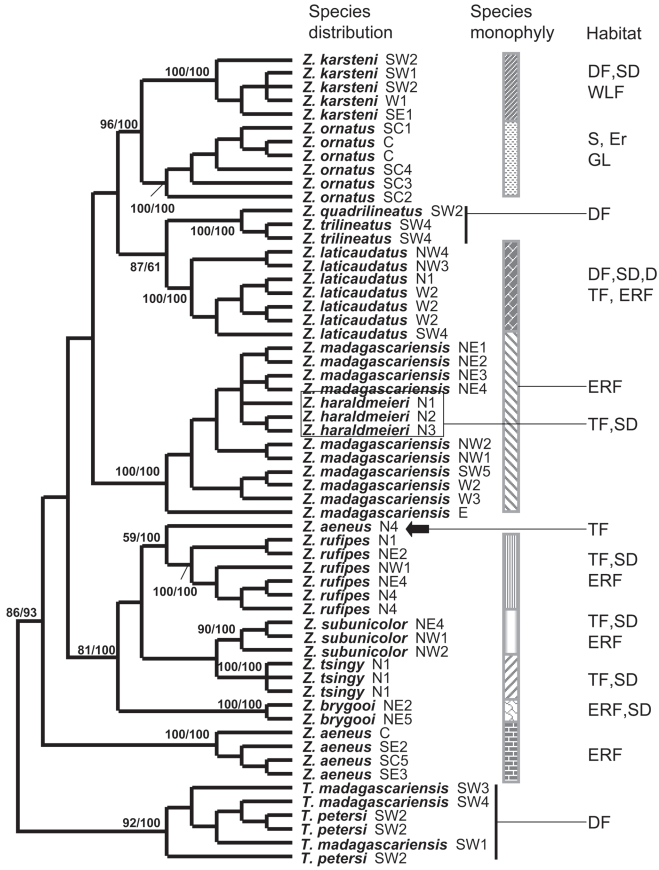


FIGURE 11.1 Molecular phylogeny of plated lizards. Shown is a parsimony tree based on full-length cytochrome *b* sequences. Numbers indicate bootstrap support and posterior probability scores for clades representing various species. Sampling localities are as follows: Northern Madagascar (N) 1, Ankarana; N2, Analamera; N3, Ambre; N4, Daraina. Northeastern Madagascar (NE) 1, Anjanaharibe-Sud; NE2, Antalaha; NE3, Betaolana; NE4, Marojejy; NE5, Tampolo. Northwestern Madagascar (NW) 1, Lokobe; NW2, Manongarivo; NW3, Ambanja; NW4, Ankarafantsika. Eastern Madagascar (E), Mantadia. Central Madagascar (C), Andranomay. South central Madagascar (SC) 1, Itremo; SC2, Vinanintelo; SC3, Manambolo; SC4 Vohipaha; SC5, Ivohibe. Western Madagascar (W) 1, Ambatomainty; W2, Bemaraha; W3, Ambohijanahary. Southeastern Madagascar (SE) 1, Petriky; SE2, Andohahela; SE3, Midongy-Sud. Southwestern Madagascar (SW) 1, Kirindy Mitea; SW2, Mike-Abrahama/Andalandomo/Ankotapike; SW3, Andotabo; SW4, Tsimanampetsotsa; SW5, Analavelona. Habitat abbreviations are: DF (dry forest); ERF (evergreen rainforest); WLS (woodland savannah); S (scrub and heath land); TF (transitional forest); SD (semideciduous forest); D (deciduous forest); GL (grass land); Er (ericoid forest). Sequences are deposited in the GenBank database under accession nos. DQ004403–DQ004461. (Adapted from A.P.R., K.P.K., and A.D.Y., unpublished work.)

haraldmeieri is a semideciduous species found in a small area in the extreme north.

In the case of *Z. aeneus*, an individual from Daraina (indicated with an arrow in Fig. 11.1) branches with *Zonosaurus rufipes* instead of with the rest of *Z. aeneus* populations. In addition, the Daraina population of *Z. aeneus* is geographically closer to *Z. rufipes* than to the rest of *Z. aeneus*. These findings, if upheld with additional sampling of the Daraina population, suggest the presence of a cryptic species among *Z. rufipes*, with the overall morphological similarity between true *Z. aeneus* and ''*Z. aeneus*'' from Daraina due to convergence. Alternately, we must hold open the possibility that this is simply a case of specimen misidentification until additional specimens can be collected. It is essential that we return to this locality to more fully sample this potentially crucial population, an issue that probably would not have emerged in the absence of this phylogenetic analysis. Because Daraina is ecologically different from the other areas where *Z. rufipes* is distributed (Daraina is drier transitional forest instead of humid evergreen rain forest), the situation is quite intriguing in that it may potentially point to an example of parapatric speciation.

The cytochrome *b* phylogeny also confirms the species status of the closely related species pair *Zonosaurus subunicolor* and *Z. rufipes*, which have overlapping distributions. Z. *subunicolor* has long been considered a subspecies of *Z. rufipes* and was recently elevated to species level, although without strong support (Glaw and Vences, 1994). Two years later, in 1996, Vences *et al*. (1996) provided more detailed information based on coloration and habitat preference to justify the resurrection of *Z. subunicolor*. Our phylogenetic analysis of cytochrome *b* supports the reciprocal monophyly of these two species, thus supporting the species designations based on morphological studies.

Continued investigation will also focus on *Z. rufipes* and the *Zonosaurus quadrilineatus–Zonosaurus trilineatus* complex for which we presently have limited geographic sampling. *Z. rufipes* presents geographically variable color patterns, and populations from some localities show fragile body scales, a physiological character used to distinguish some plated lizard species (Raselimanana *et al*., 2000). *Z. quadrilineatus* and *Z. trilineatus* are presently recognized as separate species although they are morphologically quite similar. The number of light stripes on the back used to distinguish them show intermediary forms within the same population, and these two species are never found sympatrically. Their respective distributions are separated by the Onilahy River. Thus, it is presently unclear whether they are perhaps conspecific, although geographically isolated. It will be important to sample more individuals from the two putative species across their geographic ranges to more fully address their hypothesized species status.

CASE 2: DEFINING GEOGRAPHIC BOUNDARIES AMONG SPECIES AND RECONSTRUCTING THE HISTORY OF TRIDENT BATS IN MADAGASCAR

An ongoing study of trident bats (genus *Triaenops*, family Hipposideridae) demonstrates the ways in which extensive sampling within Madagascar yields biogeographic insights both within and beyond the island's physical limits (A.L.R., J.R., E. Palkovacs, S.M.G., and A.D.Y., unpublished work). On the basis of a recent morphological study, three species are currently recognized: *Triaenops rufus*, *Triaenops furculus*, and *Triaenops auritus* (J.R. and S.M.G., unpublished work). As illustrated in Fig. 11.2, this result is supported by molecular phylogenetic analysis. Moreover, when the Malagasy species are analyzed with their African congener (*Triaenops persicus*), the phylogeny reveals that the Malagasy members of this genus are paraphyletic with respect to the African species. Thus, two dispersal events between Africa and Madagascar must be invoked to explain this distribution. The unanswered question at present is whether Africa served as the center of origin, with two dispersal events to Madagascar, or whether Madagascar served as the center of diversification, with (presumably) a back migration to Africa.

We are presently employing population genetic methods to address these competing dispersal hypotheses, as well as to test the hypothesis that one of the northern rivers in western Madagascar may act as a biogeographic barrier separating *T. auritus* and *T. furculus*. Neutrality tests, FS (Fu and Li, 1997) and R2 (Ramos-Onsins and Rozas, 2002), and mismatch distributions (Slatkin and Hudson, 1991) support a history of population expansion in both *T. rufus* and *T. furculus*, with the strong indication that expansion was much more recent in *T. rufus*. Results from *T. auritus* are consistent with a history of constant population size through time, and may represent an older lineage that is at mutation-drift equilibrium. These results therefore seem to support two allochronic dispersals from Africa to Madagascar. The more northern populations of *T. furculus* (Namoroka and Anjohibe) are significantly differentiated from those in the south, but genetic variation within the two regions, respectively, is considerably lower, lending support to the north/south biogeographic structuring observed in some other Malagasy mammals (Yoder *et al.*, 2000). Conversely, analyses of genetic structure within *T. rufus* show a complete lack of geographic structure. Pastorini *et al*. (2003) found that the Betsiboka River formed a major barrier separating populations and species in several different lemur groups. The *Triaenops* data, however, are not consistent with that pattern. Given the vastly different life history and dispersal characteristics in lemurs and bats, it should not be surpris-

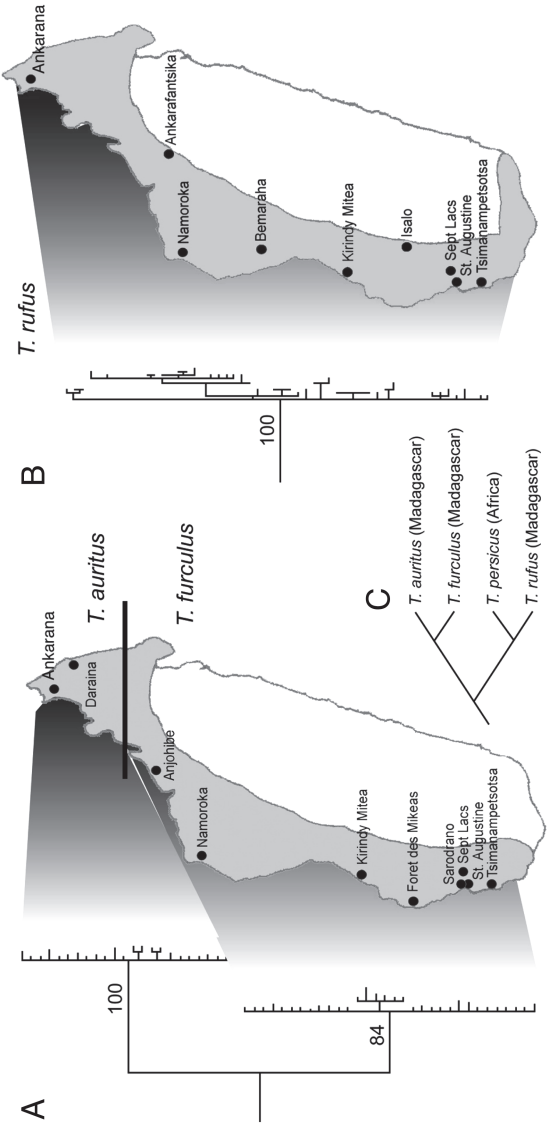


FIGURE 11.2 Comparison of phylogenetic and biogeographic structure in Malagasy trident bats (genus *Triaenops*). (*A*) *T. auritus* (resurrected from synonom) (J.R. and S.M.G., unpublished work) and *T. furculus* show very distinct segregation into northern and southern distributions (indicated by solid line). (*B*) *T. rufus* shows diffuse distribution with no apparent biogeographic structure. (C) Comprehensive *Triaenops* phylogeny reveals that Malagasy taxa are paraphyletic with respect to African species, *T. persicus*. Sequences are deposited in the GenBank database under accession nos. DQ005718–DQ005850. (Adapted from A.L.R., E. Palkovacs, J.R., S.M.G., and A.D.Y., unpublished work.)

ing that rivers might present significant barriers to dispersal for one group (lemurs), but not for another (bats).

Case 3: Revealing Unexpected Geographic and Evolutionary Patterns in Long-Tailed Shrew Tenrecs

Recently, Olson *et al*. (2004) used an integrative approach for clarifying species boundaries in one of the more broadly distributed terrestrial mammals on Madagascar (summarized in Fig. 11.3). The lesser long-tailed shrew tenrec (*Microgale longicaudata*), like many members of this most speciose genus of Malagasy mammal, has a complicated taxonomic history. Tenrecs, and shrew tenrecs in particular, exhibit numerous ontogenetic peculiarities that have stymied taxonomists for the better part of a century (see Jenkins *et al*., 1997; MacPhee, 1987). The number of recognized shrew tenrec species has jumped from 10 to 20 in the past two decades alone (Goodman and Soarimalala, 2004b; Jenkins, 2003; MacPhee, 1987; Olson *et al*., 2004), largely due to a notable increase in museum specimens from previously unsurveyed portions of the island, and to a better understanding of patterns of intra- and inter-population variation. Although traditional comparative morphology continues to advance our knowledge of shrew tenrec diversity (Goodman and Soarimalala, 2004b; Jenkins and Goodman, 1996), molecular methods offer the advantage of rapidly uncovering cryptic genetic lineages. This proved to be the case with *M. longicaudata*, a forest-dwelling species with purportedly substantial levels of intraspecific morphometric variation (MacPhee, 1987). Some authors had suggested that a second described species (*Microgale majori*), subsequently synonymized with *M. longicaudata*, may warrant resurrection. The distribution of a putative *M. majori* was suspected to be limited to a small handful of localities. Moreover, the surveys of Ecology Training Program team members found that, at certain sites, intraspecific variation did not seem to be continuous, suggesting that two separate species might be existing in sympatry. Phylogeographic evidence from the mitochondrial ND2 gene recovered two deeply divergent and reciprocally monophyletic haplotype lineages with broad distributional overlap (indeed, members of the two respective clades have been collected in syntopy). The area of sympatric overlap spans the latitudinal extent of Madagascar's remaining forested areas (see Fig. 11.3). Morphometric analyses conducted on the same specimens corroborated the molecular findings, resulting in the resurrection of *M. majori* from synonymy. Several lines of evidence suggest that at least one additional cryptic species exists within this species complex, but current sample sizes are insufficient to rigorously test this hypothesis at present.

Recently proposed methods that integrate molecular and morpho-

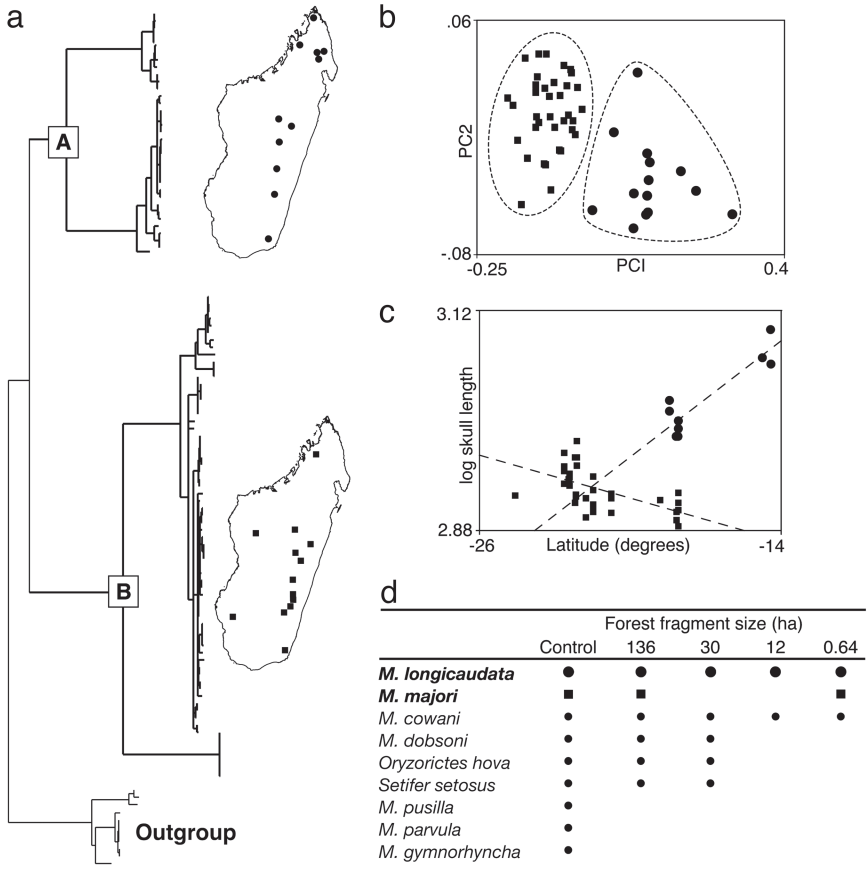


FIGURE 11.3 Overview of the approach used to clarify species limits in long-tailed shrew tenrecs and the subsequent insights into geographic variation and community structure. (*a*) Phylogeographic analysis of mtDNA recovers two cryptic, highly divergent, yet broadly sympatric (and in many cases syntopic), haplotype clades (clades A and B) within the single nominal species of long-tailed shrew tenrec, *M. longicaudata*. (*b*) Despite their striking morphological similarity, members of each haplotype clade are readily distinguished by both *a priori* and *a posteriori* morphometric analyses, supporting the recognition of two cryptic species, *M. longicaudata* (clade A, round symbols) and *M. majori* (clade B, square symbols). (*c*) The revised species-level taxonomy provides insights into biogeography and geographic variation. For example, contrasting patterns of clinal variation in body size were previously obscured. (*d*) Reevaluation of a published study of tenrec community assembly in fragmented forest patches (Goodman and Rakotondravony, 2000) in light of the revised taxonomy shows that both species coexist in remarkably small habitat patches. Sequences are deposited in the GenBank database under accession nos. AY193297–AY193416. (Adapted from figures and text in Olson *et al.*, 2004.)

logical data to identify cryptic species (e.g., Wiens and Penkrot, 2000) do not take potential sympatry among reproductively isolated lineages into account. As was shown with *M. longicaudata* and *M. majori*, widespread sympatry among cryptic species can and does occur (and, indeed, may have contributed to the continued recognition of one rather than two or more species in this case). Second, the revised taxonomy of long-tailed shrew tenrecs revealed contrasting patterns of geographic morphological variation. The failure to recognize two distinct species would have obscured the statistically significant trend toward larger body size at higher latitudes in *M. majori* (Fig. 11.3). Emerging evidence suggests that such latitudinal clinal variation may be much more widespread among Madagascar's tenrecs than has been previously appreciated (L.E.O., unpublished work). Additional insight into elevational segregation and habitat partitioning was also revealed by the phylogeographic analysis (Olson *et al.*, 2004). Finally, the resurrection of *M. majori* from *M. longicaudata* uncovered surprising evidence that these remarkably similar species are able to coexist not only in syntopy, but within the same tiny, isolated forest fragments (Fig. 11.4). Whether this ability to coexist is due to subtle

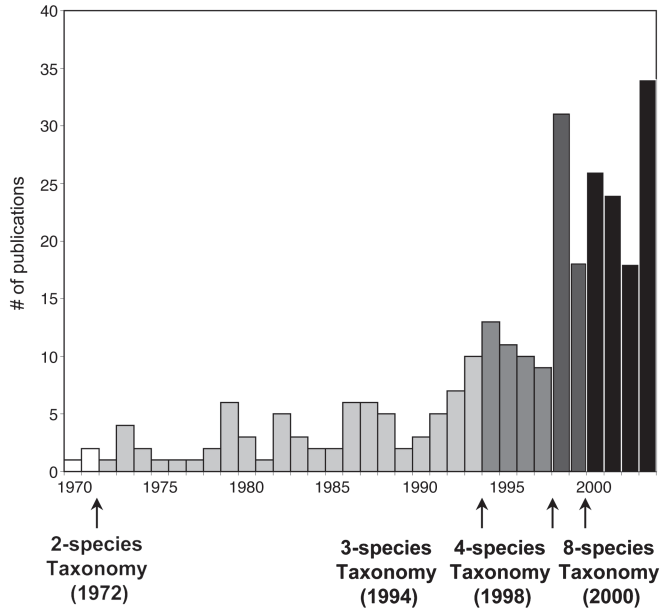


FIGURE 11.4 Graph of the number of publications focusing on genus *Microcebus* from 1970 through 2003. Increased publication activity seems to correlate with increased number of recognized species. Publication numbers were determined by means of a survey of ISI Web of Science.

differences in body size or to some other mechanism, perhaps relating to differential resource utilization, remains unknown. The clarification of potential habitat partitioning is thus a question to be resolved with future field investigations. Collectively, these results challenge the assertion that, within shrew tenrecs, ''no further fine adjustments to taxonomic boundaries are likely to uncover convincing examples of heretofore unknown adaptive types'' (MacPhee, 1987, p. 33). Rather, continued taxonomic refinements offer the best promise for understanding adaptation and diversification in these diminutive yet unequivocally successful mammals.

Case 4: Adding a Temporal Dimension to Species Diversification in Mouse Lemurs, Moving from the Field to the Laboratory and Back Again

The genus *Microcebus* was considered monotypic by most authorities, containing the single species *murinus* (Schwarz, 1931), from the time of its original description in 1795 until the 1970s. After increased research activity and broader geographic review of mouse lemur populations, several investigators reached the conclusion that there were actually at least two distinct forms (Martin, 1972; Petter *et al.*, 1977; Tattersall, 1982): *murinus*, a long-eared gray animal from the western regions of Madagascar and *rufus*, a short-eared reddish animal from the east. Martin (1972), in particular, made note of the differing habitats and ecological constraints defining the two species, with *murinus* inhabiting dry deciduous and spiny desert forest and specializing on insectivory, and *rufus* inhabiting humid rain forest and showing dietary tendencies toward omnivory. Thus, the idea that both ecological and biogeographic mechanisms maintain species separation was an implicit assumption of the two-species taxonomy. The two-species classification remained stable until the early 1990s, after which time studies on mouse lemur biology increased sharply (Atsalis, 2000; Fietz, 1999; Ganzhorn and Schmid, 1998; Genin *et al.*, 2003; Ortmann *et al.*, 1997; Perret and Aujard, 2001; Radespiel *et al.*, 1998; Randrianambinina *et al.*, 2003; Schwab and Ganzhorn, 2004; Wright and Martin, 1995; Zimmermann *et al.*, 2000). The discovery that multiple species co-occurred at several localities in the west was one result of this enhanced research activity, thereby yielding an increase in the number of recognized species from two to four (Schmid and Kappeler, 1994; Thalmann and Rakotoarison, 1994; Zimmermann *et al.*, 1998). Most recently, Rasoloarison *et al*. (2000) described three new species from the western regions of Madagascar, and resurrected another two from synonymy, bringing the total count of recognized species to eight. The Rasoloarison *et al*. species designations were based on a combination of natural history observations, distributional data, and detailed morpho-

metric analysis, with Yoder *et al*. (2000) testing the species designations with mtDNA data.

Combined analysis of three mitochondrial data partitions (HV1 of the control region, cytochrome *b*, and cytochrome oxidase II) consistently yielded reciprocally monophyletic clades congruent with the various species recognized in the Rasoloarison *et al*. study (Yoder *et al.*, 2000). One of the more surprising results from the mtDNA study, however, concerns the phylogenetic placement of two population samples from eastern localities. Although it had been assumed that these two populations would belong to a single clade, congruent with their species designation of *Microcebus rufus*, the two populations were instead found to be paraphyletic with respect to the western species. This result strongly suggests that, contrary to the current recognition of a single species *Microcebus rufus*, there are at least two species of mouse lemurs in the eastern regions of Madagascar, and potentially many more (Yoder *et al.*, 2000).

The quadrupling of recognized mouse lemur species within a period of 6 years (1994–2000) undoubtedly relates (as both cause and consequence) to the enormous amount of investigative energy that has been directed toward these animals during this time interval. As Fig. 11.4 illustrates, the number of empirical studies focused on mouse lemurs has increased in step with the number of recognized species. One might therefore ask whether the increased scrutiny has driven the numerical expansion of recognized species, or the converse, that the appreciation of unexpected species diversity has inspired a renewed interest in these primates. Undoubtedly, both are true, although it is virtually certain that the rather uniform mouse lemur phenotype has until recently retarded our appreciation of their biological diversity. Indeed, recent studies have indicated that the process of mouse lemur diversification began at least 5 million years ago (Yang and Yoder, 2003), making their morphological uniformity all the more intriguing.

In surveying lemur diversity at the species level, we can now appreciate that mouse lemurs are probably the most speciose of all of the Lemuriformes (with the possible exception of genus *Lepilemur*). Genus *Eulemur* ranks a close second, with at least five recognized species. The species count for *Eulemur* has remained nearly stable for at least the past 20 years (Tattersall, 1982) whereas that for *Microcebus* has changed dramatically in the past several years (Rasoloarison *et al.*, 2000; Schmid and Kappeler, 1994; Zimmermann *et al.*, 1998). Ostensibly, the differences in taxonomic stability relate to the fact that the various *Eulemur* species are readily identifiable according to their variety of coloration patterns and other morphological features, whereas *Microcebus* species are not. But why is this so? Given that we now suspect that the temporal origins of the two groups are nearly contemporaneous (Yoder and Yang, 2004), why

should rates of apparent morphological evolution have been markedly more rapid in one genus than in the other? The answer probably relates to the fact that *Eulemur* is primarily diurnal whereas *Microcebus* is strictly nocturnal.

For mammals, visual signals will be most efficiently transmitted and received by day, and other signals, such as acoustic or olfactory, will be required for nocturnal signaling. Thus, mate choice criteria will tend to mirror the signal transmission favored in a given environment (Endler, 1992). As discussed above, the species diversity within the genus *Microcebus* had been underestimated for many years, and, although we can now identify subtle patterns of coloration and morphometric variation as distinguishing among species, it is not a stretch to refer to them as a cryptic species radiation (*sensu* of Mayr, 1995). Although the various species contained within the diurnal genus *Eulemur* show a notable array of sexually dichromic pelage variation, with males in particular showing species-specific head ornamentation, mouse lemurs are uniformly drab, showing no sexual dichromatism. These patterns perfectly fit with the prediction that diurnal animals will emphasize visual cues for mate selection whereas nocturnal animals will emphasize olfactory and auditory signals (Jones, 1997).

This prediction as applied to mouse lemurs seems to be born out by studies demonstrating that olfactory and hormonal signals conveyed by means of urine exposure can have powerful effects on both behavior and on basic physiological and reproductive functions in these mammals. For example, exposure to female urine can significantly increase testosterone levels in males, just as exposure to the urine of dominant males can suppress testosterone production in other males (Perret and Schilling, 1995). Acoustic studies in particular have revealed subtleties in signaling, with two results noteworthy in their implications for potential mate-choice mechanisms. First, acoustic signals in mouse lemurs seem to evolve extremely rapidly, and second, the greatest levels of acoustic separation occur in the sexual advertisement calls of males. Relevant to the issue of rapid rates, studies of captive mouse lemur colonies in Europe reveal that colonies that have been separated for only a few generations have already begun to develop distinct dialects in their acoustic signals (Zimmermann and Hafen, 2001). Similarly, a detailed field study conducted in Madagascar revealed that the sexual advertisement calls of males occurring in demes separated by only 1.5 km or so showed distinct differences, even though there were no apparent biogeographic barriers separating the demes (Hafen *et al.*, 1998). Moreover, when sexual advertisement calls were compared with predator advertisement calls in two species from widely separated habitats, it was found that, although there was a great deal of overlapping interspecific variation in the predator calls, the sexual

advertisement calls were entirely and profoundly distinct (Zimmermann *et al.*, 2000).

Thus, auditory and chemosensory data lend support to the morphological and mitochondrial hypotheses of mouse lemur species diversity. Moreover, numerous field studies are reporting that sympatric mouse lemur species practice microhabitat partitioning, either in their choice of nesting sites (Radespiel *et al.*, 2003; Weidt *et al.*, 2004) and/or with regard to competitive exclusion relating to as-yet-undetermined resources (Schwab and Ganzhorn, 2004). It should be noted, however, that, up to the present, all published studies that have examined sympatric overlap between mouse lemur species have demonstrated sympatry only for *Microcebus murinus* plus another species (Radespiel *et al.*, 2003; Schmid and Kappeler, 1994; Schwab and Ganzhorn, 2004; Weidt *et al.*, 2004; Yoder *et al.*, 2002; Zimmermann *et al.*, 1998). Sympatric overlap of species not including *Microcebus murinus* has yet to be reported.

Even in light of the accumulating morphological, genetic, and behavioral evidence supporting the species level status of the eight mouse lemur groups (Rasoloarison *et al.*, 2000; Yoder *et al.*, 2000), more can be done to investigate the evolutionary barriers among these putative species. For example, careful scrutiny of morphological variation and genetic divergence in this group indicates that the patterns of covariation are not uniform. It seems that high levels of genetic divergence do not necessarily predict clear-cut morphological divergence, just as clearcut levels of morphological divergence do not necessarily indicate high levels of genetic distance (Fig. 11.5). For these and other reasons, we are further exploring mouse lemur species patterns with nuclear loci, as advocated by many investigators (Brumfield *et al.*, 2003; Edwards and Beerli, 2000; Zhang and Hewitt, 2003). Here, the results are proving interesting, although far from decisive with relation to the question of species identity. As exemplified in Fig. 11.6, the nuclear markers that we have investigated thus far tend to show patterns of either incomplete lineage sorting, or perhaps persistent hybridization among several species. Only with continued genetic sampling and analysis will we be able to differentiate between these (and potentially other) explanations.

Finally, among the most intriguing of the genetic results is the observation that *Microcebus ravelobensis* is genetically quite diverged from other mouse lemur species (Fig. 11.6). This result becomes particularly interesting when considered in the light of emerging physiological data. It has long been known that mouse lemurs demonstrate a physiological specialization for torpor (Aujard *et al.*, 1998; Ganzhorn and Schmid, 1998; Genin and Perret, 2003; Martin, 1973; Randrianambinina *et al.*, 2003). Numerous studies of naturally occurring mouse lemur populations have shown a strong seasonal pattern of daily torpor in response both to changes in

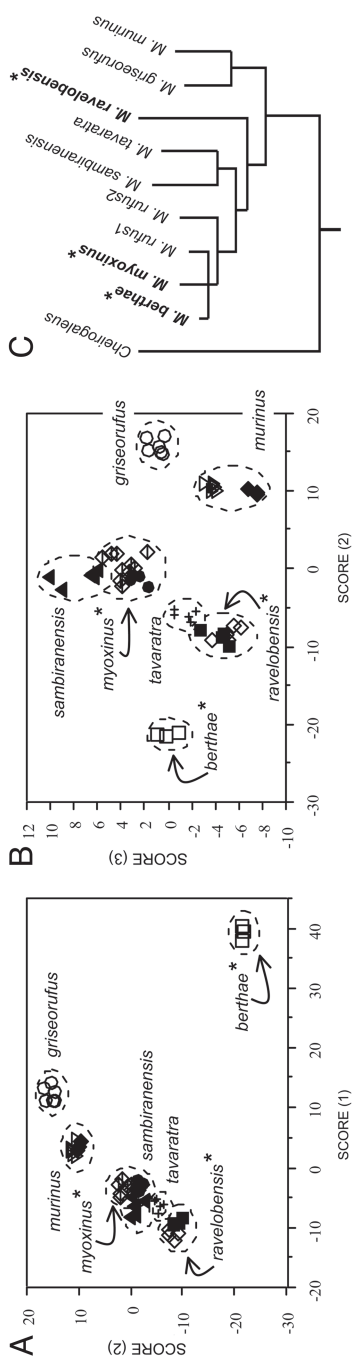


FIGURE 11.5 This figure illustrates the lack of precise correlation between morphometric distinctiveness and genetic divergence in mouse lemur species. The results of discriminant function analysis of 34 cranial, dental, and external morphometric characters are redrawn from Yoder *et al.* (2000). Functions 1 and 2 (*A*) show conspicuous discrimination of *Microcebus berthae* from other species, but otherwise do not discriminate well among species. Functions 2 and 3 (*B*) show discrimination of all species, with *Microcebus berthae* remaining as highly distinct. (C) A maximum likelihood phylogram of species-specific haplotypes derived from the fibrinogen α intron 4 locus illustrates that *Microcebus berthae* and *Microcebus myoxinus* are genetically very similar. Conversely, *Microcebus ravelobensis* is genetically and phylogenetically divergent.

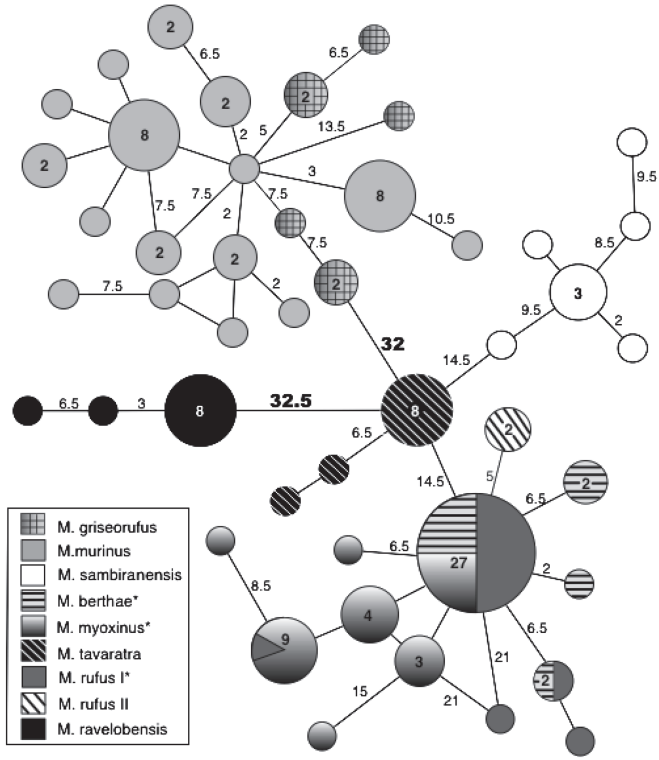


FIGURE 11.6 Minimum spanning network of the fibrinogen α intron 4 (609 bp) in genus *Microcebus*. This network was calculated in ARLEQUIN 2.000 (Schneider *et al.*, 2000) using pairwise differences between haplotypes. Each shading represents an individual species. Numbers inside circles and squares are the number of individuals sharing a haplotype; empty circles equal one individual. Numbers on connecting lines are the number of nucleotide changes separating each haplotype; empty lines equal one change. Note that all alleles for *Microcebus ravelobensis* are species-specific and are highly diverged from alleles sampled from other species. Conversely, the allele in greatest frequency within the *myoxinus-berthae-rufus*1 clade is identical among the three species. Sequences are deposited in the GenBank database under accession nos. DQ003345–DQ003479. (Adapted from K.L.H., R.R., S.M.G., and A.D.Y., unpublished work.)

ambient temperature and to photoperiodic variations, with this behavior presumably adapting them for extreme resource limitation during Madagascar's dry season (Aujard *et al.*, 1998; Ganzhorn and Schmid, 1998; Genin and Perret, 2000). Mouse lemurs progressively accumulate fat stores during the wet season (Genin and Perret, 2000), after which time

they spontaneously enter a period of daily torpor during the dry season. Presently, however, our information is limited to a few localities in western Madagascar, and predominantly, to a single species, *Microcebus murinus*. It is therefore of evolutionary consequence to ask whether this unusual system is characteristic of all mouse lemur species and populations, or only to a subset of species and habitats. At present, there is preliminary indication that *Microcebus ravelobensis* is perhaps unique among mouse lemurs in that it does not enter torpor (Randrianambinina *et al.*, 2003). Given our new-found appreciation for the genetically divergent position of this species, this physiological anomaly becomes all the more meaningful. It will be a fascinating exercise to return to the field, giving refined scrutiny to the ecological and other biological characteristics of this highly derived mouse lemur species.

SUMMARY

With the case studies above, we hope to have illustrated both the complexity and the importance of resolving species boundaries in nature. As a process of discovery, the identification of species requires a multidimensional approach that employs tools spanning everything from human intuition to molecular phylogenetics. In the cases highlighted above, we have attempted to illustrate the roles that basic biological inventory, the collection of specimen data, and the careful analysis of morphological and genetic variation can play in delimiting species boundaries, or at least for hypothesizing their existence. As we have shown, the impacts of this exercise can be far-reaching, spanning disciplines of behavioral ecology to biogeography to conservation biology. By providing biologists with a hypothesis of species identity, distribution, and abundance, a cascade of investigation is prompted that will ultimately reflect back upon and refine the initial hypothesis (Hey *et al.*, 2003). As described above, systematic clarification, coupled with the revelation of cryptic biological diversity, is yielding insight into the myriad of ecological, evolutionary, and biogeographic forces that have shaped Madagascar's vertebrate diversity. The need for species discovery and documentation in Madagascar and other biodiversity hotspots is urgent. There remains a vast wealth of species yet to be identified, and their future is uncertain at best. To address these problems, we advocate a collaborative structure among biologists of differing organismal, methodological, and analytical skills, as well as of differing cultural backgrounds, so that the process can move rapidly and expertly from field, to lab, and back again. Furthermore, as we are discovering in our work, Mayr's ''three basic tasks'' of the systematist provide an ideal framework for collecting, synthesizing, and implementing the acquired data for both analytical and applied goals.

ACKNOWLEDGMENTS

A.D.Y. thanks Francisco Ayala, Walter Fitch, and Jody Hey for their invitation to participate in this colloquium celebrating the life and works of Ernst Mayr. This work has been supported by a National Science Foundation-CAREER Award and a Biodiversity Leadership Award from the Bay and Paul Foundations (to A.D.Y.). S.M.G. has been supported by awards from the MacArthur Foundation, the Volkswagen Foundation, and the Critical Ecosystem Partnership Fund of Conservation International.

REFERENCES

Andreone, F. & Greer, A. E. (2002) Malagasy scincid lizards: Descriptions of nine new species, with notes on the morphology, reproduction, and taxonomy of some previously described species (Reptilia, Squamata: Scincidae). *J. Zool.* **258,** 139–181.

Andreone, F. & Luiselli, L. M. (2003) Conservation priorities and potential threats influencing the hyper-diverse amphibians of Madagascar. *Ital. J. Zool.* **70,** 53–63.

Atsalis, S. (2000) Spatial distribution and population composition of the brown mouse lemur (*Microcebus rufus*) in Ranomafana National Park, Madagascar, and its implications for social organization. *Am. J. Primatol.* **51,** 61–78.

Aujard, F., Perret, M. & Vannier, G. (1998) Thermoregulatory responses to variations of photoperiod and ambient temperature in the male lesser mouse lemur: A primitive or an advanced adaptive character? *J. Comp. Physiol. B* **168,** 540–548.

Brumfield, R. T., Beerli, P., Nickerson, D. A. & Edwards, S. V. (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol. Evol.* **18,** 249–256.

Carleton, M. D. & Goodman, S. M. (1996) Systematic studies of Madagascar's endemic rodents (Muroidea: Nesomyinae): A new genus and species from the Central Highlands. *Fieldiana (Zool.)* **85,** 231–256.

Coffin, M. F. & Rabinowitz, P. D. (1992) The Mesozoic East African and Madagascan conjugate continental margins. In *Geology and Geophysics of Continental Margins*, eds. Watkins, J. S., Zhiqiang, F. & McMillen, K. J. (American Association of Petroleum Geologists Memoir), Vol. 53, pp. 207–246.

Crandall, K. A., Bininda-Emonds, O. R. P., Mace, G. M. & Wayne, R. K. (2000) Considering evolutionary processes in conservation biology. *Trends Ecol. Evol.* **15,** 290–295.

de Wit, M. (2003) Madagascar: Heads it's a continent, tails its an island. *Annu. Rev. Earth Planet. Sci.* **31,** 213–248.

Dufils, J.-M. (2003) Remaining forest cover. In *The Natural History of Madagascar*, eds. Goodman, S. M. & Benstead, J. P. (Univ. of Chicago Press, Chicago), pp. 88–96.

Edwards, S. V. & Beerli, P. (2000) Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution Int. J. Org. Evolution* **54,** 1839–1854.

Endler, J. A. (1992) Signals, signal conditions, and the direction of evolution. *Am. Nat.* **139,** S125–S153.

Fietz, J. (1999) Mating system of *Microcebus murinus*. *Am. J. Primatol.* **48,** 127–133.

Fu, Y.-X. & Li, W.-H. (1997) Estimating the age of the common ancestor of a sample of DNA sequences. *Mol. Biol. Evol.* **14,** 195–199.

Ganzhorn, J. U. & Schmid, J. (1998) Different population dynamics of *Microcebus murinus* in primary and secondary deciduous dry forests of Madagascar. *Int. J. Primatol.* **19,** 785–796.

Genin, F. & Perret, M. (2000) Photoperiod-induced changes in energy balance in gray mouse lemurs. *Physiol. Behav.* **71,** 315–321.

Genin, F. & Perret, M. (2003) Daily hypothermia in captive grey mouse lemurs (*Microcebus murinus*): Effects of photoperiod and food restriction. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **136,** 71–81.

Genin, F., Nibbelink, M., Galand, M., Perret, M. & Ambid, L. (2003) Brown fat and nonshivering thermogenesis in the gray mouse lemur (*Microcebus murinus*). *Am. J. Physiol.* **284,** R811–R818.

Glaw, F. & Vences, M. (1994) *A Fieldguide to the Amphibians and Reptiles of Madagascar* (Moos Druck, Leverkussen, Germany).

Goodman, S. M. & Benstead, J. P., eds. (2003) *The Natural History of Madagascar* (Univ. of Chicago Press, Chicago).

Goodman, S. M. & Cardiff, S. G. (2004) A new species of *Chaerephon* (Molossidae) from Madagascar with notes on other members of the family. *Acta Chiropt.* **69,** 75–81.

Goodman, S. M. & Rakotondravony, D. (2000) *J. Zool.* **250,** 193–200.

Goodman, S. M. & Soarimalala, V. (2004a) A new species of *Macrotarsomys* (Rodentia: Muridae: Nesomyinae) from the Forêt des Mikea of southwestern Madagascar. *Proc. Biol. Soc. Washington* **117,** 265–279.

Goodman, S. M. & Soarimalala, V. (2004b) A new species of *Microgale* (Lipotyphla: Tenrecidae: Oryzorictinae) from the Forêt des Mikea of southwestern Madagascar. *Proc. Biol. Soc. Washington* **117,** 250–264.

Goodman, S. M., Hawkins, A. F. A. & Domergue, C. A. (1997) A new species of vanga (Family Vangidae: Calicalicus) from southwestern Madagascar. *Bull. Br. Ornithol. Club* **117,** 4–10.

Green, G. M. & Sussman, R. W. (1990) Deforestation history of the eastern rain forests of Madagascar from satellite images. *Science* **248,** 212–214.

Hafen, T., Neveu, H., Rumpler, Y., Wilden, I. & Zimmermann, E. (1998) Acoustically dimorphic advertisement calls separate morphologically and genetically homogenous populations of the grey mouse lemur (*Microcebus murinus*). *Folia Primatol.* **69,** Suppl. 1, 342–356.

Hey, J., Waples, R. S., Arnold, M. L., Butlin, R. K. & Harrison, R. G. (2003) Understanding and confronting species uncertainty in biology and conservation. *Trends Ecol. Evol.* **18,** 597–603.

Horning, N. (2000) Changes in forest cover from 1993/94–1997/98 in target zones around protected areas. *Lemur News* **5,** 28–30.

Jenkins, P. D. (1992) Description of a new species of *Microgale* (Insectivora: Tenrecidae) from eastern Madagascar. *Bull. Br. Mus. Nat. Hist. (Zool).* **58,** 53–59.

Jenkins, P. D. (2003) *Microgale*, shrew tenrecs. In *The Natural History of Madagascar*, eds. Goodman, S. M. & Benstead, J. P. (Univ. of Chicago Press, Chicago), pp. 1273–1278.

Jenkins, P. D. & Goodman, S. M. (1996) A new species of *Microgale* (Lipotyphla, Tenrecidae) from isolated forest in southwestern Madagascar. *Bull. Nat. Hist. Mus. London* **65,** 155–164.

Jenkins, P. D., Raxworthy, C. J. & Nussbaum, R. A. (1997) A new species of *Microgale* (Insectivora, Tenrecidae), with comments on the status of four other taxa of shrew tenrecs. *Bull. Nat. Hist. Mus. London (Zool.)* **63,** 1–12.

Jones, G. (1997) Acoustic signals and speciation: The roles of natural and sexual selection in the evolution of cryptic species. *Adv. Study Behav.* **26,** 317–354.

Karl, S. A. & Bowen, B. W. (1999) Evolutionary significant units versus geopolitical taxonomy: Molecular systematics of an endangered sea turtle (genus *Chelonia*). *Conserv. Biol*. **13,** 990–999.

MacPhee, R. D. E. (1987) The shrew tenrecs of Madagascar: Systematic revision and Holocene distribution of *Microgale* (Tenrecidae, Insectivora). *Am. Mus. Novit*. **2889,** 1–45.

Martin, R. D. (1972) A preliminary field-study of the lesser mouse lemur (*Microcebus murinus* J.F. Miller, 1777). *Z. Tierpsychol*. **9,** Suppl., 43–89.

Martin, R. D. (1973) A review of the behaviour and ecology of the lesser mouse lemur (*Microcebus murinus* J.F. Miller, 1777). In *Comparative Ecology and Behaviour of Primates*, eds. Michael, R. P. & Crook, J. H. (Academic, London), pp. 1–68.

Mayr, E. (1942) *Systematics and the Origin of Species from the Viewpoint of a Zoologist* (Harvard Univ. Press, Cambridge, MA).

Mayr, E. (1995) What is a species, and what is not? *Philos. Sci*. **63,** 262–277.

Moritz, C. (1994) Defining "evolutionarily significant units" for conservation. *Trends Ecol. Evol*. **9,** 373–375.

Moritz, C. (1995) Uses of molecular phylogenies for conservation. *Philos. Trans. R. Soc. London B* **349,** 113–118.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature* **403,** 853–858.

Olson, L. E., Goodman, S. M. & Yoder, A. D. (2004) Illumination of cryptic species boundaries in long-tailed shrew tenrecs (Mammalia: Tenrecidae; *Microgale*) provides insights into geographic variation and distributional constraints. *Biol. J. Linn. Soc*. **83,** 1–22.

Ortmann, S., Heldmaier, G., Schmid, J. & Ganzhorn, J. U. (1997) Spontaneous daily torpor in Malagasy mouse lemurs. *Naturwissenschaften* **84,** 28–32.

Pastorini, J., Thalmann, U. & Martin, R. D. (2003) A molecular approach to comparative phylogeography of extant Malagasy lemurs. *Proc. Natl. Acad. Sci. USA* **100,** 5879–5884.

Perret, M. & Aujard, F. (2001) Daily hypothermia and torpor in a tropical primate: synchronization by 24-h light-dark cycle. *Am. J. Physiol*. **281,** R1925–R1933.

Perret, M. & Schilling, A. (1995) Sexual responses to urinary chemosignals depend on photoperiod in a male primate. *Physiol. Behav*. **58,** 633–639.

Petter, J.-J., Albignac, R. & Rumpler, Y. (1977) Mammifères Lémuriens (Primates Prosimiens). In *Faune de Madagascar* (ORSTOM and CNRS, Paris), Vol. 44, pp. 1–513.

Radespiel, U., Cepok, S., Zietemann, V. & Zimmermann, E. (1998) Sex-specific usage patterns of sleeping sites in grey mouse lemurs (*Microcebus murinus*) in northwestern Madagascar. *Am. J. Primatol*. **46,** 77–84.

Radespiel, U., Ehresmann, P. & Zimmermann, E. (2003) Species-specific usage of sleeping sites in two sympatric mouse lemur species (*Microcebus murinus* and *M. ravelobensis*) in northwestern Madagascar. *Am. J. Primatol*. **59,** 139–151.

Ramos-Onsins, S. E. & Rozas, J. (2002) Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol*. **19,** 2092–2100.

Randrianambinina, B., Rakotondravony, D., Radespiel, U. & Zimmermann, E. (2003) Seasonal changes in general activity, body mass and reproduction of two small nocturnal primates: A comparison of the golden brown mouse lemur (*Microcebus ravelobensis*) in northwestern Madagascar and the brown mouse lemur (*Microcebus rufus*) in eastern Madagascar. *Primates* **44,** 321–331.

Raselimanana, A. P., Raxworthy, C. J. & Nussbaum, R. A. (2000) A revision of the dwarf *Zonosaurus boulenger* (Reptilia: Squamata: Cordylidae) from Madagascar, including Descriptions of three new species. *Sci. Pap. Univ. Kansas Nat. Hist. Mus*. **18,** 1–16.

Rasoloarison, R. M., Goodman, S. M. & Ganzhorn, J. U. (2000) Taxonomic revision of mouse lemurs (*Microcebus*) in the western portions of Madagascar. *Int. J. Primatol*. **21,** 963–1019.

Ryder, O. A. (1986) Species conservation and systematics: The dilemma of subspecies. *Trends Ecol. Evol.* **1,** 9–10.

Schmid, J. & Kappeler, P. M. (1994) Sympatric mouse lemurs (*Microcebus* spp.) in western Madagascar. *Folia Primatol.* **63,** 162–170.

Schneider, S., Roessli, D. & Excoffier, L. (2000) ARLEQUIN: A Software for Population Genetics Data Analysis (Genetics and Biometry Lab., Dept. of Anthropology, Univ. of Geneva, Geneva), Version 2.000.

Schwab, D. & Ganzhorn, J. U. (2004) Distribution, population structure and habitat use if *Microcebus berthae* compared to those of other sympatric cheirogaleids. *Int. J. Primatol.* **25,** 307–330.

Schwarz, E. (1931) A revision of the genera and species of Madagascar Lemuridae. *Proc. Zool. Soc. London* **1931,** 399–428.

Sechrest, W., Brooks, T. M., da Fonseca, G. A., Konstant, W. R., Mittermeier, R. A., Purvis, A., Rylands, A. B. & Gittleman, J. L. (2002) Hotspots and the conservation of evolutionary history. *Proc. Natl. Acad. Sci. USA* **99,** 2067–2071.

Slatkin, M. & Hudson, R. R. (1991) Pairwise comparisons of mitochondrial-DNA sequences in stable and exponentially growing populations. *Genetics* **129,** 555–562.

Storey, M., Mahoney, J. J., Saunders, A. D., Duncan, R. A., Kelley, S. P. & Coffin, M. F. (1995) Timing of hot spot-related volcanism and the breakup of Madagascar and India. *Science* **267,** 852–855.

Tattersall, I. (1982) *The Primates of Madagascar* (Columbia Univ. Press, New York).

Thalmann, U. & Rakotoarison, N. (1994) Distribution of lemurs in central western Madagascar, with a regional distribution hypothesis. *Folia Primatol.* **63,** 156–161.

Vences, M. & Glaw, F. (2002) Two new treefrogs of the *Boophis rappiodes* group from eastern Madagascar (Amphibia Mantellidae). *Trop. Zool.* **15,** 141–163.

Vences, M. & Glaw, F. (2004) Revision of the subgenus *Chonomantis* (Anura: Mantellidae: Mantidactylus) from Madagascar, with description of two new species. *J. Nat. Hist.* **38,** 77–118.

Vences, M., Müller-Jung, J., Glaw, F. & Böhme, W. (1996) Review of the *Zonosaurus aeneus* species group, with resurrection of *Zonosaurus subunicolor* (Boettger 1881). *Senckenb. Biol.* **76,** 47–59.

Vences, M., Andreone, F., Glaw, F. & Mattioli, F. (2002) New dwarf species of *Mantidactylus* from northwestern Madagascar (Anura: Mantellidae). *Copeia* **2002,** 1057–1062.

Weidt, A., Hagenah, N., Randrianambinina, B., Radespiel, U. & Zimmermann, E. (2004) Social organization of the golden brown mouse lemur (*Microcebus ravelobensis*). *Am. J. Phys. Anthropol.* **123,** 40–51.

Wiens, J. J. & Penkrot, T. A. (2000) Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). *Syst. Biol.* **51,** 69–91.

Wright, P. C. & Martin, L. B. (1995) Predation, pollination and torpor in two nocturnal prosimians: *Cheirogaleus major* and *Microcebus rufus* in the rain forest of Madagascar. In *Creatures of the Dark*, eds. Alterman, L., Doyle, G. A. & Izard, M. K. (Plenum, New York).

Yang, Z. & Yoder, A. D. (2003) Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. *Syst. Biol.* **52,** 705–716.

Yoder, A. D. & Yang, Z. (2004) Divergence dates for Malagasy lemurs estimated from multiple gene loci: Geological and evolutionary context. *Mol. Ecol.* **13,** 757–773.

Yoder, A. D., Rasoloarison, R. M., Goodman, S. M., Irwin, J. A., Atsalis, S., Ravosa, M. J. & Ganzhorn, J. U. (2000) Remarkable species diversity in Malagasy mouse lemurs (Primates, *Microcebus*). *Proc. Natl. Acad. Sci. USA* **97,** 11325–11330.

Yoder, A. D., Burns, M. M. & Genin, F. (2002) Molecular evidence of reproductive isolation in sympatric sibling species of mouse lemurs. *Int. J. Primatol*. **23,** 1335–1343.

Zhang, D. X. & Hewitt, G. M. (2003) Nuclear DNA analyses in genetic studies of populations: Practice, problems and prospects. *Mol. Ecol*. **12,** 563–584.

Zimmermann, E. & Hafen, T. G. (2001) Colony specificity in a social call of mouse lemurs (*Microcebus* spp.). *Am. J. Primatol*. **54,** 129–141.

Zimmermann, E., Cepok, S., Rakotoarison, N., Zietemann, V. & Radespiel, U. (1998) Sympatric mouse lemurs in north-west Madagascar: A new rufous mouse lemur species (*Microcebus ravelobensis*). *Folia Primatol*. **69,** 106–114.

Zimmermann, E., Vorobieva, E.,Wrogemann, D. & Hafen, T. G. (2000) Use of vocal fingerprinting for specific discrimination of gray (*Microcebus murinus*) and rufous mouse lemurs (*Microcebus rufus*). *Int. J. Primatol*. **21,** 837–852.

12
Examining Bacterial Species Under the Specter of Gene Transfer and Exchange

HOWARD OCHMAN,* EMMANUELLE LERAT,† AND VINCENT DAUBIN*

Even in lieu of a dependable species concept for asexual organisms, the classification of bacteria into discrete taxonomic units is considered to be obstructed by the potential for lateral gene transfer (LGT) among lineages at virtually all phylogenetic levels. In most bacterial genomes, large proportions of genes are introduced by LGT, as indicated by their compositional features and/or phylogenetic distributions, and there is also clear evidence of LGT between very distantly related organisms. By adopting a whole-genome approach, which examined the history of every gene in numerous bacterial genomes, we show that LGT does not hamper phylogenetic reconstruction at many of the shallower taxonomic levels. Despite the high levels of gene acquisition, the only taxonomic group for which appreciable amounts of homologous recombination were detected was within bacterial species. Taken as a whole, the results derived from the analysis of complete gene inventories support several of the current means to recognize and define bacterial species.

Real species are typically defined by the ability of their constituents to exchange genes. This activity (i.e., sexual reproduction) goes a long way toward explaining the maintenance of species as cohesive units whose members are closely related and are of similar genetic

Departments of *Biochemistry and Molecular Biophysics and †Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721.

Abbreviations: G + C, guanine plus cytosine; LGT, lateral gene transfer.

architecture. As such, conspecifics share numerous characteristics by which they can be grouped, even when evidence of interbreeding is limited or unknown.

By lacking a mechanism that regularly homogenizes the features of different organisms, strictly asexual organisms continuously diverge from one another as independent lineages. And although the classification of these lineages is undoubtedly useful, it could be argued that any criteria for delineating asexual species, e.g., possessing of a particular suite of phenotypic traits or attaining a prescribe degree of DNA similarity, are arbitrary, inconsistent across taxa, and biologically meaningless.

Acknowledging the problems associated with classifying groups of asexual organisms into discrete species, the situation with bacteria is even worse. Bacteria reproduce asexually, yet they are also capable of obtaining genes from other organisms, even those of different kingdoms. Moreover, the amounts, types, and sources of imported genes can vary among lineages, allowing gene transfer to blur the boundaries of bacterial groups at every taxonomic level and in ways that are impossible to predict. And if patterns of vertical descent are obscured in varied and unknown ways, then the systematic classification of bacteria might not be possible [see Rossello-Mora and Amann (2001), Lan and Reeves (2001), Young (2001), and Cohan (2002) for current reviews on the concept of bacterial species].

There is a clear advantage to examining the process of diversification in bacteria, which is the availability of complete sequences from hundreds of genomes whose relationships range in type from members of the same nominal species to representatives of groups that diverged billions of years ago. These new data allow us to follow the origin and ancestry of every gene in a genome to resolve the degree to which gene transfer has shaped the contents of bacterial genomes and has obscured the history of bacterial groups at different phylogenetic depths.

THE SCOPE OF GENE TRANSFER IN BACTERIA

There are several means by which bacteria can acquire genes: by conjugal transfer, by phage-mediated insertions and by the update of native DNA from the outside sources (Ochman *et al.*, 2000; Redfield, 2001). But given the diversity of mechanisms that are capable of planting virtually any gene in virtually any organism, bacterial genomes remain small (on the order of 500–10,000 kb) and are not simply arbitrary assortments of genes of mixed heritage. Although bacteria might be bombarded constantly with foreign genes, only evolutionarily relevant events of transfer, i.e., those resulting in genes that persist, are evident from the contents of extent genomes.

With the completion of each bacterial genome sequence, there is a

search for horizontally acquired genes. This research most commonly proceeds by scanning the genome sequence for regions of atypical base composition, a surprisingly accurate method for identifying one class of recently acquired genes. The rationale for this approach has its foundations in research performed nearly 50 years ago, when the initial goals were to characterize the nature of nucleic acids within cellular organisms (Rolfe and Meselson, 1959; Sueoka, 1961; Sueoka *et al.*, 1959). By the early 1960s, base composition [usually expressed as the relative proportion of guanine and cytosine (G + C) residues, % G + C] had been determined for hundreds of bacterial genomes, leading to the general observations: (*i*) that the diversity in base composition among bacterial genomes, which ranges from 20% to 80% G + C, is much greater than that in eukaryotes, (*ii*) that despite this variation, the base composition within an organism is fairly consistent over the entire chromosome, and (*iii*) that closely related organisms have similar G + C contents (Muto and Osawa, 1987; Sueoka, 1962, 1988). The observed heterogeneity among genomes, coupled with the compositional homogeneity within genomes, implicates gene transfer between organisms of different G + C contents as a source of intragenomic variation in base composition and codon usage patterns (Daubin *et al.*, 2003a; Guerdoux-Jamet *et al.*, 1997; Lawrence and Ochman, 1998; Medigue *et al.*, 1991; Ragan, 2001).

Naturally, other factors, such as the amino acid contents of a protein, might influence its overall nucleotide composition; however, phylogenetic information supports the use of G + C content as a way to identify acquired genes. Because acquired regions often manifest multiple features that denote their ancestry, it is thus not perhaps surprising that many genes with sporadic distributions, as might occur from a history of lateral transfer, have anomalous base compositions. To illustrate the utility and accuracy of these methods for recognizing acquired genes, Fig. 12.1 shows the amount of protein-coding DNA within each minute (≈45 kb encompassing ≈40 genes) along the *E. coli* MG1655 chromosome having atypical sequence features and/or a phylogenetic distribution indicative lateral transfer (Lawrence and Ochman, 2002).

These procedures rely on very different sorts of information and might be expected to identify somewhat different sets of acquired genes, the degree of overlap (gray portion of each bar in Fig. 12.1) is quite good: among the 755 genes originally identified as being horizontally acquired based on sequence characteristics, nearly 80% display a phylogenetic distribution compatible with lateral gene transfer (LGT). As expected, the base compositional approach does not recognize some number of acquired genes, such as those obtained from organisms of similar genomic G + C contents (black portion of bars). Taken as a whole, nearly 25% of the 4,280 protein-coding genes in this *Escherichia coli* lineage were introduced by

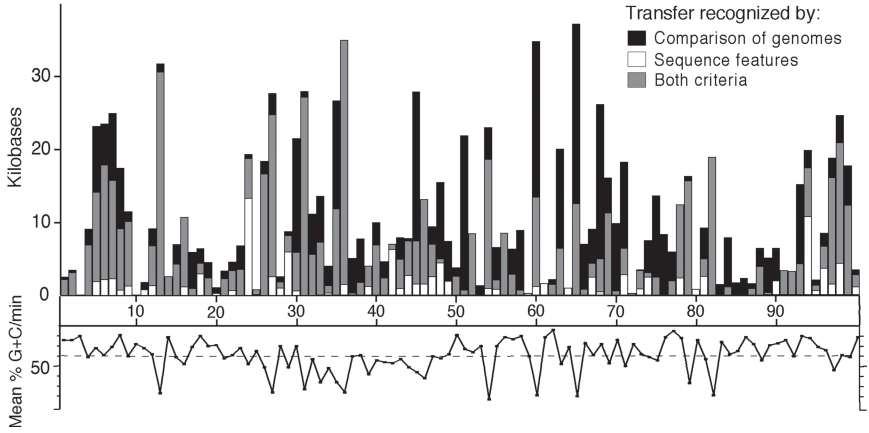


FIGURE 12.1 Linear representation of the *E. coli* MG1655 chromosome showing the distribution of horizontally acquired DNA. At each minute (1–100), vertical bars depict the amount of horizontally acquired, protein-coding DNA, as inferred by two methods: (*i*) atypical sequence features (i.e., base composition) (white) and (*ii*) the unique occurrence of a gene in *E. coli* after aligning and comparing the genome sequences of *E. coli*, *Salmonella enterica*, and *Klebsiella pneumonia* (black). Gray portions of vertical bars denote the overlap between the methods and the amount of protein-coding DNA in *E. coli* inferred to be horizontally acquired based both on its sequence features and on its phylogenetic distribution. Along the bottom of the figure is shown the base composition (% G/C) computed in discrete windows for each minute of the chromosome. The dashed horizontal line shows the overall average base composition for all protein-coding genes in this genome (51.0% G/C). Figure modified from Lawrence and Ochman (2002).

LGT because it split from *Salmonella enterica* an estimated 100 million years ago. And the amount of acquired DNA detected in *E. coli*, as inferred from the base compositional features of genes, seems to be about average for a genome of its size (Garcia-Vallve *et al.*, 2000; Ochman *et al.*, 2000).

Gauging the proportion of acquired genes within a bacterial genome by evaluating its compositional features has some distinct advantages: it is computationally simple and does not rely on the availability of any other genomes. But this method divulges predominantly one class of acquired sequences, i.e., unique genes obtained from very divergent sources, and might vastly underestimate the full extent of LGT-affecting bacterial genomes. Gene exchange can also occur between close relatives and/or between genes that are conserved among organisms. Such events result in an exceptionally high degree of similarity between genes from different

taxa and are usually uncovered by some type of comparative method. For example, genomes are surveyed regularly for genes whose best match (as detected by BLAST) lie outside their closest sequenced relatives, and in the case of certain sequenced bacteria (e.g., *Thermotoga maritima* and *Aquifex aeolicus*), substantial fractions of their genes were found to be most similar to genes present in Archaea (Deckert *et al.*, 1998; Logsdon and Faguy, 1999; Nelson *et al.*, 1999).

Because LGT will result in different phylogenies for different portions of the genome, the most common and robust way to identify cases of transfer and exchange is by searching for evidence of discordance among gene trees. This approach has been applied from the deepest to the shallowest phylogenetic levels, and thousands of transfer events have been recognized (Boucher *et al.*, 2003; Koonin *et al.*, 2001). The availability of full genome sequences has allowed the evaluation of the history of the genes distributed among all life forms, which might be thought to be highly constrained and less susceptible to replacement by LGT. Many of these genes, even ribosomal RNA, long touted and applied as the benchmark for determining organismal relationships, show evidence of LGT over some portion of their evolutionary history (Yap *et al.*, 1999).

THE COHESION OF BACTERIAL GENOMES

So far, these studies confirm that LGT is pervasive and is an ongoing process within bacterial genomes. Genes with sporadic distributions and atypical sequence features arise by LGT, and there are clear cases of gene transfer occurring at all taxonomic levels, even among the genes common to all life forms. With the potential for the LGT of any gene and among all organisms, bacterial species and other taxonomic groupings might not be definable entities. Thus, there is a need to establish whether LGT is resorting the genes in bacterial genomes, eradicating the vestiges of bacterial species, and confounding attempts at phylogenetic classification.

To assess the extent to which LGT is linked phylogenetic disruption, we considered the relationship between DNA acquisition and phylogenetic incongruence in fully sequenced bacteria at several taxonomic levels, including that occurring within species (*E. coli*, *Chlamydophila pneumoniae*, and *Staphylococcus aureus*), within genera (*Escherichia*, *Salmonella*, *Buchnera*, and *Streptococcus*), and within families (Enterobacteriaceae and Rhizobiaceae) (Daubin *et al.*, 2003b). We focused on groups at these phylogenetic depths both to assure substantial overlaps in genome contents and to minimize the risk of reconstruction artifacts due to hidden paralogy or long-branch attraction. And for each group of four genomes, we inferred both the number of recently acquired and lost genes (based on their phylogenetic distributions) and the proportion of ortholog phy-

logenies supporting lateral transfers (by asking whether an alignment significantly supports the rRNA reference topology, either of the two alternate topologies, or no topology).

For all groups and at all taxonomic levels, the proportion of ortholog phylogenies supporting a hypothesis of LGT is always small and often zero (Fig. 12.2). Only among members of the same species (*E. coli* or *C. pneumoniae*) were there increased levels of LGT (i.e., where >5% of alignments support a tree other than the rRNA reference topology). In contrast to the rarity of exchange among orthologs, levels of gene acquisition (as determined from the phylogenetic occurrence of genes) remain high, as previously inferred from the compositional features of individual genomes.

Thus, gene acquisition is frequent but gene replacement is relatively rare, resulting in fundamentally two classes of proteincoding sequences within bacterial genomes: first are the orthologs that are conserved among taxa and not prone to gene transfer and exchange among species. Next are the acquired genes, which are generally unique to a genome and, unlike orthologs, encode proteins of uncharacterized functions. So despite high levels of LGT, bacteria seem to form coherent groups at the shallower taxonomic levels because LGT is concentrated in a class of genes that are not suitable candidates for phylogenetic analysis (Daubin *et al.*, 2003b).

DETERMINANTS OF GENE EXCHANGE IN BACTERIAL SPECIES

Despite the massive influx of new genes into bacterial genomes, the only taxonomic group for which appreciable amounts of homologous recombination were detected was within bacterial species. This finding is remarkably similar to the concept of species that is applied to sexually reproducing eukaryotes, i.e., groups of organisms that exchange genes. But assuming that there is the potential for any sequence to be transferred among bacteria, what factors abide the integration and exchange of homologs from some sources and prevent those from others?

The extent of homologous exchange, as indexed by multilocus enzyme electrophoresis and by multilocus sequence typing, has been shown to vary enormously across bacterial species (Feil and Spratt, 2001; Selander and Musser, 1990), making it unlikely that a single mechanism regulates recombination efficiency in all bacteria. However, the process has been analyzed in some detail in *E. coli* and *Salmonella typhimurium* (*Salmonella enterica* serovar Typhimurium), in which gene exchange depends on the degree of similarity between donor and recipient sequences. Homologous genes from *E. coli* and *Salmonella typhimurium* differ by ≈15% in sequence, and recombination rates, as assayed in conjugal matings, are $\approx 10^5$ lower for intergeneric than for intraspecies crosses (Baron *et al.*, 1968; Matic *et*

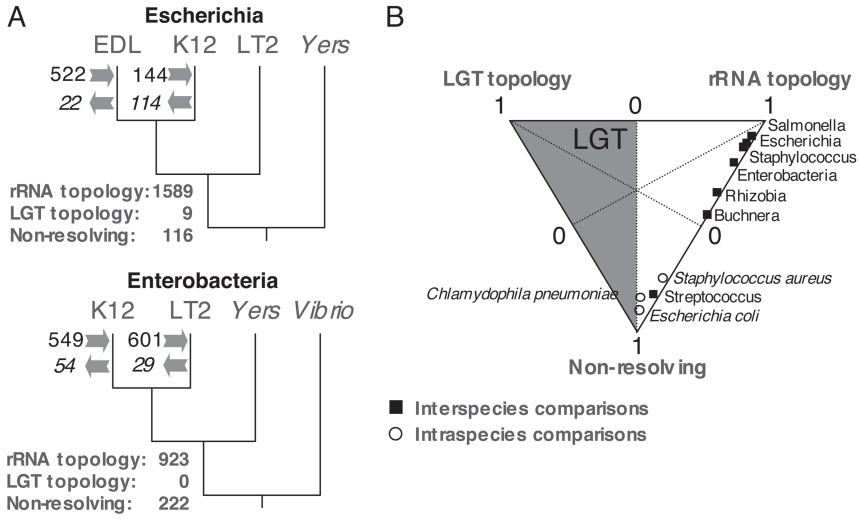


FIGURE 12.2 Phylogenetic inference of cohesion within bacterial genomes. (*A*) Relationship between gene acquisition and loss and the amount of phylogenetic incongruence observed in fully sequenced bacterial genomes. For each quartet of genomes, we inferred the number of recently acquired and lost genes (shown at arrows on the corresponding branches). In addition, for each quartet of genomes, orthologous genes were inferred, aligned, and evaluated at the nucleic sequence level based on the Shimodaira–Hasegawa test implemented in Puzzle 5.1 (Shimodaira and Hasegawa, 1999; Strimmer and von Haeseler, 1996). Shown are the numbers of orthologs supporting each category of alignment: (*i*) those supporting the reference phylogeny (rRNA topology), (*ii*) those supporting either alternate phylogeny (LGT topology), and (*iii*) those with no statistical support for any phylogeny (nonresolving). (*B*) Relative frequencies of the three categories of alignments in diverse bacterial groups at several taxonomic levels. The shaded zone of this plot represents the area where LGT predominates. Only for intraspecies comparisons within *E. coli* and *C. pneumoniae* are the frequencies of LGT + 5%. Data and figure from Daubin *et al.* (2003b).

al., 1995). This barrier to gene exchange is effected, in part, by mismatch repair enzymes, which inhibit recombination between divergent sequences, thereby allowing gene exchange among close relatives and preventing it among more distant strains (Matic *et al.*, 1996; Rayssiguier *et al.*, 1989; Vulic *et al.*, 1997).

If similar mechanisms that limit homologous recombination are operating in other taxa, then bacterial species can be viewed as assemblages of lineages that are sufficiently closely related to potentially exchange shared genes. Then, depending on the actual rates of recombination, population

structure, and patterns of lineage extinction, these assemblages will eventually assort into distinct species that have diverged sufficiently at the DNA level to form a genetic barrier to gene exchange. In this case, the practice of delineating bacterial species on the basis of some prescribed level of sequence divergence seems to be well justified.

WHY SPECIES?

There is still an overarching issue, which stems from the fact that all of the genetic and genomic properties discussed so far were characterized in groups of lineages that were already designated as distinct bacterial species. *E. coli* and *Salmonella typhimurium* were each discovered more than a century ago, and their classification is founded on schemes devised before there was any knowledge of genes or genetics.

Bacterial species are typically recognized according to their cellular properties and metabolic capabilities; for example, *E. coli*, a mammalian commensal, ferments lactose but not citrate, whereas *Salmonella enterica*, a mammalian pathogen, is lactose negative and citrate positive. Examining the genetic basis of these traits, the *lac* operon is a G + C-rich region unique to *E. coli*, whereas the citrate utilization (as well as many *Salmonella* virulence determinants) is conferred by low G + C genes present only in *Salmonella*. Therefore, assignment of isolates to each of these bacterial species seems to have been largely on traits that were introduced by LGT [but see Stoebel (2005) for a contrasting view of *lac* operon evolution]. And consequently, the species so defined have turned out to be discrete biological entities that, because of genetic and mechanistic reasons, rarely exchange homologs.

To determine how horizontally acquired genes are able to accurately define bacterial species, we need to trace the phylogenetic history of genes that occur sporadically among multiple taxa. To accomplish this, it is necessary to step back from *E. coli* and *Salmonella enterica* (where genes are either confined to one, or present in both, species) and consider families of genes within the broader taxonomic framework that subsumes these lineages. We examined the full protein-coding gene repertoires within 13 sequenced Gammaproteobacteria, including one strain each of *E. coli* and *Salmonella enterica* (Lerat *et al.*, 2005). Previously it was shown that >99% of the 205 single-copy genes that are shared by all genomes supported the same relationships for the 13 species examined (Lerat *et al.*, 2003), thereby providing a robust organismal phylogeny against which the trees based on less conserved genes can be tested (Fig. 12.3*A*).

Considering single-copy genes that are absent from one of more of these genomes (i.e., those whose distributions may result from LGT, gene loss, or some combination of these processes), we found that very few

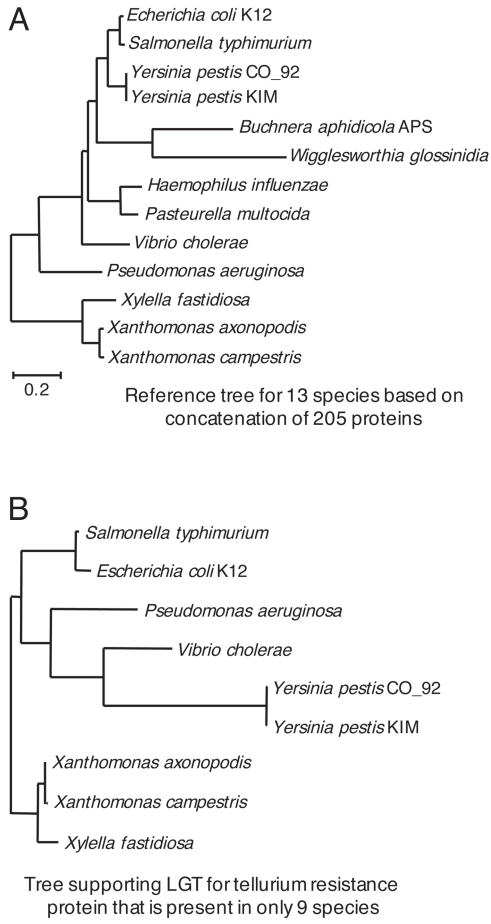


FIGURE 12.3 Example of a lateral gene transfer detected by phylogenetic discordance. (*A*) Neighbor-joining tree based on the concatenation of 205 single-copy genes common to all 13 Gammaproteobacterial species. Note that 203 of the 205 genes individually supported the same topology. Figure adapted from Lerat *et al.* (2003). (*B*) Example of a tree that conflicts with the reference topology. Among the small proportion of proteins showing statistical support for an alternate topology was tellurium resistance protein. Homologs of the gene encoding this protein have been detected in only 9 of the 13 species. Note that the topology of this tree departs from that of the reference tree (depicted in *A*) because of a single LGT event that occurred in the ancestor to *E. coli* and *Salmonella enterica*. Data in *B* are from Lerat *et al.* (2005).

display statistically supported incongruence with the organismal phylogeny (Table 12.1). For genes present in a single copy in only a subset of the 13 genomes, the incidence of LGT is very low and not significantly different from that observed for the 205 single-copy genes present in all species. And furthermore, those few cases of LGT can usually be accounted for by a single event, an example of which is shown in Fig. 12.3*B*.

Although LGT has been the major source of new genes in these bacterial lineages, as reflected by the large number of gene families restricted to one or two genomes (i.e., 10,728 of 14,158 total families), the lack of phylogenetic inconsistencies among the sporadically distributed genes reveals that (*i*) acquired genes come from sources outside of this group, and (*ii*) subsequent to their initial acquisition, genes are by and large transmitted vertically. These findings explain why properties introduced by LGT can serve as stable markers of bacterial species and of phylogenetics. First, genes acquired from distant sources are more likely to supply a novel trait that would set the recipient apart from its relatives. Next, those acquired genes that confer a useful (and defining) trait will persist within the descendant clade and only rarely be transferred laterally to related species.

A BACTERIAL SAGA (SPECIATION ATTRIBUTABLE TO GENE ACQUISITION)

Taken as a whole, the effects of gene transfer and exchange on bacterial evolution and classification are opposite, or at least orthogonal, to what one might anticipate. High levels of gene transfer should, in the words of Gogarten *et al.* (2002) ''obliterate the patterns of vertical descent'' and erase the boundaries between species or any other taxonomic units. But despite massive amounts of LGT, bacteria seem to form more or less cohesive groups at many taxonomic levels (Daubin *et al.*, 2003b; Kurland *et al.*, 2003). These groupings are the result of a nonarbitrary process of gene acquisition in which divergent organisms serve as a persistent source of novel genes in a genome, and the levels of recombinational exchange among homologs shared by related species are low.

As such, LGT can sometimes be viewed as an agent that promotes and maintains bacterial species (Lawrence, 2002; Lawrence and Ochman, 1998). Acquired genes play a major role in bacterial diversification by supplying previously unavailable traits, which can allow the rapid exploitation of new environments. Such capabilities, which are strictly vertically transmitted once they are acquired, can serve to subdivide the population, allowing the phenotypically distinct lineages to diverge at the sequence level to the point where there is a recombinational barrier to gene exchange. Although our saga has been based largely on the analyses

TABLE 12.1 Incidence of Lateral Gene Transfer Among Single-Copy Genes with Different Phylogenetic Distributions

No. of species	6	7	8	9	10	11	12	13
No. of single-copy genes	167	188	137	78	69	109	98	205
No. rejecting reference tree	5	2	3	3	1	3	5	2
Trees supporting LGT, %	2.99	1.06	2.19	3.85	1.45	2.80	5.10	0.98

Only those genes whose homologs are present in 6–13 of the species are shown. Orthologs were aligned and evaluated to recover topologies that differed significantly from the reference tree.

of a single taxon, these results show that it is still possible to make inferences about the origin and nature of bacterial species in light of substantial lateral gene transfer.

ACKNOWLEDGMENTS

We thank Nancy Moran and two anonymous reviewers for comments on the manuscript and Becky Nankivell for preparing the figures. Department of Energy Grant DEFG0301ER6314 and National Institutes of Health Grant GM56120 supported much of the work presented in this article.

REFERENCES

Baron, L. S., Gemski, P., Johnson, E. M. & Wohlhieter, J. A. (1968) Intergeneric bacterial matings. *Bacteriol. Rev.* **32,** 362–369.

Boucher, Y., Douady, C. J., Papke, R. T., Walsh, D. A., Boudreau, M. E., Nesbo, C. L., Case, R. J. & Doolittle, W. F. (2003) Lateral gene transfer and the origins of prokaryotic groups. *Annu. Rev. Genet.* **37,** 283–328.

Cohan, F. M. (2002) What are bacterial species? *Annu. Rev. Microbiol.* **56,** 457–487.

Daubin, V., Lerat, E. & Perriere, G. (2003a) The source of laterally transferred genes in bacterial genomes. *Genome Biol.* **4,** R57.

Daubin, V., Moran, N. A. & Ochman, H. (2003b) Phylogenetics and the cohesion of bacterial genomes. *Science* **301,** 829–832.

Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., *et al.* (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392,** 353–358.

Feil, E. J. & Spratt, B. G. (2001) Recombination and the population structures of bacterial pathogens. *Annu. Rev. Microbiol.* **55,** 561–590.

Garcia-Vallve, S., Romeu, A. & Palau, J. (2000) Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res.* **10,** 1719–1725.

Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. (2002) Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* **19,** 2226–2238.

Guerdoux-Jamet, P., Henaut, A., Nitschke, P., Risler, J. L. & Danchin, A. (1997) Using codon usage to predict genes origin: Is the *Escherichia coli* outer membrane a patchwork of products from different genomes? *DNA Res.* **4,** 257–265.

Koonin, E. V., Makarova, K. S. & Aravind, L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* **55,** 709–742.

Kurland, C. G., Canback, B. & Berg, O. G. (2003) Horizontal gene transfer: A critical view. *Proc. Natl. Acad. Sci. USA* **100,** 9658–9662.

Lan, R. & Reeves, P. R. (2001) Comparison of two major forms of the *Shigella* virulence plasmid pINV: positive selection is a major force driving the divergence. *Trends Microbiol.* **9,** 419–424.

Lawrence, J. G. (2002) Gene transfer in bacteria: speciation without species? *Theor. Popul. Biol.* **61,** 449–460.

Lawrence, J. G. & Ochman, H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA* **95,** 9413–9417.

Lawrence, J. G. & Ochman, H. (2002) Reconciling the many faces of lateral gene transfer. *Trends Microbiol.* **10,** 1–4.

Lerat, E., Daubin, V. & Moran, N. A. (2003) From gene trees to organismal phylogeny in prokaryotes: The case of the gamma-Proteobacteria. *PLoS Biol.* **1,** E19.

Lerat, E., Daubin V., Ochman, H. & Moran, N. A. (2005) Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.*, in press.

Logsdon, J. M. & Faguy, D. M. (1999) *Thermotoga* heats up lateral gene transfer. *Curr. Biol.* **9,** R747–R751.

Matic, I., Rayssiguier, C. & Radman, M. (1995) Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. *Cell* **80,** 507–515.

Matic, I., Taddei, F. & Radman, M. (1996) Genetic barriers among bacteria. *Trends Microbiol.* **4,** 69–72.

Medigue, C., Rouxel, T., Vigier, P., Henaut, A. & Danchin, A. (1991) Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J. Mol. Biol.* **222,** 851–856.

Muto, A. & Osawa, S. (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl. Acad. Sci. USA* **84,** 166–169.

Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson J. D., Nelson W. C., Ketchum K. A., *et al.* (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399,** 323–329.

Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405,** 299–304.

Ragan, M. A. (2001) On surrogate methods for detecting lateral gene transfer. *FEMS Microbiol. Lett.* **201,** 187–191.

Rayssiguier, C., Thaler, D. S. & Radman, M. (1989) The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature* **342,** 396–401.

Redfield, R. J. (2001) Do bacteria have sex? *Nat. Rev. Genet.* **2,** 634–639.

Rolfe, R. & Meselson, M. (1959) The relative homogeneity of microbial DNA. *Proc. Natl. Acad. Sci. USA* **45,** 1039–1042.

Rossello-Mora, R. & Amann, R. (2001) The species concept for prokaryotes. *FEMS Microbiol. Rev.* **25,** 39–67.

Selander, R. K. & Musser, J. M. (1990) Population genetics of bacterial pathogenesis. In *The Evolution of Bacterial Pathogens* (Academic, New York) Vol. XI, pp. 11–36.

Shimodaira, H. & Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16,** 1114–1116.

Stoebel, D. M. (2005) Lack of evidence for horizontal transfer of the lac operon into *Escherichia coli*. *Mol. Biol. Evol.* **22,** 683–690.

Strimmer, K. & von Haeseler, A. (1996) Quartet puzzling: A quartet maximum-liklihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13,** 964–969.

Sueoka, N. (1961) Variation and heterogeneity of base composition of deoxyribonucleic acids: a compilation of old and new data. *J. Mol. Biol.* **3,** 31–40.

Sueoka, N. (1962) On the genetic basis of variation and heterogeneity of DNA base composition. *Proc. Natl. Acad. Sci. USA* **48,** 582–592.

Sueoka, N. (1988) Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci. USA* **85,** 2653–2657.

Sueoka, N., Marmur, J. & Doty, P. (1959) Heterogeneity in deoxyribonucleic acids. II. Dependence of the density of deoxyribonucleic acids on guanine-cytosine. *Nature* **183,** 1429–1431.

Vulic, M., Dionisio, F., Taddei, F. & Radman, M. (1997) Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proc. Natl. Acad. Sci. USA* **94,** 9763–9767.

Yap, W. H., Zhang, Z. & Wang, Y. (1999) Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *J. Bacteriol.* **181,** 5201–5209.

Young, J. M. (2001) Renaming of *Agrobacterium larrymoorei* Bouzar and Jones 2001 as *Rhizobium larrymoorei* (Bouzar and Jones 2001) comb. nov. *Int. J. Syst. Evol. Microbiol.* **51,** 945–953.

13

Ernst Mayr and the Modern Concept of Species

KEVIN DE QUEIROZ*

Ernst Mayr played a central role in the establishment of the general concept of species as metapopulation lineages, and he is the author of one of the most popular of the numerous alternative definitions of the species category. Reconciliation of incompatible species definitions and the development of a unified species concept require rejecting the interpretation of various contingent properties of metapopulation lineages, including intrinsic reproductive isolation in Mayr's definition, as necessary properties of species. On the other hand, the general concept of species as metapopulation lineages advocated by Mayr forms the foundation of this reconciliation, which follows from a corollary of that concept also advocated by Mayr: the proposition that the species is a fundamental category of biological organization. Although the general metapopulation lineage species concept and Mayr's popular species definition are commonly confused under the name "the biological species concept," they are more or less clearly distinguished in Mayr's early writings on the subject. Virtually all modern concepts and definitions of the species category, not only those that require intrinsic reproductive isolation, are to be considered biological according to the criterion proposed by Mayr. Definitions of the species category that identify a particular contingent property of metapopulation lineages (including intrin-

*Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560-0162.

sic reproductive isolation) as a necessary property of species reduce the number of metapopulation lineages that are to be recognized taxonomically as species, but they cause conflicts among alternative species definitions and compromise the status of the species as a basic category of biological organization.

Species are one of the fundamental units of comparison in virtually all subfields of biology, from anatomy to behavior, development, ecology, evolution, genetics, molecular biology, paleontology, physiology, and systematics. In large part, the importance of species in biology derives from their importance in systematics, which is responsible for the taxonomic framework used in all branches of biology. Systematics is one of the oldest scientific disciplines and, from its beginning, one of its central concepts has been the concept of species. Systematics can be characterized generally as the branch of science devoted to the study of the different kinds of organisms (biological diversity, in contemporary terms), and the term ''species'' is Latin for ''kind.'' Moreover, systematics, for the last 250 years, has been strongly influenced by the familiar hierarchy of taxonomic categories originating from the work of Carolus Linnaeus (1753, 1758), of which the species is the lowest, and in some sense the most fundamental, of the principal categories (de Queiroz, 1997). According to one major dictionary, it is ''the basic category of biological classification'' (Flexner and Hauck, 1993).

The central role of species in systematics is reinforced by the relationship of systematics to evolutionary biology. Modern systematics continues to become thoroughly integrated with evolutionary biology, and evolutionary biology has, from its inception, granted a central role to species. This situation should be evident from the fact that the most important book in the history of this field, the one that more or less initiated the field itself, is titled *On the Origin of Species* (Darwin, 1859). The central role of species has continued into the more recent history of the discipline, including the period of the Modern Evolutionary Synthesis (Huxley, 1942; Mayr and Provine, 1980), which laid the foundation for much current research in systematics and evolutionary biology. Evidence for the central role of species is provided by the titles of two of the most important publications from this period, both of which highlight species through reference to Darwin's title: Dobzhansky's *Genetics and the Origin of Species* (1937) and Mayr's *Systematics and the Origin of Species* (1942).

In the case of Mayr's (1942) book, the importance of species is also attested to by the fact that one of the most important and enduring influences of this book, along with subsequent repetitions and elaborations (Mayr, 1963, 1970, 1982), concerns its discussion of species concepts, including a proposed definition of the species category that became a text-

book standard. Most contemporary biologists are familiar with the idea that ''species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups'' (Mayr, 1942), now commonly known as the ''biological species definition'' or the ''biological species concept.'' However, despite the important influence that Mayr's writings concerning species had on systematics in particular, and on evolutionary biology in general, and despite the wide adoption of his proposed species definition in textbooks, these contributions did not solve the long-standing problem concerning the nature of species. In fact, Mayr's proposed definition seems to have led to the emergence of new sources of disagreement.

In this paper, I will argue that the reconciliation of alternative and incompatible definitions of the species category is a natural outgrowth of the general concept of species for which Ernst Mayr was one of the primary developers and advocates. On the other hand, this proposed reconciliation is at odds with most contemporary species definitions, including the popular definition proposed by Mayr himself, at least as commonly interpreted. The incompatibility results from treating certain properties acquired by diverging population lineages as necessary properties of species, including potential interbreeding and its counterpart, intrinsic reproductive isolation, in the case of Mayr's definition. Reinterpreting these properties as neither necessary nor sufficient for the definition of the species category eliminates the incompatibilities among alternative concepts of species, resulting in a unified species concept that brings taxonomic practice in line with common claims about the theoretical significance of species, and that is highly consistent with the general concept of species for which Ernst Mayr was arguably the most articulate and prolific advocate.

THE SPECIES PROBLEM

Despite the wide acceptance of Mayr's proposed species definition (and perhaps partly because of it), this definition stimulated critiques as well as the proposal of alternatives. An early critique, including an alternative definition, was published by George Gaylord Simpson (1951, 1961), another leader of the Modern Synthesis (Fitch and Ayala, 1994). However, alternative species concepts did not really begin to proliferate until the 1970s, starting with a paper by Sokal and Crovello (1970), which proposed a phenetic species concept. By the late 1990s, literally dozens of alternatives had been proposed. Mayden (1997), for example, identified 24 named species concepts, including the now-familiar biological, phenetic, evolutionary, ecological, and phylogenetic (three versions) concepts and 16 others. The diversity of contemporary species concepts has been

reviewed in several recent publications (Coyne and Orr, 2004; de Queiroz, 1998; Harrison, 1998; Mayden, 1997) and will not be repeated here. For the present discussion, the important thing to recognize is that different contemporary species concepts are based, in part, on different biological properties. For example, the biological species concept emphasizes the property of reproductive isolation (Dobzhansky, 1970; Mayr, 1942), the ecological species concept emphasizes occupation of a distinct niche or adaptive zone (Andersson, 1990; Van Valen, 1976), one version of the phylogenetic species concept emphasizes diagnosability (Cracraft, 1983; Nixon and Wheeler, 1990) and another, monophyly (Donoghue, 1985; Rosen, 1979). For a more extensive list of properties that form the basis of alternative species concepts, see de Queiroz (1998).

As a consequence of these differences, many alternative contemporary species concepts are incompatible in that they lead to the recognition of different species taxa depending on which concept is adopted. In other words, they lead to different species boundaries and different numbers of recognized species. For example, adopting the diagnosable version of the phylogenetic species concept commonly leads to the recognition of many more species taxa than adopting the biological species concept (Cracraft, 1983; Martin, 1996; Zink, 1996). The existence of alternative. and at least partially incompatible, definitions of the species category, hereafter referred to as the ''species problem,'' creates difficulties given that species are used as basic units of comparison in diverse types of studies. On the one hand, species taxa recognized according to different species concepts often will not be comparable to one another with regard to the biological properties they possess. On the other hand, a study that uses species taxa based on a single species concept may yield very different results from one that uses species taxa based on a different species concept. This is not to deny that particular concepts are preferred by particular groups of biologists. Some such groups argue passionately about the superiority of their preferred concept over the alternatives. However, other groups argue just as passionately in favor of different species concepts. In addition, the species problem seems to be getting worse rather than better, which is to say the number of alternative species concepts has been growing rather than diminishing. Moreover, judging by the increasing numbers of critiques and proposed alternatives, Mayr's species definition, although still perhaps the most widely adopted, seems to be less popular now than 20–30 years ago.

The existence of diverse species concepts is not altogether unexpected, because different concepts are based on properties that are of greatest interest to different subgroups of biologists (de Queiroz, 1998). For example, biologists who study hybrid zones tend to emphasize reproductive barriers, whereas systematists tend to emphasize diagnosability and monophyly, and ecologists tend to emphasize niche differences. Paleon-

tologists and museum taxonomists tend to emphasize morphological differences, and population geneticists and molecular systematists tend to emphasize genetic ones. Nevertheless, for those biologists who are able to set aside their own personal investments and research interests, all of the concepts seem to have some merit. It is certainly the case that all are based on important biological properties.

THE METAPOPULATION LINEAGE CONCEPT OF SPECIES

The reconciliation of alternative and incompatible species concepts derives from the recognition of a more general concept of species that is shared by all contemporary species concepts and definitions (de Queiroz, 1998, 1999, 2005). This general concept of species originated at least as early as the beginning of the 20th century [Mayr (1955, 1982) cited papers by Jordan (1896, 1905) and Poulton (1903) as early examples], but it became well established during the period of the Modern Evolutionary Synthesis (Huxley, 1942; Mayr and Provine, 1980), through the writings of the great leaders of that movement, including Sewall Wright, Theodosius Dobzhansky, George Gaylord Simpson, and particularly Ernst Mayr. All modern species concepts and definitions conform to this general species concept and can therefore be considered variants of it. This general species concept, not Mayr's more restricted species definition, is the true biological species concept (see below).

Several influential discussions of the general species concept that became established during the Modern Synthesis emphasized the correspondence of species with metapopulations or gene pools. Species were equated with groups of interconnected populations that form an extended reproductive community and an unevenly distributed but unitary gene pool or field for gene recombination. The equation of species with metapopulations or gene pools is evident in a number of species definitions from the period of the Modern Synthesis, including those proposed by several of the most influential contributors to that movement. Thus, according to Sewall Wright (1940) ''[species are] groups within which all subdivisions interbreed sufficiently freely to form intergrading populations wherever they come in contact, but between which there is so little interbreeding that such populations are not found.'' Similarly, according to Theodosius Dobzhansky (1950), ''The biological species is the largest and most inclusive Mendelian population'' (a ''Mendelian population is a reproductive community of sexual and crossfertilizing individuals which share in a common gene pool''). And finally, according to Ernst Mayr (1942), ''Species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups.''

All three of these definitions equate species with metapopulations (sets of connected subpopulations, maximally inclusive populations), and all of them suggest that the limits of species as metapopulations are set in one way or another by the limits of interbreeding (which requires not only mating but also the production of viable and fertile offspring), thus implying sexual reproduction. For cases involving purely asexual reproduction, there are two possibilities regarding species. One possibility is that purely asexual organisms do not form species (Dobzhansky, 1937; Ghiselin, 1997; Hull, 1980). The other possibility is there are processes other than the exchange of genetic material, such as natural selection, that determine the limits of species in purely asexual organisms (Meglitsch, 1954; Templeton, 1989; Van Valen, 1976). Both of these views are consistent with the equation of species with metapopulations. Either asexual organisms do not form metapopulations, and therefore they do not form species, or they do formmetapopulations (as the result of some process or processes other than interbreeding), and therefore they also form species.

The general metapopulation concept of species is also evident in species definitions that describe species as lineages rather than as populations. For example, according to George Gaylord Simpson (1961), ''[a] species is a lineage (an ancestral-descendant sequence of populations) evolving separately from others and with its own unitary evolutionary role and tendencies.'' Similarly, according to Leigh Van Valen (1976), ''A species is a lineage (or a closely related set of lineages) which occupies an adaptive zone minimally different from that of any other lineage in its range and which evolves separately from all lineages outside its range.'' The reason definitions that characterize species as lineages can be considered to represent the same general species concept as those that characterize species as (meta)populations is there is a close relationship between populations and lineages at the same level of biological organization (Fig. 13.1). On the one hand, populations can be considered to extend through time, in which case a population is equivalent to a lineage. Alternatively, populations can be considered to exist at an instant in time, in which case a population is equivalent to an instantaneous cross section of a lineage, and a lineage corresponds to a continuous series of populations (the ancestral-descendant sequence of populations in Simpson's definition). In either case, this type of lineage (a single line of ancestry and descent) is not the same as a monophyletic group (a group of entities sharing an exclusive common ancestry), which is also commonly referred to as a lineage (although the term ''clade'' is more appropriate).

The important point is that virtually all contemporary definitions of the species category are based on a common general concept of species: the concept of species as (segments of) metapopulation lineages (de Queiroz, 1998, 1999, 2005). Definitions that describe species as popula-

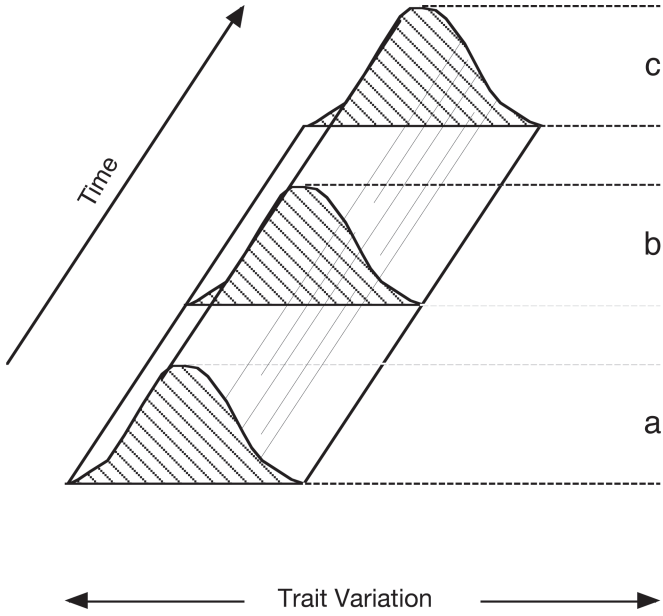


FIGURE 13.1 Populations as lineages (modified from Simpson, 1951). The population or population level lineage at a given instant in time is represented as a frequency distribution in two dimensions [*x* axis, trait variation; *y* axis, frequency (not shown)], whereas the time-extended population or population level lineage is represented by the 3D solid formed by extending the 2D frequency distribution through a third dimension (*z* axis, time). The three shaded distributions (*a*–*c*) represent cross sections of the time-extended population lineage at three different times. The population itself can be interpreted either as the 3D (time-extended) object, in which case it is equivalent to the lineage, or as one of the 2D (time-limited) objects, in which case it is equivalent to a cross section of the lineage.

tions simply view species over some relatively short interval of time, usually the present. In contrast, definitions that describe species as lineages tend to view species over longer time intervals. (In this context, it is not surprising that the first of the lineage definitions was proposed by Simpson, a paleontologist.) Because the population vs. lineage terminology simply reflects a different temporal perspective on entities of the same basic kind, I will hereafter use the term ''metapopulation lineage'' to encompass both views. Virtually all contemporary species definitions (by which I mean those advocated by some contemporary group of biologists) conform to this general metapopulation lineage concept of species (de Queiroz, 1998). They differ with regard to emphasizing (in addition to different temporal perspectives) the theoretical concept itself vs. the em-

pirical evidence and operational procedures that are used to apply it. They also differ with regard to the properties of metapopulation lineages that are considered necessary for those lineages to be regarded as species. Because the metapopulation lineage concept of species is general in the sense that it is shared by all modern concepts and definitions of species, it has previously been referred to as the ''general lineage concept of species'' (de Queiroz, 1998, 1999, 2005), although it might be more accurately termed the ''general metapopulation lineage concept of species.''

THE SPECIES AS A FUNDAMENTAL CATEGORY OF BIOLOGICAL ORGANIZATION

An important corollary of the metapopulation lineage concept of the species is that the species is a fundamental category of biological organization. Although this corollary is now often taken for granted, it is important to recognize that it represents a significant departure from an older view of the species category. Under the older view, the species category was simply a rank in the hierarchy of taxonomic categories. More specifically, the taxa at all levels in the hierarchy were viewed as being of the same basic kind, namely, groups of organisms that shared particular traits (de Queiroz, 1997), but they were assigned to different ranks to indicate differences in relative inclusiveness. Species were included within genera, genera were included within families, and so forth. Thus, species were not viewed as constituting a fundamentally different kind of entity than genera or families; they were just smaller groups separated by smaller degrees of difference. This perspective was held by Darwin (1859), who stated that he viewed ''. . . the term species, as one arbitrarily given for the sake of convenience to a set of individuals closely resembling each other, . . . it does not essentially differ from the termvariety, which is given to less distinct and more fluctuating forms'' and that ''the grades of acquired difference [between taxonomic groups are] marked by the terms varieties, species, genera, families, orders, and classes.'' In short, Darwin viewed the species category as just another categorical rank, in particular, one applied to groups of organisms that differed more than varieties but less than genera.

In contrast, the view of species that emerged during the Modern Synthesis was that species are fundamentally different from the taxa above and below them in the taxonomic hierarchy. Species differ from genera (for example) not only in inclusiveness but also in kind. Species are metapopulation lineages, whereas genera are groups of species sharing a relatively recent common ancestry. Thus, according to Simpson (1961), ". . . there are units in nature that have a special evolutionary status not fully shared with taxa either above or below them in the hierar-

chy. . . . Many of them . . . recognized before Darwin had been called species, and it was inevitable that the term should be transferred to the evolutionary units.'' Similarly, according to Mayr (1969a), ''The unique position of species in the hierarchy of taxonomic categories has been pointed out by many authors. . . . It is the only taxonomic category for which the boundaries between taxa at that level are defined objectively.''

As a fundamental category of biological organization, the species category is roughly analogous to other such categories, including the cell and the organism. In Mayr's (1982) words, '' . . . the species is as important a unit of biology as is the cell at a lower level of integration.'' Species may differ from cells and organisms in terms of the processes responsible for uniting their parts to form larger wholes, and they may exist at a higher level of organization (the population level), but all three categories are fundamentally similar in identifying particular kinds of biological entities that replicate or reproduce in the sense of generating other entities of the same kind. Cells divide to produce new cells, organisms reproduce to generate new organisms, and species speciate to produce new species.

Despite the relatively wide acceptance of the idea that the species represents a fundamental category of biological organization, some of the ways in which biologists continue to treat species taxonomically are inconsistent with that proposition. The practices in question appear to be holdovers from the earlier interpretation of the species category as a rank in the hierarchy of taxonomic categories (i.e., as opposed to its later interpretation as a fundamental category of biological organization). Moreover, those practices also appear to be responsible for the current species problem, that is, for the existence of alternative and at least partially incompatible definitions of the species category.

If the species category is to be truly analogous to the categories cell and organism and of similar general theoretical significance, then the species category must be the most general category at its particular level of biological organization. It cannot be a less general subset of this general category. Consider the analogous category organism. This category includes all living (and formerly living) beings, which are not required to possess any property beyond those used to define the general category, such as being born, or sexually mature, or fully grown, to be considered organisms. Embryos, juveniles, and adults are all considered organisms. Similarly, in the case of cells, once the entities are physically separated, they need not possess any additional property, such as having replicated their DNA or condensed their chromosomes, to be considered cells. Cells at various phases in the cell cycle are all considered cells. The general manner in which the categories cell and organism are conceptualized is what confers on these categories their general theoretical significance (which is greater than that of more restricted categories, such as prophase

TABLE 13.1 Properties, in Addition to Existence as a Separately Evolving Metapopulation Lineage, Commonly Treated as Necessary Properties of Species

Property	Species Concept and/or Definition[a]
Potential interbreeding (intrinsic reproductive isolation)	Biological species concept/definition (isolation species concept) (Mayr, 1942; Dobzhansky, 1970)
Shared specific mate recognition or fertilization system	Recognition species concept (Paterson, 1985)
Same niche or adaptive zone	Ecological species concept (Van Valen, 1976)
Monophyly (as inferred from apomorphy or exclusive coalescence of gene trees)	Monophyly version of the phylogenetic species concept (Rosen, 1979; Donoghue, 1985), genealogical species concept (Baum and Shaw, 1995)
Form a phenetic cluster (quantitative difference)	Phenetic species concept (Sokal and Crovello, 1970; Michener, 1970; Sneath and Sokal, 1973)
Form a diagnosable group (fixed qualitative difference)	Diagnosable version of the phylogenetic species concept (Cracraft, 1983; Nixon and Wheeler, 1990), some interpretations of the evolutionary species concept (Grismer, 1999, 2001)
Form a genotypic cluster	Genotypic cluster species definition (Mallet, 1995)

[a]According to the classification of de Queiroz (1998).

cell and adult organism). In the case of the species, however, most biologists continue to treat this category in a way that is at least partially inconsistent with the general theoretical significance commonly attributed to it. Under most current species definitions, the species is not the most general category at its particular level of organization. The reason is that those definitions commonly require separately evolving metapopulation lineages to possess some additional property before they are considered species.

Table 13.1 lists some of the additional properties that are commonly interpreted as necessary for a separately evolving metapopulation lineage to be considered a species (i.e., as defining properties of the species category). Those properties include intrinsic reproductive isolation, as in Mayr's well known species definition (Mayr, 1942, 1963, 1970); the occupancy of a distinct niche or adaptive zone, as in the ecological species

definition proposed by Van Valen (1976); monophyly, as in definitions proposed by certain phylogenetic systematists (Donoghue, 1985; Rosen, 1979); and a number of others. Hereafter, I will refer to these properties as contingent properties, because they are properties that a metapopulation lineage may or may not acquire during the course of its existence (the longer it persists, the more likely it is to acquire them).

The problem with the interpretation of these contingent properties of metapopulation lineages as necessary properties of species is that it compromises the generality of the species concept. By requiring that a separately evolving metapopulation lineage be intrinsically reproductively isolated, or ecologically differentiated, or monophyletic, or anything else, before it is considered a species, the concept of species is restricted to only some members of the general category at the level of biological organization in question (i.e., to only some metapopulation lineages). As a consequence, the generality of the species concept is restricted. To use analogies from other levels in the hierarchy of biological organization, current definitions of the species category are analogous to considering only those membrane-bound parcels of cytoplasm that have reached the S phase of the cell cycle to be cells or only those living beings that have reached sexual maturity to be organisms.

The interpretation of properties such as intrinsic reproductive isolation, ecological distinctiveness, monophyly, and so forth, as necessary properties of species seems to represent a holdover from the earlier view of the species as a taxonomic rank used to distinguish among groups that differ only in inclusiveness (i.e., as opposed to representing a distinct category of biological organization). The reason is that, in effect, those properties are being used to decide which metapopulation lineages deserve to be ranked as species. Those metapopulation lineages that are separately evolving but have not yet acquired the stipulated property are not considered to merit the rank of species. Instead, they are commonly ranked as subspecies. Thus, even if the species category is being treated as fundamentally different from the genus, the family, and the other higher taxonomic categories, it is still effectively being treated as a different rank rather than a different kind relative to the lower ones, in particular, relative to the subspecies category. In other cases, metapopulation lineages that have not yet acquired the stipulated property are not granted any formal taxonomic recognition whatsoever (i.e., as opposed to being ranked as subspecies). In such cases, the subspecies rank has been eliminated, and the species has effectively become the only rank assigned to metapopulation lineages. But regardless of whether one considers this situation to represent a holdover from the interpretation of the species category as a rank, it still means the species category is not the most general category at its particular level of organization.

THE CAUSE OF THE SPECIES PROBLEM

In addition to restricting the theoretical significance of the species category, the interpretation of various contingent properties of metapopulation lineages as necessary properties of species is also the cause of the species problem. That is, it is the reason for the existence of incompatible alternative definitions of the species category. Because different authors adopt species definitions that treat different contingent properties of metapopulation lineages as necessary properties of species (e.g., intrinsic reproductive isolation, diagnosability, or exclusive coalescence of alleles), those authors commonly disagree about which metapopulation lineages deserve to be ranked as species (de Queiroz, 1998, 1999, 2005).

The reason emphasis on different contingent properties leads to incompatible species definitions is that those properties arise at different times during the process of separation and divergence among metapopulation lineages (i.e., speciation). Lineage separation and divergence can be conceptualized in terms of a few general evolutionary processes: mutation, natural selection, migration (or the lack thereof), and genetic drift. In contrast, the properties affected by those processes are highly diverse. They may be genetic or phenotypic, qualitative or quantitative, selectively advantageous, disadvantageous, or neutral, and they may involve many different aspects of biology, including genetics, development, morphology, physiology, and behavior.

With regard to the species problem, the important point is that the process of evolutionary divergence leads to the acquisition of a number of different properties by diverging lineages, including those that have been emphasized by different groups of biologists in their definitions of the species category. Thus, as lineages diverge, they (or their component organisms) become distinguishable in terms of quantitative traits. They become diagnosable in terms of fixed character states. Their genitalia, gametes, genomes, and developmental systems become incompatible. Their mate-recognition systems diverge to the point where they no longer recognize one another as potential mates. They evolve distinctive ecologies. And they pass through polyphyletic, paraphyletic, and monophyletic stages in terms of their component genes and organisms. These changes commonly do not occur at the same time, and they are not even necessarily expected to occur in a regular order. The problem is that each alternative species definition adopts a different one of these properties as a defining or necessary property of species. This is the reason that the alternative species definitions, despite their general agreement regarding the conceptualization of species as metapopulation lineages, imply different conclusions concerning which lineages deserve to be recognized as species.

The highly simplified diagram in Fig. 13.2 represents the process of metapopulation lineage divergence. The progressive darkening and light-

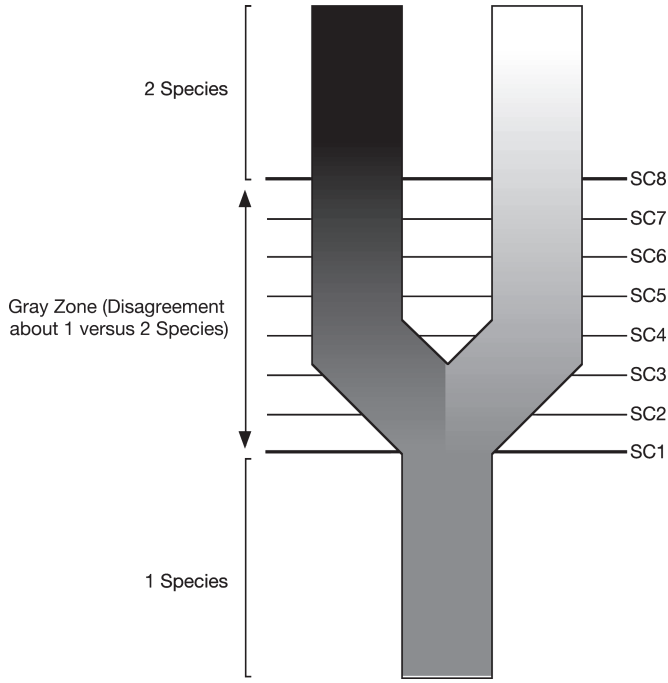


FIGURE 13.2 A highly simplified representation of the process of metapopulation lineage divergence (speciation) illustrating the conflicts caused by adopting different contingent properties of metapopulation lineages as necessary properties of species (modified from de Queiroz, 1998). Progressive darkening and lightening of the daughter lineages represent their progressive divergence through time (bottom to top), and the numbered lines labeled SC (species criterion) 1–8 represent the times at which the daughter lineages acquire different properties relative to one another (e.g., when they become phenetically distinguishable, diagnosable by a fixed character difference, reciprocally monophyletic, reproductively incompatible, ecologically distinct, etc.). Before evolution of the first property (SC1), authors will agree there is a single species, and after evolution of the last property (SC8), they will agree there are two. Between these events, however, there will be disagreement among authors about whether one vs. two species are involved. Those disagreements result from authors adopting different contingent properties (species criteria) as the basis for their species definitions.

ening of the daughter lineages represent their increasing divergence through time, and the numbered lines represent the times at which they acquire different properties relative to one another, for example, when they become phenetically distinguishable, diagnosable, reciprocally monophyletic, reproductively incompatible, ecologically distinct, and so

forth. The entire set of properties forms a broad gray zone where alternative species concepts come into conflict. On either side of the gray zone, there will be unanimous agreement about the number of species: before the acquisition of the first property, everyone will agree there is one species, and after the acquisition of the last property, everyone will agree there are two. In between, however, there will be disagreement. Some people will place the cutoff for treating the diverging lineages as separate species relatively early in the sequence, perhaps where loss or fixation of a character in one of the lineages makes them diagnosable. Others will place the cutoff later, perhaps where the lineages develop an intrinsic reproductive barrier. Still others will place the cutoff later still, perhaps where both lineages form exclusive groups in terms of multiple gene trees. This adoption of different thresholds as criteria for treating diverging lineages as different species is the cause of the species problem. It is the reason for the existence of multiple incompatible definitions of the species category despite widespread agreement about the general nature of species.

A SOLUTION TO THE SPECIES PROBLEM

Both the species problem itself and the discrepancy between the general theoretical significance commonly attributed to species and the treatment of species in taxonomic practice can be solved by making a simple yet fundamental shift in the way species are conceptualized (de Queiroz, 1998, 1999, 2005). This shift is highly consistent with the general concept of species that became established during the Modern Evolutionary Synthesis and for which Ernst Mayr was arguably the most important spokesman. It also represents the more complete acceptance of Mayr's (1982) proposition that the species is one of the fundamental categories of biological organization. The proposed solution retains the element common to all contemporary concepts of species, and it eliminates the conflicts between those rival concepts without denying the importance of the properties that underlie their obvious differences.

The proposal has two components. First, it retains the element common to all contemporary concepts and definitions of species by adopting the general concept of species as separately evolving metapopulation lineages. Second, it eliminates the conflicts among rival concepts by treating this property, existence as a separately evolving metapopulation lineage, as the only necessary property of species. In other words, all of the other properties that have previously been treated as necessary properties of species, which created incompatibilities among alternative species concepts, are reinterpreted as no longer being defining properties of the species category. Instead, they are interpreted as contingent properties not

only of metapopulation lineages but also of species, properties that species as metapopulation lineages may or may not acquire during the course of their existence. In other words, metapopulation lineages do not have to be phenetically distinguishable, or diagnosable, or monophyletic, or reproductively isolated, or ecologically divergent, to be species. They only have to be evolving separately from other such lineages. Because the interpretation of various secondary properties of lineages as necessary properties of species is the cause of the incompatibilities among alternative species concepts, their reinterpretation as contingent rather than necessary properties also removes the incompatibilities. The result is a single, general, unified concept of species.

The reason the resulting species concept can be considered unified is that it does not deny the importance of any of the properties that have been emphasized in previous definitions of the species category. Under a general and unified concept of species, the various contingent properties, although no longer treated as necessary properties of species, remain important in two ways. First, they continue to serve as important lines of evidence relevant to assessing the separation of metapopulation lineages. Indeed, the properties in question (e.g., phenetic distinguishability, reciprocal monophyly, pre- and postzygotic reproductive isolation, fixed character state differences, etc.) are among the best lines of evidence regarding the separation of metapopulation lineages. Second, the various contingent properties can be used to define subcategories of the general species category, that is, to recognize different classes of species based on the properties possessed by those entities. Just as different subcategories of the general category organism are recognized based on properties possessed by organisms (e.g., sexually mature organisms, fully grown organisms, socially dominant organisms, etc.), similarly, different subcategories of the general category species can be recognized based on properties possessed by species (e.g., diagnosable species, reproductively isolated species, monophyletic species, etc.). Thus, under a general and unified species concept, all of the properties that have been considered important by previous authors remain important for determining the numbers and boundaries of species, and they take on new importance in identifying those species most relevant to addressing particular questions. The main difference is they are no longer treated as necessary properties of species.

Another beneficial consequence of this proposal is that it removes the inconsistency between the proposition that the species is a fundamental category of biological organization and the way in which species are treated taxonomically. Under the general and unified species concept described above, the species would be the most general category at its particular level of biological organization. Consequently, species would be more directly analogous to the members of other fundamental categories

of biological organization, such as cells and organisms. Just as living beings need not (for example) be born, or sexually mature, or fully grown to be considered organisms, metapopulation lineages would not (for example) have to be diagnosable by fixed character differences, or monophyletic, or intrinsically reproductively isolated to be considered species. In other words, all separately evolving metapopulation lineages would be species (de Queiroz, 2005). Reinterpreting the properties in question as contingent rather than necessary properties of species would thus increase consistency between taxonomic practice and common assertions about the general theoretical significance of species.

ERNST MAYR AND THE MODERN CONCEPT OF SPECIES

The proposed resolution of the conflicts among alternative definitions of the species category described above is at odds with the common interpretation of Ernst Mayr's popular species definition, which treats intrinsic reproductive isolation as a necessary property of species. Nevertheless, the proposal is highly compatible with, and might even be considered the culmination of, the general metapopulation lineage concept of species for which Ernst Mayr was arguably the most important spokesman. Evidence for the greater theoretical significance of the general metapopulation lineage concept of species relative to Mayr's concise species definition can be found in at least three components of Mayr's own writings on species: First, the reason he used the adjective ''biological'' to describe his species concept and definition; second, the properties of species Mayr viewed as important for distinguishing the new biological concept he advocated from older species concepts; and third, the distinction Mayr made between species concepts and species definitions, particularly in his early writings, that is, before his views were challenged by a proliferation of alternatives.

Mayr's choice of the adjective ''biological'' as in the terms ''biological species definition'' and ''biological species concept'' has sometimes been criticized for being overly general (Simpson, 1961; Van Valen, 1976), thus raising the suspicion that it was chosen more for its rhetorical value than for its descriptive accuracy. However, an examination of Mayr's writings on species reveals he had good reason for selecting this adjective. According to Mayr (1969a,b), ''This species concept is called biological not because it deals with biological taxa, but because the definition is biological. It utilizes criteria that are meaningless as far as the inanimate world is concerned.'' The important idea for Mayr was that earlier concepts of species were based on properties, such as degree of difference, that could be applied just as easily to inanimate objects as to living things. Linnaeus (1766–1768), for example, recognized species not only of plants and ani-

mals but also of rocks and minerals. In contrast, a truly biological concept of species must be based on properties that are unique to biological systems, properties such as reproduction and interbreeding.

It should be noted, however, that all contemporary species definitions (i.e., all definitions based on the general conceptualization of species as metapopulation lineages) are biological in the sense just described. The reason is that inanimate objects such as rocks and minerals lack reproduction and thus do not form populations or lineages in the uniquely biological way that organisms do. Moreover, with regard to the contingent properties commonly adopted as necessary properties of species, potential interbreeding (and its counterpart, intrinsic reproductive isolation) are no more biological than are a number of alternative properties, such as mate recognition, monophyly, and heterozygote deficits, all of which are also unique to biological systems. Thus, although it is appropriate to use the term ''biological species concept'' for the general concept of species as metapopulation lineages adopted by Ernst Mayr and virtually all other contemporary biologists, it is misleading to use this term for Mayr's proposed species definition, that is, for the idea that a metapopulation lineage is not a species until it has become intrinsically reproductively isolated from all other such lineages.

Equation of the term ''biological species concept'' with the general concept of species as metapopulation lineages rather than the specific criterion of intrinsic reproductive isolation is supported by Mayr's own writings, in particular, by the properties he identified as being important for distinguishing the concept of species in its newer and uniquely biological sense from older conceptualizations of the species category. According to Mayr (1963, 1970), there are three properties that ''raise the species above the typological interpretation of a 'class of objects''' and represent ''one of the earliest manifestations of the emancipation of biology from an inappropriate philosophy based on the phenomena of inanimate nature.'' These properties are first, that the members of a species constitute a reproductive community; second, that a species is an ecological unit that interacts with other species in its environment; and third, that a species is a genetic unit consisting of a large intercommunicating gene pool. These three properties are very general ones that apply to species (at least those composed of sexually reproducing organisms) under all definitions that conform to the general metapopulation lineage concept, that is, not only to those species recognized on the basis of the additional criterion of intrinsic reproductive isolation. One might even question whether potentially (rather than actually) interbreeding organisms are part of the same reproductive community and intercommunicating gene pool (Hull, 1965).

Further support for the equation of the term ''biological species con-

cept'' with the general metapopulation lineage concept of species, rather than a definition based on the property of intrinsic reproductive isolation, can be found in the terminology Mayr used to distinguish among these distinct ideas, at least in his early writings on the species. In those writings, Mayr distinguished more or less clearly between the general biological or metapopulation lineage concept of species and his attempt to describe that concept with a concise definition. Thus, in *Systematics and the Origin of Species* (1942), Mayr referred to the general concept as ''the new species concept,'' and he referred to his proposed definition as ''a biological species definition'' (not ''the biological species definition'') (1942, p. 120). He also called it ''a practical species definition,'' emphasizing its utility for the practicing taxonomist and implying it was a compromise between theoretical and practical considerations (1942, p. 120). Later, in *Animal Species and Evolution* (1963), Mayr stated, ''A study of all of the species definitions published in recent years indicates that they are based on three theoretical concepts, neither more nor less'' (Mayr, 1963, p. 16). He called these three concepts ''the typological species concept'' (used to refer to older species concepts that could be applied to inanimate objects), ''the nondimensional species concept'' (for the concept adopted by naturalists working at a single time and place), and the ''interbreeding-population concept'' (for the concept he advocated). He referred to concise descriptions proposed by both Dobzhansky and himself as ''biological species definitions'' (Mayr, 1963, pp. 19–20), apparently viewing both as falling under the general interbreeding-population concept.

In Mayr's later writings (Grismer, 2001; Mayr, 1969a; Simpson, 1951), the nondimensional species concept was replaced with the nominalistic species concept (used to refer to the view that species are mental constructs invented to permit reference to several individuals collectively), and the terminterbreeding-population concept was replaced with biological species concept, thus demonstrating the equivalence between these terms. However, in these later works, Mayr's own concise definition was presented as if it followed more or less directly from the general concept (''The species definition which results from this theoretical species concept is: [Mayr's definition]''), and Dobzhansky's definition was no longer mentioned. Subsequently, particularly in the writings of other authors (both pro and con), the distinction between the general biological (interbreeding-population) concept of species and Mayr's species definition became further obscured by common reference to both ideas as the biological species concept.

In sum, there is an important distinction between the general concept of species as metapopulations or metapopulation lineages (the true biological species concept) and Ernst Mayr's concise species definition. The

former is a very general theoretical concept that underlies virtually all modern views on species, including all contemporary species definitions. The latter describes a more restricted concept that uses the property of intrinsic reproductive isolation to facilitate taxonomic decisions concerning which metapopulation lineages are to be recognized as species (a practice that retains elements of an older view of the species category as a taxonomic rank and prevents full acceptance of the proposition that the species is a fundamental category of biological organization). Although this distinction is fairly clear in Mayr's early writings, it has become obscured in the recent literature on species concepts. In any case, the concise definition with its treatment of intrinsic reproductive isolation as a necessary property of species is an important part of the species problem (the existence of alternative and partially incompatible definitions of the species category), whereas the general theoretical concept of species as metapopulation lineages forms the basis of a solution to this problem that results in a unified concept of species.

CONCLUSION

Ernst Mayr is almost certainly the greatest of all biologists in terms of his contributions to the development and acceptance of modern views on species. However, with regard to theoretical advances and their practical consequences, his most important contribution in this area was not his widely adopted definition of species but rather the major role he played in the development and advocacy of the general metapopulation lineage concept of species. This contribution had tremendous significance both for systematics in particular and for biology in general. It represented a fundamental shift in the conceptualization of the species category that resulted in a uniquely biological concept of species and changed the species category from a more-or-less arbitrary rank in the hierarchy of taxonomic categories to a basic category of biological organization. Moreover, because this important change in the conceptualization of the species category still has not been fully accepted, it continues to have important consequences. For example, as discussed in this paper, its more complete acceptance provides a simple solution to the species problem, and this solution, in turn, brings the way in which species are treated in taxonomic practice into line with claims about the general theoretical significance of the species category. In sum, Ernst Mayr's ideas had tremendous importance, among many other things, for the development and acceptance of the modern metapopulation lineage concept of species, and they continue to provide the foundation for advances regarding the theoretical concept of species and its practical application.

ACKNOWLEDGMENTS

I thank Jody Hey, Walter Fitch, and Francisco Ayala for the invitation to present this paper at the Colloquium "Systematics and the Origin of Species: On Ernst Mayr's 100th Anniversary." Jody Hey provided valuable comments on the original manuscript.

REFERENCES

Andersson, L. (1990) The driving force: Species concepts and ecology. *Taxon* **39,** 375–382.

Baum, D. A. & Shaw, K. L. (1995) Genealogical perspectives on the species problem. In *Experimental and Molecular Approaches to Plant Biosystematics*, eds. Hoch, P. C. & Stephenson, A. G. (Mo. Bot. Gard., St. Louis, MO), pp. 289–303.

Coyne, J. A. & Orr, H. A. (2004) *Speciation* (Sinauer, Sunderland, MA).

Cracraft, J. (1983) Species concepts and speciation analysis. *Curr. Ornithol.* **1,** 159–187.

Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection* (John Murray, London).

de Queiroz, K. (1997) The Linnaean hierarchy and the evolutionization of taxonomy, with emphasis on the problem of nomenclature. *Aliso* **15,** 125–144.

de Queiroz, K. (1998) The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations. In *Endless Forms: Species and Speciation*, eds. Howard, D. J. & Berlocher, S. H. (Oxford Univ. Press, Oxford, U.K.), pp. 57–75.

de Queiroz, K. (1999) The general lineage concept of species and the defining properties of the species category. In *Species: New Interdisciplinary Essays*, ed. Wilson, R. A. (MIT Press, Cambridge, MA), pp. 49–89.

de Queiroz, K. (2005) A unified concept of species and its consequences for the future of taxonomy. *Proc. Calif. Acad. Sci.* **56,** in press.

Dobzhansky, T. (1937) *Genetics and the Origin of Species* (Columbia Univ. Press, New York).

Dobzhansky, T. (1950) Mendelian populations and their evolution. *Am. Nat.* **84,** 401–418.

Dobzhansky, T. (1970) *Genetics of the Evolutionary Process* (Columbia Univ. Press, New York).

Donoghue, M. J. (1985) A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist* **88,** 172–181.

Fitch, W. M. & Ayala, F. J. (1994) Tempo and mode in evolution. *Proc. Natl. Acad. Sci. USA* **91,** 6717–6720.

Flexner, S. B. & Hauck, L. C. (1993) *Random House Unabridged Dictionary* (Random House, New York).

Ghiselin, M. T. (1997) *Metaphysics and the Origin of Species* (State Univ. of New York Press, Albany, NY).

Grismer, L. L. (1999) An evolutionary classification of reptiles on islands in the Gulf of California, Mexico. *Herpetologica* **55,** 446–469.

Grismer, L. L. (2001) An evolutionary classification and checklist of amphibians and reptiles on the Pacific islands of Baja California, Mexico. *Bull. South. Calif. Acad. Sci.* **100,** 12–23.

Harrison, R. G. (1998) Linking evolutionary pattern and process. In *Endless Forms: Species and Speciation*, eds. Howard, D. J. & Berlocher, S. H. (Oxford Univ. Press, New York), pp. 19–31.

Hull, D. L. (1965) The effect of essentialism on taxonomy—Two thousand years of stasis (II). *Brit. J. Philos. Sci.* **16,** 1–18.

Hull, D. L. (1980) Individuality and selection. *Annu. Rev. Ecol. Syst.* **11,** 311–332.

Huxley, J. (1942) *Evolution, the Modern Synthesis* (Allen and Unwin, London).

Jordan, K. (1896) On mechanical selection and other problems. *Novit. Zool.* **3,** 426–525.

Jordan, K. (1905) Der Gegensatz zwischen geographischer und nichtgeographischer Variation. *Z. Wiss. Zool.* **83,** 151–210.
Linnaeus, C. (1753) *Species Plantarum* (Laurentii Salvii, Stockholm).
Linnaeus, C. (1758) *Systema Naturae,* 10th Ed. (Laurentii Salvii, Stockholm).
Linnaeus, C. (1766–1768) *Systema Naturae,* 12th Ed. (Laurentii Salvii, Stockholm).
Mallet, J. (1995) A species definition for the Modern Synthesis. *Trends Ecol. Evol.* **10,** 294–299.
Martin, G. (1996) Birds in double trouble. *Nature* **380,** 666–667.
Mayden, R. L. (1997) A hierarchy of species concepts: The denouement in the saga of the species problem. In *Species: The Units of Biodiversity*, eds. Claridge, M. F., Dawah, H. A. & Wilson, M. R. (Chapman & Hall, London), pp. 381–424.
Mayr, E. (1942) *Systematics and the Origin of Species* (Columbia Univ. Press, New York).
Mayr, E. (1955) Karl Jordan's contribution to current concepts in systematics and evolution. *Trans. R. Entomol. Soc. Lond.* **107,** 45–66.
Mayr, E. (1963) *Animal Species and Evolution* (Harvard Univ. Press, Cambridge, MA).
Mayr, E. (1969a) *Principles of Systematic Zoology* (McGraw–Hill, New York).
Mayr, E. (1969b) The biological meaning of species. *Biol. J. Linn. Soc.* **1,** 311–320.
Mayr, E. (1970) *Populations, Species, and Evolution* (Harvard Univ. Press, Cambridge, MA).
Mayr, E. (1982) *The Growth of Biological Thought: Diversity, Evolution, and Inheritance* (Belknap Press of Harvard Univ. Press, Cambridge, MA).
Mayr, E. & Provine, W. B., eds. (1980) *The Evolutionary Synthesis: Perspectives on the Unification of Biology* (Harvard Univ. Press, Cambridge, MA).
Meglitsch, P. A. (1954) On the nature of species. *Syst. Zool.* **3,** 49–68.
Michener, C. D. (1970) Diverse approaches to systematics. *Evol. Biol.* **4,** 1–38.
Nixon, K. C. & Wheeler, Q. D. (1990) An amplification of the phylogenetic species concept. *Cladistics* **6,** 211–223.
Paterson, H. E. H. (1985) The recognition concept of species. In *Species and Speciation*, ed. Vrba, E. S. (Transvaal Museum, Pretoria, South Africa), pp. 21–29.
Poulton, E. B. (1903) What is a species? *Trans. Entomol. Soc. London* **1903,** lxxvii–cxvi.
Rosen, D. E. (1979) Fishes from the uplands and intermontane basins of Guatemala: Revisionary studies and comparative geography. *Bull. Am. Mus. Nat. Hist.* **162,** 267–376.
Simpson, G. G. (1951) *Evolution (Lawrence, Kans.)* **5,** 285–298.
Simpson, G. G. (1961) *Principles of Animal Taxonomy* (Columbia Univ. Press, New York).
Sneath, P. H. A. & Sokal, R. R. (1973) *Numerical Taxonomy: The Principles and Practice of Numerical Classification* (Freeman, San Francisco).
Sokal, R. R. & Crovello, T. J. (1970) The biological species concept: A critical evaluation. *Am. Nat.* **104,** 127–153.
Templeton, A. R. (1989) The meaning of species and speciation: A genetic perspective. In *Speciation and Its Consequences*, eds. Otte, D. & Endler, J. A. (Sinauer, Sunderland, MA), pp. 3–27.
Van Valen, L. (1976) Ecological species, multispecies, and oaks. *Taxon* **25,** 233–239.
Wright, S. (1940) The statistical consequences of Mendelian heredity in relation to speciation. In *The New Systematics*, ed. Huxley, J. (Oxford Univ. Press, London), pp. 161–183.
Zink, R. M. (1996) Bird species diversity. *Nature* **381,** 566.

Part IV

GENOMIC APPROACHES AND NEW INSIGHTS ON DIVERSITY

Because Mayr was not a geneticist, we do not count among his direct legacies our current era of genomics. But, in some respects, genomic studies of biological diversity are just the next step on a ladder that Mayr helped to hoist. Furthermore, it is fair to ask whether genomic tools are changing our view of biological diversity. One example of the way our view has changed is provided in the article by Ochman *et al.* (Chapter 12) that is mentioned above. Another example lies in the paper by James Lake and colleagues, "Decoding the Genomic Tree of Life" (Chapter 14), who use genomic data to reconstruct the process by which eukaryotes arose from prokaryotes. Unlike typical phylogenetic events, such as the splitting of lineages, eukaryotes appear to have arisen by the fusion of genomes. The authors describe the development and application of a new phylogenetic method, called ''conditioned reconstruction,'' which is designed to detect fusion events.

As the number of sequenced genomes grows, so will the number and availability of tools for identifying the genes responsible for important variation. This is a major point of the paper by Scott Edwards *et al.* (Chapter 6) that was discussed above. Two other papers in this volume demonstrate some of the latest techniques for finding genes responsible for phenotypes of interest. Stuart Macdonald and Anthony Long, in "Prospects for Identifying Functional Variation across the Genome" (Chapter 15), describe a new method for reducing the number of single nucleotide polymorphisms that are required in association mapping studies for genes that contribute to traits that have recently been under natural selection. The idea follows

from the expectation that recent selection will have shaped divergence, and especially polymorphism, in and around the relevant sites. Using population genetic predictions of the response to selection of linked sites, it should be possible to conduct genomic scans of variation and divergence to identify the subset of polymorphic sites upon which to base an association mapping study. They demonstrate the method by looking at polymorphism within and divergence between *Drosophila* species.

Trudy Mackay *et al.*, in "Genetics and Genomics of *Drosophila* Mating Behavior" (Chapter 16), also used a *Drosophila* model to identify genomic sites with interesting functions—mating behavior, in this case. Traditionally, genes that are directly involved in reproduction are not the easiest to study genetically, simply because mutants often have low reproductive success. These authors took the artificial selection approach and generated, over the course of 20 generations, two lines of *Drosophila melanogaster* that had high and low mean values for mating speed. They then conducted a microarray study to see which genes differed in expression level between the two divergently selected lines of flies.

The final paper in the volume takes an explicitly forward look and describes the ongoing and future changes that are happening to the biological sciences. With genomic sequences for many organisms having been available for several years, many biologists are turning to the highly integrated study of cellular processes and networks, a field that is called Systems Biology (Hood, 2003). Mónica Medina, "Genomes, Phylogeny, and Evolutionary Systems Biology" (Chapter 17), writes about how this nascent field is being shaped by the availability of genome sequences throughout the tree of life and of the kinds of questions about the evolution of networks that we can anticipate. Surely, just as Systems Biology emerges and qualitatively new kinds of insights emerge about how cells function, so too will emerge the field of Evolutionary Systems Biology with concomitant insights on the evolution of cell function.

REFERENCE

Hood, L. (2003) Systems biology: Integrating technology, biology, and computation. *Mech. Ageing Dev.* **124,** 9–16.

14

Decoding the Genomic Tree of Life

ANNE B. SIMONSON,[†§] JACQUELINE A. SERVIN,[*] RYAN G. SKOPHAMMER,[†] CRAIG W. HERBOLD,[*] MARIA C. RIVERA,[†§] AND JAMES A. LAKE[*†‡§]

Genomes hold within them the record of the evolution of life on Earth. But genome fusions and horizontal gene transfer (HGT) seem to have obscured sufficiently the gene sequence record such that it is difficult to reconstruct the phylogenetic tree of life. HGT among prokaryotes is not random, however. Some genes (informational genes) are more difficult to transfer than others (operational genes). Furthermore, environmental, metabolic, and genetic differences among organisms restrict HGT, so that prokaryotes preferentially share genes with other prokaryotes having properties in common, including genome size, genome G + C composition, carbon utilization, oxygen utilization/sensitivity, and temperature optima, further complicating attempts to reconstruct the tree of life. A new method of phylogenetic reconstruction based on gene presence and absence, called conditioned reconstruction, has improved our prospects for reconstructing prokaryotic evolution. It is also able to detect past genome fusions, such as the fusion that appears to have created the first eukaryote. This genome fusion between a deep branching eubacterium, possibly an ancestor of the cyanobacterium and a proteobacterium, with an archaeal eocyte (crenarchaea), appears to be the result of an early symbiosis. Given new tools and new

*Molecular Biology Institute, Departments of [†]Molecular, Cell, and Developmental Biology and [‡]Human Genetics, and [§]National Aeronautics and Space Administration Astrobiology Institute, University of California, 242 Boyer Hall, Los Angeles, CA 90095.

Abbreviations: HGT, horizontal gene transfer; CR, conditioned reconstruction.

genes from relevant organisms, it should soon be possible to test current and future fusion theories for the origin of eukaryotes and to discover the general outlines of the prokaryotic tree of life.

Today there is enormous interest in discovering the tree of life. But as we get closer to reconstructing it, new experimental and theoretical challenges appear that cause us to reexamine our goals. New obstacles may initially seem insurmountable, but in reality they enrich our understanding of the evolution of life on Earth.

One of the most recent evolutionary mechanisms to challenge our view of genome evolution is the massive horizontal gene transfer (HGT) that has recently become so apparent (Campbell, 2000; Doolittle, 1999a; Gogarten *et al.*, 2002; Karlin *et al.*, 1997; Koonin *et al.*, 2001; Lawrence and Ochman, 1998, 2001; Rivera *et al.*, 1998). This genetic crosstalk theoretically has the potential to erase much of the history of life that has been recorded in DNA. Indeed, some scientists think that HGT has already effectively erased the phylogenetic history contained within prokaryotic genomes (reviewed in Doolittle, 1999b).

Although sympathetic to many of these points, we think the best way to decide whether the tree of life is knowable is to try one's hardest to determine it. This article reviews the progress made using whole-genome analyses but does so primarily from the unique perspective of our laboratory. When Darwin uttered his famous quote, ''The time will come I believe, . . . when we shall have fairly true genealogical trees of each great kingdom of nature,'' (1887) he was not describing prokaryotic life. Rather, he probably envisioned understanding the trees of animal and plant life. In that sense, part of his dream is already a reality. We currently understand the major radiations of the bilateral animals (Aguinaldo *et al.*, 1997; Halanych *et al.*, 1995), and the relationships linking the major plant groups are starting to be understood (Karol *et al.*, 2001; Pryer *et al.*, 2001, 2002; Nickrent *et al.*, 2000; Soltis *et al.*, 1999). This review, however, focuses on understanding the radiations that occurred even before those of the plants and animals, namely the enigmatic evolution of prokaryotes and the emergence of eukaryotes.

The origin of the eukaryotes was a milestone in the evolution of life, because eukaryotes are utterly different from prokaryotes in their spatial organization. Eukaryotes, for example, possess an extensive system of internal membranes that traverse the cytoplasm and enclose organelles, including the mitochondrion, chloroplast, and nucleus. This compartmentalization has required a number of unique eukaryotic innovations. The most dramatic innovation is the nucleus, a specific compartment for storing and transcribing DNA, for processing DNA and RNA, and possibly even for translating mRNAs (Hentze, 2001). The nucleus is unique to

eukaryotes, hence it and the nuclear genome are the defining characters for which eukaryotes are named (eu, good or true; karyote, kernel, as in nucleus).

The prokaryotes, with their simple cellular organization, are generally thought to have preceded the eukaryotes (although see Poole *et al.*, 1999). Which prokaryotic groups branched first, however, is not clear, because the root of the tree of life is uncertain and in flux due to a concern that artifacts of phylogenetic reconstruction may have unduly influenced the location of even the root that has the most experimental support (Penny and Poole, 1999; Philippe and Forterre, 1999).

THE HGT REVOLUTION

The possibility of analyzing complete genomes awakened interest in prokaryotic genome evolution and profoundly changed our understanding of genome evolution. Before the first genomes were sequenced, there was nearly unanimous scientific agreement that prokaryotic genomes were evolving clonally, or approximately so. In other words, as generation after generation of bacteria divided, each bacterium would contain the DNA it inherited from its parent, except that occasionally a single DNA nucleotide might have mutated, causing a minor change in the daughter genome. Thus it was thought that the family tree derived from any one gene would look like the family tree from any other gene. Diploid eukaryotic cells with two copies of each gene per cell slightly complicated this picture, but they, too, were thought to be evolving clonally. Most researchers felt comfortable with the premise that reliable organismal trees could be calculated from sequences of individual genes. In particular, rRNA genes were favored, because rRNA was easy to sequence, and it was assumed trees calculated from rRNA would probably be the same as those calculated from any other genes. However, it was not acknowledged that HGT had the potential to significantly alter gene trees. For example, if a gene were horizontally transferred from a prokaryote to a human, then the tree reconstructed from that gene would place humans in the midst of prokaryotes. Furthermore, each gene tree would show a different set of relationships. (Sometimes one keeps track of whether the transferred genes are new to the genome or whether they replace existing genes. Although this distinction can be important, in this paper, we will refer to both types of exchange as HGT.) Because so much attention was focused on the approximately clonal evolution of rRNA in the pregenomic era, only a few genes other than rRNA were sequenced from multiple organisms, and HGT was largely overlooked.

Once complete genomes were available, the pace of discovery accelerated, as highlighted in early analyses of complete, or nearly complete,

genome studies from the laboratories of R. Doolittle (Doolittle and Handy, 1998), W. F. Doolittle (Brown and Doolittle, 1999), Gogarten (Gogarten *et al.*, 1999), Golding (Ribeiro and Golding, 1998), Ochman (Lawrence and Ochman, 1998), and ourselves (Rivera *et al.*, 1998). These and even more recent studies of the evolution of life, based on analyses of complete genomes, described below, revealed the flaws in the old view of clonal evolution. Scientific opinion has now shifted and favors a significant role for HGT in prokaryotic genome evolution.

HGT HAS PROFOUNDLY AFFECTED OUR UNDERSTANDING OF PROKARYOTIC GENOME EVOLUTION

Three remarkable new findings, based on analyses of whole genomes, have engendered appreciation for the important role of HGT in prokaryotic evolution. First, HGT is now generally recognized to be rampant among genomes (rampant at least on a geological timescale). Second, not all genes are equally likely to be horizontally transferred. Informational genes (involved in transcription, translation, and related processes) are rarely transferred, whereas operational genes (involved in amino acid biosynthesis, and numerous other operational activities) are readily transferred. Third, biological and physical factors appear to have altered HGT. These include intracellular structural constraints among proteins (the complexity hypothesis), interactions among organisms, and interactions with the physical environment. These three findings are described below.

EVIDENCE FOR EXTENSIVE HGT

As early as 1996, the complete sequence of the methanogen *Methanococcus janaschii* (Bult *et al.*, 1996) revealed that its genome consisted of certain groups of genes that were much more similar to eukaryotic genes than those from bacteria, whereas other groups of genes were much more closely related to their bacterial homologs. Koonin *et al.* (1997) substantiated that the *M. jannaschii* genes for translation, transcription, replication, and protein secretion were more similar to eukaryotes than to bacteria. They interpreted this finding to mean that archaea were a chimera of eukaryotic and eubacterial genes (Koonin *et al.*, 1997). Using whole-genome phylogenetic methods, our laboratory discovered the presence of two superclasses of genes in prokaryotes that had different relationships to eukaryotic genes. In that study (Rivera *et al.*, 1998) of the *Escherichia coli*, *Synechocystis* PCC6803 (a cyanobacterium), *M. jannaschii*, and *Saccharomyces cerevisiae* genomes (Blattner *et al.*, 1997; Bult *et al.*, 1996; Goffeau *et al.*, 1996; Kaneko *et al.*, 1996), the *M. jannaschii* informational genes, consisting of gene products responsible for such processes as trans-

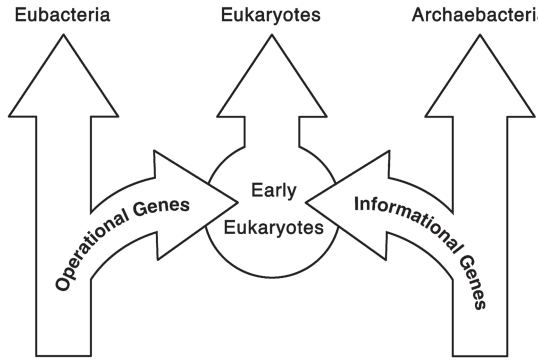


FIGURE 14.1 Early genome studies indicated that eukaryotes were a mixture of eubacterial and archaebacterial genes with an unusual distribution. The operational genes were primarily from the eubacteria, and the informational genes were from the archaebacteria.

lation and transcription, were found to be most closely related to those found in eukaryotes. The operational genes of the eukaryote, responsible for the day-to-day operation of the cell (operational genes), on the other hand, were most closely related to their counterparts found in *E. coli* and *Synechocystis* (Rivera *et al.*, 1998). Of the yeast genes analyzed, approximately one-third were informational genes, and two-thirds were operational genes. This provided good evidence that the 16S rRNA tree does not reflect the evolution of all of the genes in a genome and also supplied evidence that early eukaryotes were a chimera of eubacteria and archaebacterial genes. A stylized illustration of these results is shown in Fig. 14.1. Recently, a thorough comprehensive analysis involving large numbers of genomes and genes has documented the strength of this correlation (Esser *et al.*, 2004).

Further evidence for extensive HGT came from the observation that another methanogen, *Methanobacterium thermoautotrophicum*, contains several regions that have an ≈10% lower G + C content than the G + C content of the whole genome on average (Smith *et al.*, 1997). ORFs in these regions exhibit a codon usage pattern atypical of *M. thermoautotrophicum*, suggesting that the DNA sequences may have been acquired by HGT (Smith *et al.*, 1997).

Additional evidence for HGT came from a thermophilic relative of the methanogens, *Archaeoglobus fulgidus*. ORFs in the functional categories of translation, transcription, replication, and some essential biosynthetic pathways in this prokaryote are very similar to those in *M. jannaschii*. However, these two genomes differ in many of their opera-

tional genes, such as those for environmental sensing, transport, and energy metabolism (Klenk *et al.*, 1997). The tryptophan biosynthesis pathway in *A. fulgidus* seems very closely related to the eubacterium *Bacillus subtilis*, even though these two are separated by large distances on the 16S tree (Klenk *et al.*, 1997). These observations suggested that the extent of gene exchange that has occurred in the methanogens and their relatives is tremendous.

Among the extreme thermophiles, some of which live in temperatures in excess of the boiling temperature of water, HGT is equally prevalent (Makarova *et al.*, 1999). Lecompte *et al.* (2001) compared the three closely related proteomes from the high-temperature methanogen relatives *Pyrococcus abyssi*, *Pyrococcus furiosus*, and *Pyrococcus horikoshii*. In their gene analysis, the ORFs encoding translation proteins and transcription proteins (informational genes) fairly consistently indicated that the distances among the three species were uniform, as would happen if these genes were evolving approximately clonally. However, most other ORFs (mainly operational genes) gave a wide distribution of distances. The existence of a distribution was interpreted as evidence of HGT (Lecompte *et al.*, 2001), because the horizontal transfer of genes from closely and distantly related organisms would be expected to correspond to heterogeneous distances. In addition, *P. furiosus* is capable of transporting and metabolizing maltose/maltodextrin, properties that are absent in *P. horikoshii*. Of two maltose/maltodextrin import systems in *P. furiosus*, one has the greatest similarity to the transport system in *E. coli*, a finding most parsimoniously explained as a lateral transfer of the entire system from *E. coli* to *P. furiosus* (DiRuggiero *et al.*, 2000; Maeder *et al.*, 1999). Comparison between *P. furiosus* and *P. abyssi* has revealed linkage between restriction-modification genes. Because codon usage is different in various organisms, the codon biases of some restriction-modification systems in the *Pyrococcus* genomes suggest that these systems have been acquired by horizontal transfer (Chinen *et al.*, 2000).

HGT is also widely prevalent in the eubacteria [see the article by Ochman *et al.*, "Examining Bacterial Species Under the Specter of Gene Transfer and Exchange" (Chapter 12)]; this has been demonstrated in *Aquifex aeolicus*, where little consistency was seen among trees reconstructed from a number of operational genes (Deckert *et al.*, 1998). Comparative analyses of *E. coli* ORFs showed that 675 *E. coli* ORFs have greatest similarity to *Synechocystis*, 231 to *M. jannaschii*, and 254 to the eukaryote *S. cerevisiae* (Blattner *et al.*, 1997). Using skewed base composition and codon usage as a measure of an alien gene, Ochman and coworker (Lawrence and Ochman, 1998) argued that 755 of 4,288 *E. coli* ORFs have been horizontally acquired in 234 lateral transfer events, because *E. coli* diverged from *Salmonella* ≈100 million years ago (Lawrence and Ochman, 1998).

Classically, the three principal molecular mechanisms known to produce horizontal transfer are transformation, conjugation, and transduction. Numerous authors have found evidence of transduction. For example, the *B. subtilis* genome harbors a number of foreign genes, as evidenced by many prophage-like regions encompassing ≈15% of the genome (Kunst *et al.*, 1997). Like its close relative *B. subtilis*, *Bacillus halodurans,* an alkaliphilic prokaryote, also possesses regions with a G + C content similar to that of some viruses (Takami *et al.*, 2000). As a consequence of this similarity, those DNA sequences were proposed to have been obtained by lateral transfer (Takami *et al.*, 2000). The genome of *Clostridium acetobutylicum* contains genes missing in *B. subtilis*. These genes have a number of different phylogenetic relationships. For example, 49 genes reveal an immediate relationship between *C. acetobutylicum* and eukaryotes, and another 195 are most closely related to archaeal extremophiles (Nolling *et al.*, 2001).

The cyanobacterium *Synechocystis PCC6803* is another bacterium whose genome supports extensive HGT among prokaryotes. The genome of *Synechocystis* contains a number of insertion sequence (IS) elements. The DNA in the vicinity of the IS elements displays features of *E. coli* DNA, indicative of horizontal genetic acquisitions (Cassier-Chauvat *et al.*, 1997).

ALTHOUGH HGT IS RAMPANT, IT IS NOT RANDOM: THE COMPLEXITY HYPOTHESIS

In a subsequent phylogenetic analysis (Jain *et al.*, 1999), our laboratory examined the frequency of horizontal/lateral transfer of operational genes among six prokaryotic proteomes, *E. coli*, *Synechocystis* PCC6803, *B. subtilis*, *A. aeolicus*, *M. jannaschii,* and *A. fulgidus*, using three different topology-based tests of gene ortholog relationships to measure the extent of HGT in informational and operational genes. All three tests showed that operational genes have been continually transferred much more frequently among prokaryotes since the last common ancestor of life or cenancestor (Fitch and Upper, 1987). To explain at least partially why operational genes undergo HGT more frequently than informational genes, we proposed the complexity hypothesis (Jain *et al.*, 1999), which posits that informational genes are less likely to undergo horizontal transfer, because their products are members of large complexes with many intricate interactions. Operational genes, on the other hand, are generally not parts of large complexes, and thus are more readily transferred. Obviously the complexity hypothesis is not the sole factor relating differential horizontal transfer rates between informational and operational genes, because many other factors, including environmental ones, can also

modify horizontal transfer. At the same time, the data are forcing us to recognize that gene exchange is not simply occurring within species, but extensive exchanges also occur within larger groups of prokaryotes consisting of multiple species as well.

HGT ACCELERATES GENOME INNOVATION AND EVOLUTION

It is becoming clear that HGT has had great impact on the evolution of life on Earth. It is a key agent, perhaps the major agent, responsible for spreading genetic diversity among prokaryotes by moving genes across species boundaries (Jain *et al.*, 2003). By rapidly introducing newly evolved genes into existing genomes, HGT circumvents the slow step of *ab initio* gene creation and thereby accelerates genome innovation (the acquisition of novel genes by organisms), although not necessarily gene evolution. We refer to a collection of organisms that can share genes by HGT but need not be in physical proximity as an exchange community. In effect, when organisms are exchanging genes, genome innovation is increased in proportion to the effective population sizes of their exchange groups.

We were interested in the structure of exchange communities and in the environmental and other factors that help define them. In an analysis of ≈20,000 genes contained in eight free-living prokaryotic genomes, we assessed which geographic, environmental, and internal parameters have influenced genetic exchange by HGT and found that HGT is not random but depends critically upon these internal and environmental factors. The statistically significant parameters were similar genome sizes, genome G + C compositions, carbon utilization methods, oxygen tolerance, and maximum, optimal, and minimum temperatures (Jain *et al.*, 2003). By identifying and quantifying those parameters, we were able to delineate exchange community boundaries, estimate the effective population size of exchange groups, and thereby estimate the extent to which HGT has accelerated genome innovation. By correlating the extent of HGT among specific organisms with the degree of phylogenetic clustering of those organisms observed on all possible gene trees, one can determine the effect of various environmental or other parameters on HGT. We found that HGT preferentially occurs among organisms that have environmental and genomic factors in common, a phenomenon we termed positive associativity (Jain *et al.*, 2003). In short, like prokaryotes preferentially exchanged genes by HGT with like prokaryotes. It is difficult to ascertain precisely how much HGT has accelerated prokaryotic genome innovation, but the acceleration is significant. It has been estimated there are 109 prokaryotic species on Earth containing 10^{30} prokaryotes (Whitman *et al.*, 1998). The sizes of exchange communities are unknown, but some of the

parameters characterizing them are not too different from those of some terrestrial ecosystems. The median prokaryotic population of 12 diverse soil ecosystem types, as reviewed by Whitman, Coleman, and Wiebe (1998), is $\approx 10^{28}$ prokaryotes, suggesting an average exchange group could contain 10^7 species. Allowing 3 orders of magnitude for the inexactness of our estimate, the increase in innovation afforded by HGT could be as small as 104, but even this would constitute a huge HGTdependent increase in innovation. This means that a species exchanging genes only with other members of its species would take 10,000 years to obtain the amount of genome innovation that would occur for an average exchange group in just 1 year. Indeed, HGT may be responsible for a remarkable increase in genome innovation that greatly exceeds anything that could have been accomplished by clonal evolution.

HGT GREATLY COMPLICATES RECONSTRUCTING THE UNIVERSAL TREE OF LIFE

W. Ford Doolittle recently reviewed the state of ''Phylogenetic Classification and the Universal Tree'' in a thoughtful analysis (Doolittle, 1999b). He points out the specific challenges to classification that HGT presents as follows, ''If, however, different genes give different trees, and there is no fair way to suppress this disagreement, then a species (or phylum) can 'belong' to many genera (or kingdoms) at the same time: There really can be no universal phylogenetic tree of organisms based on such a reduction to genes.'' In other words, Doolittle (1999b) suggests that the gene mixing resulting from HGT is so extensive that it might preclude one from ever reconstructing the tree of life. Although it would be disingenuous to pretend that the difficulties are not sizable, our laboratory is pursuing an alternative strategy. We agree that HGT is extensive and imposes limits to phylogenetic reconstruction. However, we also think the only way to discover whether HGT could destroy Darwin's dream of understanding the great kingdoms of nature is to assume that it cannot, and then make every effort to try to determine the tree of life. Some of the barriers to reconstructing the tree of life and the progress being made to surmount them are discussed below.

PITFALLS IN RECONSTRUCTING THE TREE OF LIFE

Consider what has happened to the once-ebullient field of rRNA phylogenies. For years, phylogenies based on rRNAs were the holy grail of microbial phylogenetics. To be sure, rRNA-based phylogenies have been responsible for many successes, including the new animal phylogeny and demonstrations that the mitochondrion and chloroplast are endosym-

bionts (Adoutte *et al.*, 1999, 2000; Aguinaldo *et al.*, 1997; Gray, 1999; Halanych *et al.*, 1995; Schwarz and Kossel, 1980). However, prokaryotic phylogenies are another story. One has only to read the latest *Bergey's Manual* (Boone and Castenholz, 2001) to realize that the tree of prokaryotic life is fuzzy and unresolved, so much so that rRNA-based trees, although capable of identifying to which phylum a prokaryote belongs, in most cases cannot determine how the phyla are related to each other. Furthermore, our ability to determine phylogenies accurately depends upon how extensive HGT has been. If very little or no HGT has occurred, then current methods of analysis will allow one to reconstruct the clonal tree of life. At the other extreme, if all genes undergoHGTonce per year, then coherent gene trees will be unobtainable. Between these extremes lies a continuum of results, so that perhaps the question we should be asking is, how much phylogenetic information can one obtain, and how can it best be analyzed?

HOW CAN ONE RECONSTRUCT THE TREE OF LIFE IN THE PRESENCE OF HGT?

Presences and absences of genes and gene products have been used for more than two decades to support parsimonious conclusions about the tree of life (Charlebois *et al.*, 2000; Dickerson, 1980; Lake *et al.*, 1982; Woese *et al.*, 1986). In these analyses, the absences and presences of genes were used as character states, much in the way that nucleotides A, C, G, and T are used as character states in sequence analyses. With the availability of complete genomes, useful methods have been developed for whole-genome analyses (Fitz-Gibbon and House, 1999; Montague and Hutchison, 2000; Snel *et al.*, 1999; Tekaia *et al.*, 1999). However, when analyzed using parsimony and simple distance-based methods, these analyses can be significantly influenced by HGT (Eisen, 2000; House and Fitz-Gibbon, 2002).

Recently the prospects of recovering the tree of life in the presence of HGT have improved with the development of a new mathematical algorithm, conditioned reconstruction (CR), for whole-genome-based phylogenetic reconstructions (Lake and Rivera, 2004). Like some other whole-genome methods, CR analyses also use the absences and presences of genes as character states but, through the use of a reference genome, they can obtain additional information that is not available in other types of analyses. For example, by restricting the analyses to only the genes present in a reference genome R, one can also estimate the number of gene pairs that are missing in both genomes A and B. This is critical information that is not available without the reference genome, and it allows one to use a very general class of mathematical (Markov) models to reconstruct the tree of life.

In CR, the dynamic deletions and insertions of genes that occur during genome evolution, including the insertions introduced by HGT, actually help provide the information needed to reconstruct phylogenetic trees. CR appears to have the potential to reconstruct deeper branchings in the tree of life than is possible with sequence analyses, because whole gene characters evolve more slowly than nucleotides, amino acids, and even gene inserts.

At the same time, it is important to recognize that CRs are not a panacea. It is difficult to assign the gene ortholog sets used by CR analyses accurately, because the process is greatly complicated by the need to distinguish orthologs from paralogs and to simultaneously recognize recently duplicated genes (Lake and Rivera, 2004). Currently available methods to identify gene ortholog sets are still rudimentary, and new methods are just beginning to be developed. Because CR can be no better than the ortholog sets that it is based on, much improvement is needed in this area.

Although CR analysis provides a new tool for investigating the tree of life, other methods are also likely to provide important information about deep divergences in the tree of life. These include such important emerging techniques as phylogenetic analyses of concatenated gene sequences (Baldauf *et al.*, 2000; Brown *et al.*, 2001) or of sets of gene sequences (Esser *et al.*, 2004; Raymond *et al.*, 2002), particularly of informational genes, and the analyses of more slowly evolving sequence-related characters such as gene inserts, gene fusions, and even structural domains (Gupta and Singh, 1994; Stechmann and Cavalier-Smith, 2002; Yang *et al.*, 2005). Like CRs, these methods also have their limitations, and much work remains to be done to improve these promising techniques as well.

One of the most remarkable properties of CR is that it can rigorously identify the merger of genomes, a process that until now could not be analyzed using gene sequence. A recently published application of this method has provided evidence that the eukaryotic genome was actually formed by a fusion of the genomes from two disparate prokaryotes.

EVIDENCE THAT AN ANCIENT GENOME FUSION FORMED THE FIRST EUKARYOTE

Various theories have been proposed for the origin of the nuclear genes of eukaryotes. These include the autogenous-, chimeric-, and genome-fusion theories. To obtain a better understanding of eukaryotic origins, we analyzed 10 complete genomes using the CR method (Rivera and Lake, 2004). The sample was comprised of two eukaryotic genomes and eight prokaryotes representing the diversity of prokaryotic life. An additional 24 prokaryotic genomes were studied in supplementary studies. The results from one analysis are shown in Fig. 14.2. In this analysis, the

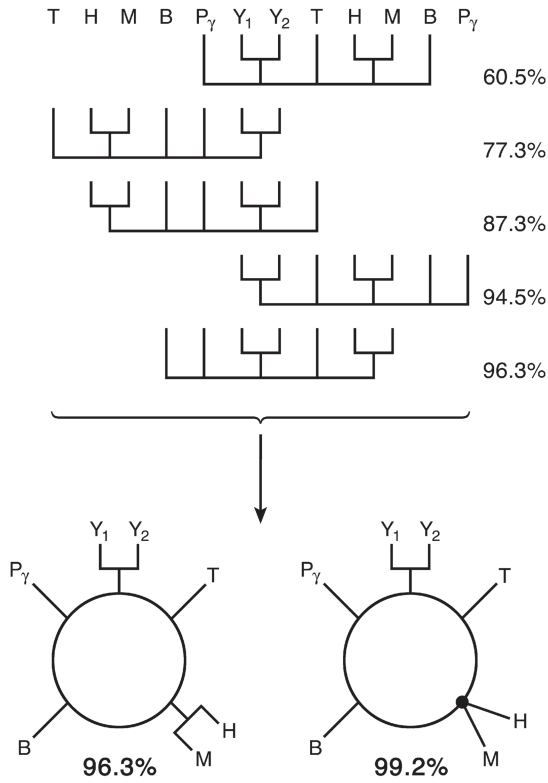


FIGURE 14.2 CRs provide evidence for the ring of life. The genomes are from two yeasts, Y_1 (*Schizosaccharomyces pombe*) and Y_2 (*S. cerevisiae*); a gammaproteobacterium, P (*Xylella fastidiosa*); a bacillus, B (*Staphylococcus aureus* MW2); a halobacterium, $\bar{H}$ (*Halobacterium* sp. NRC-1); a methanococcus, M (*Methanosarcina mazei* Goe1); an eocyte, T (*Sulfolobus tokodaii*); and an archaeoglobium not shown, the conditioning genome (*A. fulgidus* DSM4304). Cumulative probabilities are shown at the right of each tree. Fully and partially resolved rings are *Lower Left* and *Lower Right*, respectively. [Reproduced with permission from Rivera and Lake (2004) (Copyright 2004, Nature Publishing Group).]

five most probable trees are from a set of three Bacteria, three Archaea, and two eukaryotes. The cumulative probabilities of these five trees are shown at the right of each tree. We initially thought that the resolution of the tree was disappointingly poor, because the most probable tree was supported by a low bootstrap value (70% approximately corresponds to the 95% confidence level), and the other trees were supported by even lower values.

However, when the five most probable unrooted trees are aligned by

shifting each to the left or the right until their leaves match, they form a repeating pattern indicating that the five trees are simply permutations of an underlying cyclic pattern. (The five most probable unrooted trees are shown with leaves pointing upward to emphasize that each is part of a repeating pattern.) This suggested that they are derived from the single cycle graph (Lake and Rivera, 2004), or ring, shown in Fig. 14.2 *Lower Left*. When that ring is cut at any of the five central arcs and then unfolded, the resulting unrooted tree will correspond to one of the five most probable trees. In other words, the data are not tree-like; they are ring-like.

Previously, a combinatorial analysis of the genomic fusion of two organisms had shown that the CR algorithm recovers all permutations of the cycle graph (Lake and Rivera, 2004). Hence these results can be interpreted in a manner analogous to the interpretation of restriction digests of a circular plasmid or the mapping of a circular chromosome, as implying a ring of life. The fully resolved ring shown in Fig. 14.2 *Lower Left* is fully consistent with all five of the resolved trees shown in Fig. 14.2 *Upper*. That ring explains 96.3% of the bootstrap replicates, and the partially resolved ring in Fig. 14.2 *Lower Right* explains almost all (99.2%) of the bootstrap replicates. These and other control experiments provide robust evidence for the completely resolved ring (Fig. 14.2 *Lower Left*) and even stronger evidence for the less-resolved ring (Fig. 14.2 *Lower Right*).

Analyses of this type supported the ring, but other experiments were still necessary to identify the fusion organism. In particular, it was necessary to show that it was the eukaryotes, rather than a prokaryote, that resulted from the genome fusion that closed the ring of life. Hence the identity of the fusion organism was explicitly tested by systematically eliminating the eukaryotes and the individual prokaryotes for the ring of life. The ring opened into a tree only when both eukaryotes were simultaneously deleted from the analysis, indicating the eukaryotic genome had inherited genes from its prokaryotic fusion partners. This then demonstrated that eukaryotes are indeed the products of genome fusions. Furthermore, statistical support for the ring remained high for all possible choices of conditioning genome. From these results and other studies not discussed here, we inferred that the eukaryotic nuclear genome was formed from the genome fusion of either a proteobacterium or a member of a large photosynthetic clade that includes the Cyanobacteria and the Proteobacteria, with an archaeal eocyte as shown schematically in Fig. 14.3.

IMPLICATIONS OF THE RING OF LIFE

Various theories have been proposed for the origin of eukaryotes. These include autogenous, chimeric, and genome fusion theories. The

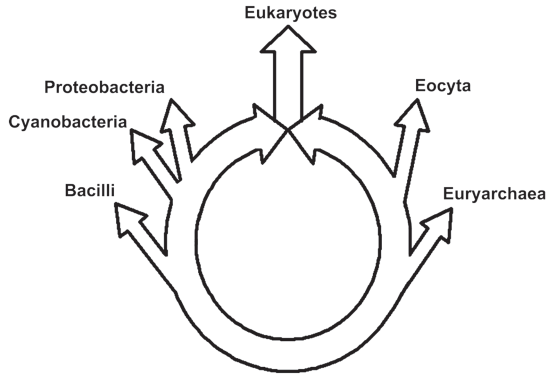


FIGURE 14.3 A schematic diagram of the ring of life. The eukaryotes include all eukaryotes plus the two eukaryotic root organisms, the operational and informational ancestors. Ancestors defining major prokaryotic groups are represented by branching points from the ring. *Archaea* (Woese *et al.*, 1990), shown on the bottom right, includes the *Euryarchaea*, the *Eocyta*, and the informational eukaryotic ancestor. *Karyota* (Lake, 1988), shown on the upper right of the ring, includes the *Eocyta* and the informational eukaryotic ancestor. The upper left circle includes the *Proteobacteria* (Woese *et al.*, 1990) and the operational eukaryotic ancestor. The most basal node on the left represents the photosynthetic prokaryotes and the operational eukaryotic ancestor.

results derived in the CR analyses argue against autogenous theories, i.e., tree of life theories, in which eukaryotes evolved clonally from a single, possibly very ancient, prokaryote. Chimeric theories refer to the acquisition of genes by eukaryotes from multiple sources through unspecified mechanisms. The data presented here argue against them, except of course chimeric theories that specifically propose genome fusions.

At least half a dozen genome fusion theories have been proposed in which the eukaryotic genome originated from two diverse genomes (Gupta *et al.*, 1994; Horiike *et al.*, 2001; Lake and Rivera, 1994; Lake *et al.*, 1982; Martin and Muller, 1998; Moreira and Lopez-Garcia, 1998). These are strongly supported by CR analyses. By default, an endosymbiosis (Margulis, 1970) between two prokaryotes is probably the mechanism responsible for the genome fusion observed here, although the fusion signal may have been augmented by gene contributions from eukaryotic organelles. Symbiotic relationships are fairly common among organisms living together and, in rare cases, this leads to endosymbiosis, the intracellular capture of former symbionts (Margulis, 1970). Given a genome fusion, and in the absence of other mechanisms that could produce fusions, one concludes that an endosymbiosis was the probable cause.

Although the data reviewed here solidly support the ring of life, it is important to recognize that CR analysis is a new technique, and its usefulness is still being explored. Currently, the resolution in CR trees is still relatively low. At the same time, it seems unlikely that the ring could be caused by low phylogenetic resolution, because the ring signal monitored in CR analyses is fundamentally different from the parsimony signals that are generated by poorly resolved trees (Lake and Rivera, 2004).

The ring of life is consistent with and confirms and extends a number of previously reported results. It implies that prokaryotes predate eukaryotes, because two preexisting prokaryotes contributed their genomes to create the first eukaryotic genome. This likely places the root of the ring below the eubacterial– and eocytic–eukaryotic last common ancestors, as shown in Fig. 14.3. This partial rooting of the ring of life is consistent with the eukaryotic rooting implied by the EF-1α insert that is present in all known eukaryotic and eocytic EF-1α sequences and lacking in all paralogous EF-G sequences (Gupta, 1998; Rivera and Lake, 1992).

The ring of life also explains some previously confusing observations and raises new ones. Because the eukaryotic genome resulted from a fusion, it is expected that in some gene trees, eukaryotes will be related to *Bacteria*, whereas in other gene trees, eukaryotes will be related to *Archaea*, in accord with the results of others (Brown and Doolittle, 1997; Feng *et al.*, 1997; Gupta, 1998; Martin *et al.*, 1996). The observations of ourselves and others (Esser *et al.*, 2004; Rivera *et al.*, 1998), that the informational genes of eukaryotes are primarily derived from *Archaea* and the operational genes are primarily derived from *Bacteria*, are also consistent with the ring. Those observations suggest that the operational genes have come from the eubacterial fusion partner and the informational genes, from the archaeal fusion partner. The ring of life does not explain why the fusion happened, but it provides a broad phylogenetic framework for testing theories for the origin and evolution of the eukaryotic genome. The genome fusion that created the ring of life may in some ways be the ultimate HGT.

ACKNOWLEDGMENTS

We thank M. Kowalczyk for illustrations. This work was supported by grants from the National Science Foundation, the National Aeronautics and Space Administration Astrobiology Institute, the Department of Energy, and the National Institutes of Health (to J.A.L.).

REFERENCES

Adoutte, A., Balavoine, G., Lartillot, N. & de Rosa, R. (1999) Animal evolution—the end of the intermediate taxa? *Trends Genet.* **15,** 104–108.

Adoutte, A., Balavoine, G., Lartillot, N., Lespinet, O., Prudhomme, B. & de Rosa, R. (2000) The new animal phylogeny: Reliability and implications. *Proc. Natl. Acad. Sci. USA* **97,** 4453–4456.

Aguinaldo, A. M. A., Turbeville, J. M., Linford, L. S., Rivera, M. C., Garey, J. R., Raff, R. A. & Lake, J. A. (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387,** 489–493.

Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290,** 972–977.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., ColladoVides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., *et al.* (1997) The complete genome sequence of *Escherichia coli K-12*. *Science* **277,** 1453–1462.

Boone, D. R. & Castenholz, R. W. (2001) *The Archaea and the Deeply Branching and Phototrophic Bacteria. Bergey's Manual of Systematic Bacteriology,* ed. Garrity, G. M. (Springer, New York), Vol. 1.

Brown, J. R. & Doolittle, W. F. (1997) Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61,** 456–502.

Brown, J. R. & Doolittle, W. F. (1999) Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. *J. Mol. Evol.* **49,** 485–495.

Brown, J. R., Douady, C. J., Italia, M. J., Marshall, W. E. & Stanhope, M. J. (2001) Universal trees based on large combined protein sequence data sets. *Nat. Genet.* **28,** 281–285.

Bult, C. J., White, O., Olsen, G. J., Zhou, L. X., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., *et al.* (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273,** 1058–1073.

Campbell, A. M. (2000) Lateral gene transfer in prokaryotes. *Theor. Popul. Biol.* **57,** 71–77.

Cassier-Chauvat, C., Poncelet, M. & Chauvat, F. (1997) Three insertion sequences from the cyanobacterium *Synechocystis* PCC6803 support the occurrence of horizontal DNA transfer among bacteria. *Gene* **195,** 257–266.

Charlebois, R. L., Singh, R. K., Chan-Weiher, C. C.-Y., Allard, G. C. C., Confaloniere, F., Curtis, B., Duget, M., Erauso, G., Faguy, D., Gaasterland, T., *et al.* (2000) Gene content and organization of a 281-kbp contig from the genome of the extremely thermophilic archaeon, *Sulfolobus solfataricus* P2. *Genome* **43,** 116–136.

Chinen, A., Uchiyama, I. & Kobayashi, I. (2000) Comparison between *Pyrococcus horikoshii* and *Pyrococcus abyssi* genome sequences reveals linkage of restriction-modification genes with large genome polymorphisms. *Gene* **259,** 109–121.

Darwin, F. (1887) *The Life and Letters of Charles Darwin* (John Murray, London).

Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., *et al.* (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392,** 353–358.

Dickerson, R. E. (1980) Structural conservatism in proteins over three billion years: Cytochrome with a touch of collagen. In *Diffraction and Related Studies*, ed. Srinivasan, R. (Pergamon, Oxford), Vol. 1, pp. 227–249.

DiRuggiero, J., Dunn, D., Maeder, D. L., Holley-Shanks, R., Chatard, J., Horlacher, R., Robb, F. T., Boos, W. & Weiss, R. B. (2000) Evidence of recent lateral gene transfer among hyperthermophilic Archaea. *Mol. Microbiol.* **38,** 684–693.

Doolittle, R. F. & Handy, J. (1998) Evolutionary anomalies among the aminoacyl-tRNA synthetases. *Curr. Opin. Genet. Dev.* **8,** 630–636.

Doolittle, W. F. (1999a) Lateral genomics. *Trends Genet.* **15,** M5–M8.

Doolittle, W. F. (1999b) Phylogenetic classification and the universal tree. *Science* **284,** 2124–2128.

Eisen, J. A. (2000) Assessing evolutionary relationships among microbes from whole-genome analysis. *Curr. Opin. Microbiol.* **3,** 475–480.

Esser, C., Ahmadinejad, N., Wiegand, C., Rotte, C., Sebastiani, F., Gelius-Dietrich, G., Henze, K., Kretschmann, E., Richly, E., Leister, D., *et al.* (2004) Genome comparisons speak to the origin of mitochondria and eukaryotes. *Mol. Biol. Evol.* **21,** 1643–1660.

Feng, D. F., Cho, G. & Doolittle, R. F. (1997) Determining divergence times with a protein clock: Update and reevaluation. *Proc. Natl. Acad. Sci. USA* **94,** 13028–13033.

Fitch, W. M. & Upper, K. (1987) The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* **52,** 759–767.

Fitz-Gibbon, S. T. & House, C. H. (1999) Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic Acids Res.* **27,** 4218–4222.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., *et al.* (1996) Life with 6000 genes. *Science* **274,** 546, 563–567.

Gogarten, J. P., Murphey, R. D. & Olendzenski, L. (1999) Horizontal gene transfer: Pitfalls and promises. *Biol. Bull.* **196,** 359–361.

Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. (2002) Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* **19,** 2226–2238.

Gray, M. W. (1999) Evolution of organellar genomes. *Curr. Opin. Genet. Dev.* **9,** 678–687.

Gupta, R. S. (1998) Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.* **62,** 1435–1491.

Gupta, R. S. & Singh, B. (1994) Phylogenetic analysis of 70-Kd heat-shock protein sequences suggests a chimeric origin for the eukaryotic cell-nucleus. *Curr. Biol.* **4,** 1104–1114.

Gupta, R. S., Aitken, K., Falah, M. & Singh, B. (1994) Cloning of giardia-lamblia heat-shock protein Hsp70 homologs—implications regarding origin of eukaryotic cells and of endoplasmic-reticulum. *Proc. Natl. Acad. Sci. USA* **91,** 2895–2899.

Halanych, K. M., Bacheller, J. D., Aguinaldo, A. M. A., Liva, S. M., Hillis, D. M. & Lake, J. A. (1995) Evidence from 18s ribosomal DNA that the lophophorates are protostome animals inarticulate. *Science* **267,** 1641–1643.

Hentze, M. W. (2001) Protein synthesis—believe it or not—translation in the nucleus. *Science* **293,** 1058–1059.

Horiike, T., Hamada, K., Kanaya, S. & Shinozawa, T. (2001) Origin of eukaryotic cell nuclei by symbiosis of Archaea in baceria is revealed by homology-hit analysis. *Nat. Cell Biol.* **3,** 210–214.

House, C. H. & Fitz-Gibbon, S. T. (2002) Using homolog groups to create a whole-genomic tree of free-living organisms: An update. *J. Mol. Evol.* **54,** 539–547.

Jain, R., Rivera, M. C. & Lake, J. A. (1999) Horizontal gene transfer among genomes: The complexity hypothesis. *Proc. Natl. Acad. Sci. USA* **96,** 3801–3806.

Jain, R., Rivera, M. C., Moore, J. E. & Lake, J. A. (2003) Horizontal gene transfer accelerates genome innovation and evolution. *Mol. Biol. Evol.* **20,** 1598–1602.

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., *et al.* (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis sp.* strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3,** 109–136.

Karlin, S., Mrazek, J. & Campbell, A. M. (1997) Compositional biases of bacterial genomes and evolutionary implications. *J. Bacteriol.* **179,** 3899–3913.

Karol, K. G., McCourt, R. M., Cimino, M. T. & Delwiche, C. F. (2001) The closest living relatives of land plants. *Science* **294,** 2351–2353.

Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., *et al.* (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390,** 364–370.

Koonin, E. V., Mushegian, A. R., Galperin, M. Y. & Walker, D. R. (1997) Comparison of archaeal and bacterial genomes: Computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Mol. Microbiol.* **25,** 619–637.

Koonin, E. V., Makarova, K. S. & Aravind, L. (2001) Horizontal gene transfer in prokaryotes: Quantification and classification. *Annu. Rev. Microbiol.* **55,** 709–742.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., *et al*. (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390,** 249–256.

Lake, J. A. (1988) Origin of the eukaryotic nucleus determined by rate-invariant analysis of ribosomal RNA sequences. *Nature* **331,** 184–186.

Lake, J. A. & Rivera, M. C. (1994) Was the nucleus the 1st endosymbiont. *Proc. Natl. Acad. Sci. USA* **91,** 2880–2881.

Lake, J. A. & Rivera, M. C. (2004) Deriving the genomic tree of life in the presence of horizontal gene transfer: Conditioned Reconstruction. *Mol. Biol. Evol.* **21,** 681–690.

Lake, J. A., Henderson, E., Clark, M. W. & Matheson, A. T. (1982) Mapping evolution with ribosome structure: Intralineage constancy and interlineage variation. *Proc. Natl. Acad. Sci. USA* **79,** 5948–4952.

Lawrence, J. G. & Ochman, H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA* **95,** 9413–9417.

Lecompte, O., Ripp, R., Puzos-Barbe, V., Duprat, S., Heilig, R., Dietrich, J., Thierry, J. C. & Poch, O. (2001) Genome evolution at the genus level: Comparison of three complete genomes of hyperthermophilic Archaea. *Genome Res.* **11,** 981–993.

Maeder, D. L., Weiss, R. B., Dunn, D. M., Cherry, J. L., Gonzalez, J. M., DiRuggiero, J. & Robb, F. T. (1999) Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and *P. horikoshii* inferred from complete genomic sequences. *Genetics* **152,** 1299–1305.

Makarova, K. S., Aravind, L., Galperin, M. Y., Grishin, N. V., Tatusov, R. L., Wolf, Y. I. & Koonin, E. V. (1999) Comparative genomics of the archaea (*Euryarchaeota*): Evolution of conserved protein families, the stable core, and the variable shell. *Genome Res.* **9,** 608–628.

Margulis, L. (1970) *Origin of the Eukaryotic Cells* (Yale Univ. Press, New Haven, CT).

Martin, W. & Muller, M. (1998) The hydrogen hypothesis for the first eukaryote. *Nature* **392,** 37–41.

Martin, W., Mustafa, A. Z., Henze, K. & Schnarrenberger, C. (1996) Higher-plant chloroplast and cytosolic fructose-1,6-bisphosphatase isoenzymes: Origins via duplication rather than prokaryote-eukaryote divergence. *Plant Mol. Biol.* **32,** 485–491.

Montague, M. G. & Hutchison, C. A. (2000) Gene content phylogeny of herpesviruses. *Proc. Natl. Acad. Sci. USA* **97,** 5334–5339.

Moreira, D. & Lopez-Garcia, P. (1998) Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. *J. Mol. Evol.* **47,** 517–530.

Nickrent, D. L., Parkinson, C. L., Palmer, J. D. & Duff, R. J. (2000) Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Mol. Biol. Evol.* **17,** 1885–1895.

Nolling, J., Breton, G., Omelchenko, M. V., Makarova, K. S., Zeng, Q. D., Gibson, R., Lee, H. M., Dubois, J., Qiu, D. Y., Hitti, J., *et al.* (2001) Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* **183,** 4823–4838.

Ochman, H. (2001) Lateral and oblique gene transfer. *Curr. Opin. Genet. Dev.* **11,** 616–619.

Penny, D. & Poole, A. (1999) The nature of the last universal common ancestor. *Curr. Opin. Genet. Dev.* **9,** 672–677.

Philippe, H. & Forterre, P. (1999) The rooting of the universal tree of life is not reliable. *J. Mol. Evol.* **49,** 509–523.

Poole, A., Jeffares, D. & Penny, D. (1999) Early evolution: prokaryotes, the new kids on the block. *BioEssays* **21,** 880–889.

Pryer, K. M., Schneider, H., Smith, A. R., Cranfill, R., Wolf, P. G., Hunt, J. S. & Sipes, S. D. (2001) Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* **409,** 618–622.

Pryer, K. M., Schneider, H., Zimmer, E. A. & Banks, J. A. (2002) Deciding among green plants for whole genome studies. *Trends Plant Sci.* **7,** 550–554.

Raymond, J., Zhaxybayeva, O., Gogarten, J. P., Gerdes, S. Y. & Blankenship, R. E. (2002) Whole-genome analysis of photosynthetic prokaryotes. *Science* **298,** 1616–1620.

Ribeiro, S. & Golding, G. B. (1998) The mosaic nature of the eukaryotic nucleus. *Mol. Biol. Evol.* **15,** 779–788.

Rivera, M. C. & Lake, J. A. (1992) Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257,** 74–76.

Rivera, M. C. & Lake, J. A. (2004) The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431,** 152–155.

Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. (1998) Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. USA* **95,** 6239–6244.

Schwarz, Z. & Kossel, H. (1980) Primary structure of 16s rDNA from *Zea-mays* chloroplast is homologous to *Escherichia-coli* 16s ribosomal-RNA. *Nature* **283,** 739–742.

Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H. M., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., *et al.* (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum Delta H*: Functional analysis and comparative genomics. *J. Bacteriol.* **179,** 7135–7155.

Snel, B., Bork, P. & Huynen, M. A. (1999) Genome phylogeny based on gene content. *Nat. Genet.* **21,** 108–110.

Soltis, P. S., Soltis, D. E., Wolf, P. G., Nickrent, D. L., Chaw, S. & Chapman, R. L. (1999) The phylogeny of land plants inferred from 18S rDNA sequences: Pushing the limits of rDNA signal? *Mol. Biol. Evol.* **16,** 1774–1784.

Stechmann, A. & Cavalier-Smith, T. (2002) Rooting the eukaryote tree by using a derived gene fusion. *Science* **297,** 89–91.

Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, F., Hirama, C., Nakamura, Y., Ogasawara, N., *et al.* (2000) Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res.* **28,** 4317–4331.

Tekaia, F., Lazcano, A. & Dujon, B. (1999) The genomic tree as revealed from whole proteome comparisons. *Genome. Res.* **9,** 550–557.

Whitman, W. B., Coleman, D. C. & Wiebe, W. J. (1998) Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci. USA* **95,** 6578–6583.

Woese, C. R., Pace, N. R. & Olsen, G. J. (1986) Are arguments against archaebacteria valid. *Nature* **320,** 401–402.

Woese, C. R., Kandler, O. & Wheelis, M. L. (1990) Towards a natural system of organisms: Proposal for the domains *Archaea, Bacteria* and *Eucarya*. *Proc. Natl. Acad. Sci. USA* **87,** 4576–4579.

Yang, S., Doolittle, R. F. & Bourne, P. E. (2005) Phylogeny determined by protein domain content. *Proc. Natl. Acad. Sci. USA* **102,** 373–378.

15

Prospects for Identifying Functional Variation Across the Genome

STUART J. MACDONALD* AND ANTHONY D. LONG*

The genetic factors contributing to complex trait variation may reside in regulatory, rather than protein-coding portions of the genome. Within noncoding regions, SNPs in regulatory elements are more likely to contribute to phenotypic variation than those in nonregulatory regions. Thus, it is important to be able to identify and annotate noncoding regulatory elements. DNA conservation among diverged species successfully identifies noncoding regulatory regions. However, because rapidly evolving regulatory regions will not generally be conserved across species, these will not detected by using purely conservation-based methods. Here we describe additional approaches that can be used to identify putative regulatory elements via signatures of nonneutral evolution. An examination of the pattern of polymorphism both within and between populations of *Drosophila melanogaster*, as well as divergence with its sibling species *Drosophila simulans*, across 24.2 kb of noncoding DNA identifies several nonneutrally evolving regions not identified by conservation. Because different methods tag different regions, it appears that the methods are complementary. Patterns of variation at different elements are consistent with the action of selective sweeps, balancing selection, or population dif-

*Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697-2525.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AY863438–AY864021).

ferentiation. Together with regions conserved between *D. melanogaster* and *Drosophila pseudoobscura*, we tag 5.3 kb of noncoding DNA as potentially regulatory. Ninety-seven of the 408 common noncoding SNPs surveyed are within putatively regulatory regions. If these methods collectively identify the majority of functional noncoding polymorphisms, genotyping only these SNPs in an association mapping framework would reduce genotyping effort for noncoding regions 4-fold.

A major goal of modern biological research is to understand the relationship between genotype and phenotype. The search for genetic variation contributing to differences among individuals is exemplified by association studies that aim to identify those segregating genetic polymorphisms that confer risk to common polygenic, or complex diseases in humans (Carlson *et al.*, 2004). Association mapping involves genotyping a dense set of SNPs in a large population of individuals, and asking whether there is evidence of an association between the genotype at each SNP and the phenotype. A significant association suggests that the genotyped SNP is either itself responsible for conferring disease risk or strongly correlated, i.e., is in linkage disequilibrium, with the causal site. We refer to SNPs contributing to phenotypic variation as functional SNPs (fSNPs).

The human genome harbors 4.6–7.1 million common SNPs [minor allele frequency above 5%; Kruglyak and Nickerson (2001) and Stephens *et al.* (2001)], with the vast majority presumed to be nonfunctional. Unfortunately, it is not yet cost-effective to exhaustively test every SNP for an association with a disease phenotype. Despite a great deal of academic and private research, genotyping technology remains unable to efficiently genotype millions of SNPs in thousands of individuals at reasonable cost (Syvänen, 2001). Thus, some intelligent way of reducing the genotyping effort is needed.

One such method, the HapMap project (International HapMap Consortium, 2003), seeks to take advantage of the level of linkage disequilibrium (LD) across the genome and choose a subset of SNPs to genotype that explain the majority of haplotype information. This approach is favored for humans, with the recent suggestion that the genome exhibits a block-like LD structure (Daly *et al.*, 2001; Gabriel *et al.*, 2002; Patil *et al.*, 2001). Under the HapMap plan, between 200,000 and 1 million SNPs need to be genotyped to achieve complete genome coverage (Carlson *et al.*, 2003; Gabriel *et al.*, 2002; Goldstein *et al.*, 2003; Patil *et al.*, 2001). However, this plan is critically dependent on the degree to which available SNPs capture human haplotype diversity, which is hotly debated (Carlson *et al.*, 2003; Reich *et al.*, 2003), and on the reliability of the block definitions

across different populations, which is also unclear (Gabriel *et al.*, 2002). Perhaps a more fundamental difficulty with this methodology is that haplotypes do not cause disease. Finding an association to a haplotype block is not an endpoint, it merely delimits the search, and further genotyping is required to finally identify the causal mutation.

An alternative strategy to reduce total genotyping effort is to genotype the subset of SNPs most likely to contribute to the examined phenotype. In a seminal paper, Risch and Merikangas (1996) showed that association studies for complex traits have higher power than linkage mapping approaches, and the paper is widely cited as supporting the use of association mapping. However, an important aspect of the theoretical treatment put forward by Risch and Merikangas (1996) is often overlooked: the actual disease-causing site must be one of the sites genotyped. The power of association studies is greatly reduced if the causative site is not among those genotyped (Kruglyak, 1999; Long and Langley, 1999).

Based on data acquired from analyses of Mendelian diseases, Botstein and Risch (2003) have suggested that causal polymorphisms may generally be coding, which immediately suggests a strategy for selecting putatively disease-causing SNPs on which to focus: identify and genotype all SNPs in coding regions. This approach would ensure a large reduction in total genotyping effort, and provided complex traits are somewhat similar to Mendelian traits in their genetic architecture is likely to uncover some fraction of phenotypically relevant genetic variation. Nevertheless, some clear examples of genetic factors underlying complex trait variation suggest that the responsible polymorphisms may reside in regulatory regions (Robin *et al.*, 2002; Shapiro *et al.*, 2004; Ueda *et al.*, 2003). The strategy suggested by Botstein and Risch (2003) will be undermined if variation in complex traits is generally determined by regulatory genetic variants.

Methods that allow us to identify functional noncoding regulatory domains, such as promoters or enhancers capable of modulating spatial and temporal gene expression, would enable SNPs to be classified based on their position relative to these domains. Genotyping only those SNPs present within regulatory domains would allow for a reduction in the total genotyping effort in association studies. Such a strategy is simple in principal, but it is a major challenge to sift through the ocean of noncoding DNA to find those polymorphisms that are truly cis-regulatory in function. In *Drosophila melanogaster*, the amount of noncoding DNA is 95.9 megabases (Mb), or ≈80% of the euchromatic genome (Adams *et al.*, 2000). In humans, the disparity between coding and noncoding DNA is more extreme, with 2,817 Mb of noncoding DNA representing 98.8% of the genome (International Human Genome Sequencing Consortium, 2004). It is possible that statistical tests coopted from the fields of population ge-

netics and molecular evolution can be adapted to identify regulatory regions of the noncoding genome: if a region can be shown to have evolved in a nonneutral manner, presumably there are functional elements buried in these regions. Statistical tests are available to explicitly test for evidence of selection in coding regions (e.g., McDonald and Kreitman, 1991; Nielsen and Yang, 1998), but these tests cannot be applied to noncoding DNA because they rely on the ability to parse the sequence into synonymous and nonsynonymous sites.

Here we examine several statistics that can be applied to noncoding DNA to detect regions of sequence subject to the action of past natural selection: conservation between phylogenetically diverged species, the ratio of polymorphism to divergence between sibling species, the polymorphism frequency spectrum, and the level of population structure. We explore graphical sliding window presentations (Kreitman and Hudson, 1991) of these statistics, because our goal is to suggest regions likely to harbor fSNPs rather than to apply rigorous statistical tests. Such graphical, sliding-window tests are also more easily generalized to genomescale data. Compared to SNPs in regions that do not show evidence for past natural selection, those in regions showing departures from neutral expectation are stronger candidates for fSNPs. Because nonneutrally evolving regions are likely to be enriched for fSNPs, a reduction in genotyping effort could be achieved in association studies by preferentially genotyping SNPs from nonneutrally evolving regions.

We select 26 ≈1-kb fragments of primarily noncoding DNA near known genes distributed across the *D. melanogaster* genome. We examine the rate at which the various proposed tools are capable of ''tagging'' potential regulatory elements in these regions, and also determine the degree to which the statistics tag the same or different areas as nonneutrally evolving. Because we chose the noncoding regions for this study randomly with respect to cis-regulatory annotation, the rate at which we tag potential regulatory elements is likely typical of the genome as a whole. The tests we propose could be applied to any noncoding sequence. Proving that tagged regions are cis-regulatory elements harboring fSNPs is a more difficult problem that remains to be addressed.

MATERIALS AND METHODS

Sequenced Regions

We chose 26 loci distributed evenly with respect to genetic location along the five major chromosome arms of *D. melanogaster* (Table 15.1). These loci fall into three categories: those known to interact with the *Notch* signaling pathway (seven genes), those thought to affect development of

TABLE 15.1 Details of the Loci Examined

Gene Name	Gene Symbol	Gene Position	Amplicon Position	Functional Category (ref.)
deltex	*dx*	X, 17.0	−553	Notch
cut	*ct*	X, 20.0	+3870	PNS
dishevelled	*dsh*	X, 34.5	−385	Notch
scalloped	*sd*	X, 51.5	−2441	PNS
Beadex	*Bx*	X, 59.4	−30244	PNS (Norga *et al.*, 2003)
split ends	*spen*	2L, 0.5	−2652	PNS (Kuang *et al.*, 2000)
friend of echinoid	*fred*	2L, 11.5	−730	PNS
wingless	*wg*	2L, 21.9	+34125	PNS (Ramain *et al.*, 2001)
numb	*numb*	2L, 35.5	−13229	Notch
daughterless	*da*	2L, 41.3	−967	PNS
deadpan	*dpn*	2R, 57.5	−1709	PNS (Norga *et al.*, 2003; Bier *et al.*, 1992)
scabrous	*sca*	2R, 66.7	−768	PNS
mastermind	*mam*	2R, 70.3	−18725	Notch
cousin of atonal	*cato*	2R, 79.5	−635	PNS
smooth	*sm*	2R, 91.5	+3796	PNS (Lage *et al.*, 1997)
Distal-less	*Dll*	2R, 107.8	−808	—
extra macrochaetae	*emc*	3L, 0.0	−407	PNS
vein	*vn*	3L, 16.2	−2767	—
quemao	*qm*	3L, 23.0	+254	PNS (Lai *et al.*, 1998)
Bearded	*Brd*	3L, 42.0	+392	Notch
neuralized	*neur*	3R, 48.5	−1381	PNS
Actin 88F	*Act88F*	3R, 57.1	−1062	—
Hairless	*H*	3R, 69.5	−499	Notch
pointed	*pnt*	3R, 79.0	−4659	PNS
Serrate	*Ser*	3R, 92.0	−722	Notch
tramtrack	*ttk*	3R, 102.0	−2030	PNS

NOTE: Gene position is the chromosome arm on which the gene resides, followed by its genetic position. Amplicon position is the distance between the midpoint of the amplicon and the gene start codon in bp (+, base pair is upstream of the start codon; −, base pair is downstream). Functional categories were determined by using the Gene Ontology (www.geneontology.org) unless references are provided. Notch, these genes functionally interact with the *Notch* signaling pathway (Gene Ontology terms GO:0007219, GO:0030179, and GO:0005112); PNS, these genes are involved in peripheral nervous system and sensory organ development, or bristle morphogenesis (GO:0007422, GO:0007423, and GO:0008407). — indicates that genes are unlikely to have any involvement in neurogenesis.

the peripheral nervous system or that have been shown to have quantitative effects on bristle number (16 genes), and finally three genes selected to ensure coverage of the genome. Primers were developed for an ≈1-kb amplicon at each locus using *Primer3* (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi; sequences are provided in Table 2, which is published as supporting information on the PNAS web site). None of the developed amplicons appear to encompass known regulatory elements.

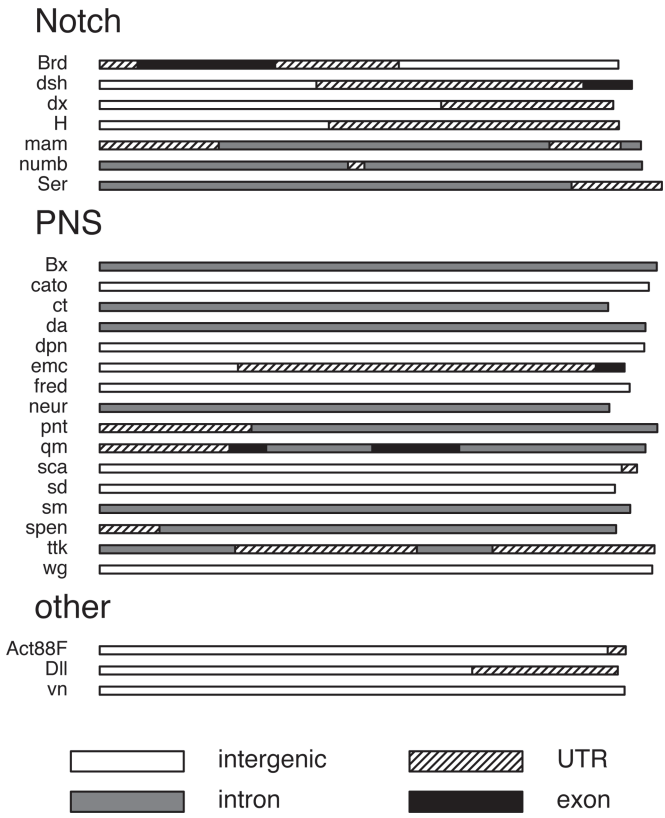


FIGURE 15.1 The type of DNA sequence surveyed. Each of the 26 amplicons is referred to by the symbol for the closest known gene, and amplicons are grouped according to functional category (see Table 15.1 for full gene names and a description of the categorization). The amplicons are each represented by a bar, scaled to the length of the *D. melanogaster* alignment, and shaded to reflect the *D. melanogaster* release 4.0 genome annotation.

Fig. 15.1 highlights the annotated regions sequenced for each amplicon (taken from Release 4.0 of the *D. melanogaster* genome sequence), Table 15.1 documents the position of the amplicon relative to the start codon of the gene, and Fig. 3, which is published as supporting information on the PNAS web site, details the exact positions of the amplicons relative to the structure of the loci.

D. melanogaster Stocks

All 26 amplicons were sequenced for 16 wild-type lines representing a worldwide sample. The stock numbers for these lines are: B1 (Canton, OH), B3839 (Bermuda), B3841 (Bogata, Colombia), B3844 (Barcelona, Spain), B3846 (Capetown, South Africa), B3852 (Koriba Dam, South Africa), B3853 (Koriba Dam, South Africa), B3864 (Israel), B3870 (Riverside, CA), B3875 (Athens), B3886 (Red Top Mountain, GA), T14021-0231.0 (Oahu, Hawaii), T14021-0231.1 (Ica, Peru), T14021–0231.4 (Kuala Lumpur, Malaysia), T14021-0231.6 (Mysore, India), and T14021-0231.7 (Ken-ting, Taiwan), where ''B'' and ''T'' refer to the Bloomington and Tucson *Drosophila* stock centers, respectively. Before sequencing, the 16 lines were propagated by using single male–female pairs for between 2 and 12 generations to reduce heterozygosity.

In addition, for each amplicon, we sequenced eight strains from a single population. For the X- and third-chromosome amplicons, we sequenced eight chromosomal extraction strains, where the natural alleles were derived from Napa Valley, CA, whereas for amplicons on the second chromosome, we sequenced eight inbred lines derived from North Carolina (kindly provided by C. H. Langley, Center for Population Biology, University of California, Davis).

Outgroup Sequences

Using shotgun sequencing assemblies provided by the Genome Sequencing Center, Washington University Medical School (http://genome.wustl.edu/projects/simulans), we obtained the homologous region in *Drosophila simulans* from one of the strains, sim4, sim6 or w501 with BLASTN, for each amplicon. We used a similar procedure to identify the homologous region for each amplicon from the *Drosophila pseudoobscura* genome assembly (release 1.03), taken as a 4-kb window centered on the position of the best BLASTN hit. Details of the regions extracted from these outgroup species are provided in Table 3, which is published as supporting information on the PNAS web site.

Sequence and Population Genetics Analyses

Sequence traces for each *D. melanogaster* strain/amplicon combination were assembled by using SEQMANII (version 5.01, DNASTAR), and for each amplicon, the *D. melanogaster* and *D. simulans* sequences were manually aligned by using BIOEDIT (www.mbio.ncsu.edu/BioEdit/bioedit.html). All *D. melanogaster* sequences were deposited in the GenBank database (accession nos. AY863438–AY864021).

After alignment, each amplicon was represented by at least six within-population *D. melanogaster* lines, and at least 13 worldwide *D. melanogaster* lines. Missing sequences were due largely to repeated PCR failures, but were also due to ambiguous sequence reads caused by heterozygous insertion/deletion polymorphisms that remained in some of the worldwide and North Carolina lines despite inbreeding. Twelve of the 26 amplicons showed between one and four sequences harboring at least one heterozygous nucleotide, and before analysis, on a per amplicon basis, each heterozygous sequence was arbitrarily split into a pair of pseudohaplotypes. This split is justified because PCR was performed on DNA extracted from single males, so the heterozygous sequence reflects the presence of two alleles. None of the diversity measures we estimate are affected by the phase of the polymorphism data.

Using a sliding window approach with a window size of 250 bp, stepping through each sequence alignment in 1-bp increments, we estimated (*i*) nucleotide diversity (π) across the *D. melanogaster* sequences, (*ii*) divergence (*K*) between *D. melanogaster* and *D. simulans*, (*iii*) Tajima's *D* (Tajima, 1989), which provides a measure of the polymorphism frequency spectrum, (*iv*) π_w estimated from the alleles obtained from the single *D. melanogaster* population (either Napa Valley or North Carolina), and (*v*) π_b estimated from the worldwide *D. melanogaster* samples. A comparison of π_w and π_b serves as a proxy for population structure in that differences between within- and among-population nucleotide diversity can be assessed. Sites segregating for more than two alleles were ignored for all calculations, with window size kept constant with respect to the remaining informative sites. Missing data and gaps were treated as a reduction in the sample size, and values were weighted accordingly. All analyses were performed by using custom scripts in the statistical programming language R (www.r-project.org).

Finally, we extracted the consensus sequence for each *D. melanogaster* alignment and used a sliding-window approach to BLAST 31-bp sections against the homologous region of the *D. pseudoobscura* genome, stepping through the consensus sequence in 1-bp increments. For each *D. melanogaster* query sequence, we recorded the position, orientation, and score of the highest BLAST hit in *D. pseudoobscura*, and considered only hits with a score ≈45 in further analyses.

RESULTS

We sequenced 26 ≈1-kb amplicons in *D. melanogaster*, primarily from noncoding regions in or near genes involved in peripheral nervous system development and/or regulation of *Notch* signaling. For each amplicon, we also identified the homologous region from the closely re-

lated *D. simulans* species, and from *D. pseudoobscura*, which is thought to have diverged from *D. melanogaster* ≈25 million years ago (Russo *et al.*, 1995). The degree to which studied amplicons harbor cis-regulatory elements is unknown. These data allowed us to examine a set of sequence attributes across each of the amplicons to examine for regions exhibiting nonneutral evolution: (*i*) the level of sequence conservation between *D. melanogaster* and *D. pseudoobscura*, (*ii*) the amount of nucleotide polymorphism within *D. melanogaster* relative to the level of divergence between *D. melanogaster* and its sibling species *D. simulans*, (*iii*) the polymorphism frequency spectrum in *D. melanogaster*, and (*iv*) the amount of population structure within *D. melanogaster*, by comparing the nucleotide diversity within a single *D. melanogaster* population to the diversity observed in a worldwide panel. Because the footprint of selection may be small, a sliding-window framework is likely to be more informative than examining the average values of the statistics for each amplicon (see Table 4, which is published as supporting information on the PNAS web site). Fig. 15.2 shows the sliding-window analyses for six selected amplicons, and Fig. 4, which is published as supporting information on the PNAS web site, presents analyses for all amplicons. Below we document those sequenced noncoding regions that have patterns in the sliding window plots suggesting deviation from neutral expectation, and also note the number of SNPs present within such regions.

Deep Sequence Conservation

Random neutral mutation will tend to erode similarity between neutrally evolving sequences in independent lineages. Thus, conservation of DNA sequence across taxa diverged by many millions of years is taken as evidence of function, as such regions are presumed to be subject to negative, or purifying selection to preserve sequence. This has become a guiding principle in the detection of functional noncoding DNA (Berman *et al.*, 2004; Boffelli *et al.*, 2003; Hong *et al.*, 2003; Kellis *et al.*, 2003).

Nine of the 26 amplicons show no fine-scale conservation using our BLAST approach [*Bx*, *da* (Fig. 15.2*B*), *dsh*, *mam*, *sca* (Fig. 15.2*E*), *sd*, *sm*, *ttk*, and *vn*], 10 show low conservation [*Brd*, *cato* (Fig. 15.2*A*), *dpn*, *dx*, *fred*, *H*, *neur*, *numb*, *pnt* (Fig. 15.2*C*), and *spen*; defined as showing three or fewer short (<60-bp) stretches of conservation], and 7 show high conservation [*Act88F*, *ct*, *Dll*, *emc*, *qm* (Fig. 15.2*D*), *Ser* (Fig. 15.2*F*), and *wg*]. In two of the amplicons with high conservation, *qm* (Fig. 15.2*D*) and *emc*, the regions of conservation map to known exons. Overall, of the 24.2 kb of sequenced noncoding DNA in *D. melanogaster*, 2.1 kb (8.6%) is highly conserved between *D. melanogaster* and *D. pseudoobscura*, suggesting that it may have regulatory significance. There are 408 common (>5% minor allele fre-

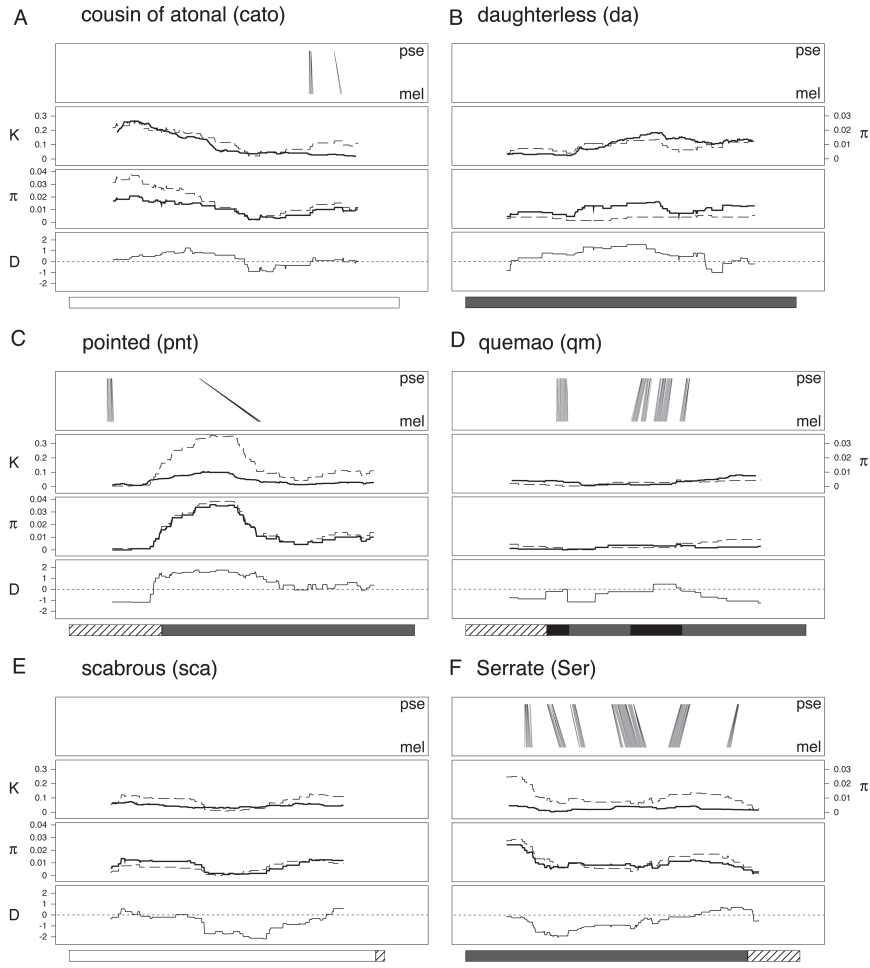


FIGURE 15.2 Signatures of selection across sequenced amplicons. Six of the 26 amplicons are detailed, and each is composed as follows. (*Top*) Conservation between *D. melanogaster* and *D. pseudoobscura*. Each line represents a BLAST hit (with a score >45) between a 31-bp subsection of the *D. melanogaster* consensus sequence and the homologous sequence from *D. pseudoobscura*, with the endpoints of each line showing the position of the hit in each genome. All BLAST hits are between subsequences in the same orientation. (*Upper Middle*) Nucleotide diversity (π) within *D. melanogaster* (dashed line), and divergence (K) between *D. melanogaster* and *D. simulans* (solid line). (*Lower Middle*) Nucleotide diversity within the lines derived from the single *D. melanogaster* population (dashed line), and within the worldwide panel of *D. melanogaster* strains (solid line). (*Bottom*) Tajima's *D* statistic (Tajima, 1989). Below each figure is an annotation bar describing the type of sequence surveyed for each amplicon, with the shading as described in Fig. 15.1.

quency) biallelic SNPs in the 24.2 kb of noncoding sequence, and 14 (3.4%) are present within the detected conserved regions.

Our BLAST approach reveals that *D. melanogaster* and *D. pseudoobscura* do not appear to differ by any conserved microinversions, as all strong BLAST hits are between subsequences in the same orientation, or by any conserved local rearrangements, because none of the hit lines cross. This finding is in accordance with previous results at the *Enhancer of split* locus (Macdonald and Long, 2005). However, we did observe at least three cases where there appears to have been a large insertion/deletion in one of the two genomes [*Act88F*, *ct*, and *pnt* (Fig. 15.2*C*)].

Patterns of Neutral Evolution

For the remaining analyses not involving *D. pseudoobscura*, sliding-window plots for 13 amplicons show no noticeable departure from the pattern of diversity predicted by neutral theory [*Brd*, *Bx*, *ct*, *Dll*, *dpn*, *dsh*, *emc*, *H*, *mam*, *qm* (Fig. 15.2*D*), *sm*, *spen*, and *ttk*]. Several criteria suggest the absence of recent, detectable selective forces acting on these regions. Within-species diversity (π) and between-species divergence (K) generally track each other, indicating no change in mutational processes between species. There are also no obvious differences between the nucleotide diversity within the single *D. melanogaster* population, and diversity across the worldwide sample of *D. melanogaster* lines, implying that no populationspecific forces are at work. Finally, for these 13 amplicons we see no clear departure from the allele frequency distribution predicted under neutrality as measured by Tajima's D statistic (Tajima, 1989).

Positive Selection

A low level of diversity coupled with a frequency spectrum skewed toward an excess of rare variants (i.e., negative Tajima's D) is generally taken as evidence for a selective sweep or positive selection (Andolfatto and Przeworski, 2001; Kim and Stephan, 2002). A selective sweep removes variation around the advantageous mutation, and observed polymorphisms are rare having arisen since the sweep.Weidentify three amplicons, *neur*, *sca* (Fig. 15.2*E*), and *sd*, showing patterns of diversity and divergence consistent with the action of a weak selective sweep. The amplicon upstream of the *sca* gene represents a particularly clear example (Fig. 15.2*E*). In a central 200-bp section of this amplicon there is a marked dip both in the level of nucleotide diversity and in Tajima's D, suggesting that a site within this short region has swept to fixation in *D. melanogaster*. A similar pattern is observed for the amplicon in an intronic region of the *neur* gene: reduced π and negative D for the second half of the amplicon.

In contrast, the entire amplicon upstream of the *sd* gene shows very low nucleotide polymorphism (just four polymorphisms exist, three of which are singletons), whereas interspecific divergence is normal, suggesting that the entire sequenced region has been impacted by a positive selection event. We estimate that 1.4 kb (5.8%) of the sequenced noncoding DNA in *D. melanogaster* has been impacted by positive selection, and these regions collectively harbor two common biallelic SNPs (0.5% of the total common SNPs discovered).

Balancing Selection

Balanced polymorphisms are segregating sites maintained in a population at intermediate frequency due to heterozygote advantage, frequency-dependent selection, by selection on alternate alleles in different environments, or by antagonistic pleiotropy. A balanced polymorphism can theoretically be maintained indefinitely and will enhance the level of neutral polymorphism surrounding it, with the size of the affected region dependent on the local recombination rate. Thus, the presence of a balanced polymorphism will generate a high level of diversity compared to divergence, and a greater number of frequent polymorphisms (i.e., positive Tajima's *D*). Three amplicons, *Act88F*, *dx*, and *pnt* (Fig. 15.2*C*), exhibit patterns suggestive of balancing selection.

The best example is provided by the amplicon in a 5′ UTR/intronic region of the *pnt* gene (Fig. 15.2*C*), where starting at the transition between 5′ UTR and intron, and continuing within the intron for ≈300 bp, the level of nucleotide diversity is very high, and *D* is positive. It is of interest that the affected region may represent a previously uncharacterized insertion relative to *D. pseudoobscura*. The amplicon about the *dx* gene is around one-third 5′ UTR, and for about 200 bp upstream of the 5′ UTR the level of nucleotide diversity is high relative to divergence, and *D* is positive. Soon after the start of the transcribed region, diversity returns to lower values, and *D* falls to its neutral expectation of zero.

The amplicon upstream of the *Act88F* gene also exhibits a pattern consistent with balancing selection for the first ≈300 bp. However, this amplicon is also noteworthy for a single sequence from the Napa Valley *D. melanogaster* population that has a unique haplotype. The presence of this sequence in the *D. melanogaster–D. simulans* alignment generates 48 singleton polymorphic sites and substantial insertion/deletion variation, such that particularly in the central portion of the *Act88F* amplicon, nucleotide diversity is high, and Tajima's *D* is negative. In comparison, analyses based on an alignment lacking the aberrant *Act88F* sequence show a Tajima's *D* and nucleotide diversity not inconsistent with neutrality, although the signature of balancing selection at the start of the amplicon

remains. Because this unique *D. melanogaster* haplotype is not similar to the *D. simulans* sequence, it is unclear whether it represents a single event or the aftermath of a series of mutational events. Rare, extremely diverged haplotypes perhaps deserve special treatment.

The amount of noncoding DNA sequenced in *D. melanogaster* that shows a pattern of nucleotide diversity consistent with balancing selection is 0.8 kb (3.4%), and these regions harbor 38 common biallelic SNPs (9.3% of the common SNPs identified in the survey).

Population Structure

Two types of *D. melanogaster* populationspecific effects are evident from our sequenced amplicons. The first type is when the single population shows lower sequence variation than does the worldwide panel. This observation is indicative of a geographically localized reduction in diversity, possibly via local adaptation. Two of the amplicons show this pattern, the central 300 bp of the intronic region sequenced for the *da* gene (Fig. 15.2*B*), and the end of the amplicon upstream of the *fred* gene. The second pattern is the reverse, where there is less variation in the worldwide sample than would be predicted based on variation within the single population. The maintenance of higher variation within a single population than across multiple populations is potentially the result of balancing selection. This pattern is apparent for the 300 bp at the end of the amplicon upstream of the *vn* gene, and for the 400 bp at the start of the amplicon upstream of the *cato* gene (Fig. 15.2*A*). Together the two patterns highlighting population structure within *D. melanogaster* encompass 1.0 kb (3.9%) of the noncoding sequence and hold 43 (10.5%) of the common biallelic SNPs uncovered in our survey.

Unexpected Patterns

Finally, two 5′ UTR/intronic amplicons, within the genes *Ser* (Fig. 15.2*F*) and *numb*, and a single amplicon downstream of the *wg* gene, show higher nucleotide polymorphism than expected given the level of sequence divergence between *D. melanogaster* and *D. simulans*. However, in no case is this accompanied by a coordinated skew in the polymorphism frequency spectrum. These three amplicons imply that, as we collect larger DNA sequence data sets from a range of sequence types, we are likely to see patterns of polymorphism that neither conform to neutral expectation nor neatly fit with our current ideas about the expected result of selective events.

DISCUSSION

There is considerable interest in developing methods to identify functional domains from primary sequence data. One goal is to detect regions likely to harbor fSNPs that contribute to intraspecific phenotypic variation in complex traits. To identify such regions, we propose employing a series of tests based on population genetics theory, which should complement approaches based purely on deep phylogenetic conservation.

Conservation

Over evolutionary time, separately evolving taxa will accumulate random neutral mutations, and only regions under functional constraint will be conserved. Comparative genome sequencing has proved quite useful for both gene prediction and for identifying conserved noncoding regions (Kellis *et al.*, 2003), which in some instances have been shown to exhibit regulatory activity (Boffelli *et al.*, 2003; Hong *et al.*, 2003; Johnson *et al.*, 2004). In the present study, 8.6% of the noncoding sequence we surveyed was conserved between the diverged species *D. melanogaster* and *D. pseudoobscura*, distributed in short sections across 17 of the 26 amplicons. The 14 common SNPs located within these identified regions are candidate fSNPs.

We have previously demonstrated that regulatory elements within the *Enhancer of split* locus in *D. melanogaster* are often conserved in *D. pseudoobscura* (Macdonald and Long, 2005), suggesting that they retain a similar regulatory function in this species. However, a recent analysis of 142 bona fide regulatory elements showed that they were only 4–8% more conserved between *D. melanogaster* and *D. pseudoobscura* than were control regions (Richards *et al.*, 2005). These results imply that the signal of function in deep pairwise species comparisons may be both weak and heterogeneous across the genome. A further difficulty with relying completely on a conservation approach to functionally annotate a genome and identify fSNPs is that, although sequence conservation may imply function, a lack of conservation does not imply the absence of function. This was elegantly shown by Ludwig *et al.* (2000) for the *even-skipped* stripe 2 embryonic expression pattern in *Drosophila*. Here, the expression pattern itself is strongly conserved between the species *D. melanogaster* and *D. pseudoobscura*, whereas the regulatory region giving rise to the pattern is very different in sequence between the two species. Thus, true regulatory regions can be missed by using phylogenetic conservation. It is entirely possible that those cis-regulatory elements that contribute to within species variation for complex traits are fast evolving, and as a result are unlikely to be conserved in wide phylogenetic comparisons.

The 8.6% of noncoding *D. melanogaster* DNA tagged by conservation with *D. pseudoobscura* harbors just 3.4% of the common SNP variation in *D. melanogaster*. This lack of polymorphism might imply that highly conserved regions are too constrained to tolerate variation, and may actually be less likely to harbor fSNPs contributing to within-species phenotypic variation than less conserved regions identified by other means.

These concerns, coupled with the fact that conservation implies the action of a single form of selection (purifying), suggests that other methods of locating noncoding regulatory domains may be helpful.

Polymorphism and Divergence

The neutral theory of molecular evolution states that, for neutrally evolving DNA, the expected ratio of polymorphism within a species to divergence between species should be constant throughout the genome (Kimura, 1983). This expectation has a large variance, because of both the sampling and the particular genealogy of the tested region, but departures from neutrality can be detected, for instance with the widely applied HKA test (Hudson *et al.*, 1987).

A few clear cases of candidate regions associated with selective sweeps have been identified in *Drosophila*, for example the *Sdic* gene that encodes a subunit of the sperm axoneme (Nurminsky *et al.*, 1998), and the cytochrome P450 gene *Cyp6g1* associated with DDT resistance (Daborn *et al.*, 2002; Schlenke and Begun, 2004). Also, cases of balanced polymorphisms have been shown, such as that centered on the *Adh* fast/slow polymorphism in *D. melanogaster* (Kreitman and Hudson, 1991). However, in these instances, the magnitude of the population genetic signature was greater than those observed in our survey.

In this study, our goal is not to test for rigorous statistical significance, but instead to suggest regions that are likely to harbor fSNPs. We made use of a graphical approach (Kreitman and Hudson, 1991) allowing visual inspection of departures from neutrality across each of the 26 amplicons. Six amplicons exhibit patterns indicative of nonneutral evolution, with three suggesting past positive selection (selective sweeps) and three implicating a balanced polymorphism. It is possible that, despite modest power to detect nonneutral events, the magnitude of departure from neutrality based on the ratio of polymorphism to divergence is predictive of the likelihood of a region being regulatory in function. In this regard, we note that within known enhancer regions in the *Drosophila* locus *Enhancer of split*, using a test adapted from the McDonald–Kreitman (1991) and HKA tests (Hudson *et al.*, 1987), the ratio of polymorphism to divergence differs significantly between transcription factor binding sites and adjacent nonbinding sites ($P = 0.004$; Macdonald and Long, 2005).

Population Structure

Wright's *F* statistics (Wright, 1951) seek to partition allelic variation into within individual, within population, and between population components, and the F_{ST} statistic represents the degree of population differentiation. Under neutrality, the same level of population subdivision should be seen across the genome, but local adaptation can result in regional departures from this genome-wide expectation. In *Drosophila*, regions showing a strong departure can be quite small, from a single site to short regions of a few hundred base pairs. For instance, the *Adh* fast/slow polymorphism and Δ1 insertion/deletion polymorphism are both functional, and both show much stronger clinal variation across *D. melanogaster* populations than do neighboring polymorphisms in the same gene (Berry and Kreitman, 1993; Stam and Laurie, 1996).

Typically, F_{ST} is based on allele frequency estimates obtained from several subpopulations each consisting of number of individuals. Such an approach may be of limited use in genome-wide scans for fSNPs, as the economics of sequencing favors generating complete sequence data from a much more limited sample. Here, we use a proxy for standard measures of F_{ST} based on comparing the nucleotide diversity in a single population sample (within-population variance) to that across a set of lines of worldwide distribution (among-population variance). In the context of genome-wide scans, this approach may have greater utility than traditional measures of F_{ST} because it requires characterizing just 24 alleles. Using this approach, regions in four amplicons showed greater or lesser worldwide variation than expected based on the variation in a single population. We hypothesize that such regions are more likely to include fSNPs than regions showing no population subdivision.

Prospects for in Silico Functional Annotation

We surveyed 24.2 kb of noncoding DNA in *D. melanogaster*, encompassing 408 common SNPs, and identified putative regulatory regions using deep phylogenetic conservation (8.6% of the sequence, 3.4% of common SNPs), the pattern of positive selection (5.8% of sequence, 0.5% of SNPs), the pattern of balancing selection (3.4% of sequence, 9.3% of SNPs), and evidence for population structure (3.9% of sequence, 10.5% of SNPs). It is clear that the different tests identify different regions as potentially harboring fSNPs. This finding suggests that any one method may fail to annotate many functionally important areas, and at present it is premature to rely on a single method (such as deep phylogenetic conservation) at the expense of the others. It is also of note that, although deep conservation and positive selection tag 14.4% of the DNA sequenced, the tagged

regions collectively harbor only 3.9% of the common SNPs. In contrast, balancing selection and population structure demarcate a much smaller portion of the sequence (7.3%) but many more SNPs (19.8%). That is, tests based on diversity-reducing forces of selection identify large regions containing few SNPs, whereas diversity-enhancing forces identify smaller regions with higher SNP density.

Collectively, we tag 5.3 kb (21.8%) of the surveyed noncoding DNAas potentially regulatory, and the identified regions harbor 97 of the 408 common biallelic SNPs discovered. Assuming that the subset of SNPs we identify includes the majority of fSNPs, if we adopted an association study approach genotyping only these 97 sites, our genotyping effort would be reduced 4-fold over genotyping all common noncoding SNPs. This value is not inconsistent with the reduction in genotyping effort for the HapMap proposal implied by some studies (Goldstein *et al.*, 2003; Patil *et al.*, 2001), although the actual level of reduction possible under the HapMap plan remains unclear. We note that the reduction in genotyping effort we propose assumes coding variants contribute little to phenotypic variation.

The major remaining question is how often each of the annotation methods identifies functional regions that influence complex phenotypes. An obvious reverse approach to answering this question is to assess the ability of the sequence-based methods we propose to identify known regulatory elements.Wehave previously examined this for the *Enhancer of split* locus in *Drosophila* (Macdonald and Long, 2005), and as more regulatory elements are identified by using molecular and developmental techniques, the ability of sequence-based methods alone to detect them can be assessed. Several forward approaches are also possible. For highly conserved regions, tests of function have taken two forms, the ability to drive gene expression in promoter–reporter constructs (Hong *et al.*, 2003; Johnson *et al.*, 2004) and the ability to bind transcription factors (Boffelli *et al.*, 2003). The population genetic approaches we present likely identify more quickly evolving regions, which may harbor regulatory elements that influence only a subset of tissues and/or developmental times. Such elements may make important contributions to complex traits, but their functional role may be difficult to confirm with promoter-reporter assays. An alternative approach may be to identify a set of putative regulatory regions using an array of methods, and exhaustively test all polymorphisms in these regions for an association with phenotype. The degree of association at the sites could then be used to assess the rate at which each annotation method falsely classifies a DNA region as harboring an fSNP. Clearly, this experiment lacks finesse, but it has the advantage of directly providing an estimate of the phenotypic effect associated with each SNP identified on the basis of primary sequence data.

A model system that is probably most amenable to this test is the

classic *D. melanogaster* bristle number quantitative trait, shown to be under stabilizing selection (García-Dorado and González, 1996), and its associated set of candidate genes (Mackay, 1995; Norga *et al.*, 2003). Many of the proteins encoded by these genes are members of the *Notch* signaling pathway, regulate members of this pathway, or are involved in the development of the peripheral nervous system in *Drosophila*. For our sequencing survey, 23 of 26 amplicons were developed in or near genes involved in these processes (Table 15.1). Thus, we have a strong *a priori* prediction that fSNPs in regions visible to selection at these candidate loci are likely to contribute to natural variation for bristle number. Clearly, SNPs in nonneutrally evolving regions around these genes do not necessarily have to affect bristle number, but mutant alleles associated with these genes regularly have pleiotropic effects on bristle number and patterning (www.flybase.org; Norga *et al.*, 2003). So, although regions of these loci experiencing recent selection are not expected to directly map to those fSNPs contributing to bristle number variation, we do expect the two sets of regions to overlap to some extent.

It is important to understand the ability of different methods of genome annotation to uncover functional regulatory variation to direct future genome sequencing studies. The current model for genome annotation employs a comparative approach, whereby annotation of a focal genome is aided by sequence comparisons to one or a set of diverged species genomes. However, depending on the performance of other annotation methods, it may be extremely valuable to sequence multiple individuals from a single species in addition to single individuals from multiple species.

ACKNOWLEDGMENT

This work was supported by National Institutes of Health Grant GM58564 (to A.D.L.).

REFERENCES

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287,** 2185–2195.

Andolfatto, P. & Przeworski, M. (2001) Regions of lower crossing over harbor more rare variants in African populations of *Drosophils melanogaster*. *Genetics* **158,** 657–665.

Berman, B. P., Pfeiffer, B. D., Laverty, T. R., Salzberg, S. L., Rubin, G. M., Eisen, M. B. & Celniker, S. E. (2004) Computational identification of developmental enhancers: Conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Genome Biol.* **5,** R61.

Berry, A. & Kreitman, M. (1993) Molecular analysis of an allozyme cline: Alcohol dehydrogenase in *Drosophila melanogaster* on the East Coast of North America. *Genetics* **134,** 869–893.

Bier, E., Vaessin, H., Younger-Shepherd, S., Jan, L. Y. & Jan, Y. N. (1992) *deadpan*, an essential pan-neural gene in *Drosophila*, encodes a helix-loop-helix protein similar to the *hairy* gene product. *Genes Dev.* **6,** 2137–2151.

Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K. D., Ovcharenko, I., Pachter, L. & Rubin, E. M. (2003) Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* **299,** 1391–1394.

Botstein, D. & Risch, N. (2003) Discovering genotypes underlying human phenotypes: Past successes for Mendelian disease, future approaches for complex disease. *Nat. Genet.* **33,** Suppl., 228–237.

Carlson, C. S., Eberle, M. A., Rieder, M. J., Smith, J. D., Kruglyak, D. & Nickerson, D. A. (2003) Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nat. Genet.* **33,** 518–521.

Carlson, C. S., Eberle, M. A., Kruglyak, L. & Nickerson, D. A. (2004) Mapping complex disease loci in whole-genome association studies. *Nature* **429,** 446–452.

Daborn, P. J., Yen, J. L., Bogwitz, M. R., Le Goff, G., Feil, E., Jeffers, S., Tijet, N., Perry, T., Heckel, D., Batterham, P., *et al.* (2002) A single P450 allele associated with insecticide resistance in *Drosophila*. *Science* **297,** 2253–2256.

Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J. & Lander, E. S. (2001) High-resolution haplotype structure in the human genome. *Nat. Genet.* **29,** 229–232.

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., *et al.* (2002) The structure of haplotype blocks in the human genome. *Science* **296,** 2225–2229.

García-Dorado, A. & González, J. A. (1996) Stabilizing selection detected for bristle number in *Drosophila melanogaster*. *Evolution (Lawrence, Kans.)* **50,** 1573–1578.

Goldstein, D. B., Ahmadi, K. R., Weale, M. E. & Wood, N. W. (2003) Genome scans and candidate gene approaches in the study of common diseases and variable drug responses. *Trends Genet.* **19,** 615–622.

Hong, R. L., Hamaguchi, L., Busch, M. A. & Weigel, D. (2003) Regulatory elements of the floral homeotic gene AGAMOUS identified by phylogenetic footprinting and shadowing. *Plant Cell* **15,** 1296–1309.

Hudson, R. R., Kreitman, M. & Aguadé, M. (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* **116,** 153–159.

International HapMap Consortium (2003) The international HapMap project. *Nature* **426,** 789–796.

International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–945.

Johnson, D. S., Davidson, B., Brown, C. D., Smith, W. C. & Sidow, A. (2004) Noncoding regulatory sequences of *Ciona* exhibit strong correspondence between evolutionary constraint and functional importance. *Genome Res.* **14,** 2448–2456.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423,** 241–254.

Kim, Y. & Stephan, W. (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160,** 765–777.

Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, U.K.).

Kreitman, M. & 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,** 565–582.

Kruglyak, L. (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* **22,** 139–144.

Kruglyak, L. & Nickerson, D. A. (2001) Variation is the spice of life. *Nat. Genet.* **27,** 234–236.

Kuang, B., Wu, S. C., Shin, Y., Luo, L. & Kolodziej, P. (2000) *split ends* encodes large nuclear proteins that regulate neuronal cell fate and axon extension in the *Drosophila* embryo. *Development (Cambridge, U.K.)* **127,** 1517–1529.

Lage, P. Z., Shrimpton, A. D., Flavell, A. J., Mackay, T. F. C. & Brown, A. J. L. (1997) Genetic and molecular analysis of *smooth*, a quantitative trait locus affecting bristle number in *Drosophila melanogaster*. *Genetics* **146,** 607–618.

Lai, C., McMahon, R., Young, C., Mackay, T. F. C. & Langley, C. H. (1998) *quemao*, a Drosophila bristle locus, encodes geranylgeranyl pyrophosphate synthase. *Genetics* **149,** 1051–1061.

Long, A. D. & Langley, C. H. (1999) The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome. Res.* **9,** 720–731.

Ludwig, M. Z., Bergman, C., Patel, N. H. & Kreitman, M. (2000) Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403,** 564–567.

Macdonald, S. J. & Long, A. D. (2005) Identifying signatures of selection at the *Enhancer of split* neurogenic gene complex in *Drosophila*. *Mol. Biol. Evol.* **22,** 1–13.

Mackay, T. F. (1995) The genetic basis of quantative variation: Numbers of sensory bristles of *Drosophila melanogaster* as a model system. *Trends Genet.* **11,** 464–470.

McDonald, J. H. & Kreitman, M. (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351,** 652–654.

Nielsen, R. & Yang, Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148,** 929–936.

Norga, K. K., Gurganus, M. C., Dilda, C. L., Yamamoto, A., Lyman, R. F., Patel, P. H., Rubin, G. M., Hoskins, R. A., Mackay, T. F. & Bellen, H. J. (2003) Quantitative analysis of bristle number in *Drosophila* mutants identifies genes involved in neural development. *Curr. Biol.* **13,** 1388–1397.

Nurminsky, D. I., Nurminskaya, M. V., De Aguiar, D. & Hartl, D. L. (1998) Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature* **396,** 572–575.

Patil, N., Berno, A. J., Hinds, D. A., Barrett, W. A., Doshi, J. M., Hacker, C. R., Kautzer, C. R., Lee, D. H., Marjoribanks, C., McDonough, D. P., *et al.* (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294,** 1719–1723.

Ramain, P., Khechumian, K., Seugnet, L., Arbogast, N., Ackermann, C. & Heitzler, P. (2001) Novel Notch alleles reveal a Deltex-dependent pathway repressing neural cell fate. *Curr. Biol.* **11,** 1729–1738.

Reich, D. E., Gabriel, S. B. & Altshuler, D. (2003) Quality and completeness of SNP databases. *Nat. Genet.* **33,** 457–458.

Richards, S., Liu, Y., Bettencourt, B. R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M. J., Chen, R., Meisel, R. P., *et al.* (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene, and *cis*-element evolution. *Genome Res.* **15,** 1–18.

Risch, N. & Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science* **273,** 1516–1517.

Robin, C., Lyman, R. F., Long, A. D., Langley, C. H. & Mackay, T. F. (2002) *hairy*: A quantitative trait locus for Drosophila sensory bristle number. *Genetics* **162,** 155–164.

Russo, C. A. M., Takezaki, N. & Nei, M. (1995) Molecular phylogeny and divergence times of Drosophilid species. *Mol. Biol. Evol.* **12,** 391–404.

Schlenke, T. A. & Begun, D. J. (2004) Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. *Proc. Natl. Acad. Sci. USA* **101,** 1626–1631.

Shapiro, M. D., Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Jónsson, B., Schluter, D. & Kingsley, D. M. (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428,** 717–723.

Stam, L. F. & Laurie, C. C. (1996) Molecular dissection of a major gene effect on a quantitative trait: The level of Alcohol dehydrogenase expression on *Drosophila melanogaster*. *Genetics* **144,** 1559–1564.

Stephens, J. C., Schneider, J. A., Tanguay, D. A., Choi, J., Acharya, T., Stanley, S. E., Jiang, R., Messer, C. J., Chew, A., Han, J.-H., *et al.* (2001) Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293,** 489–493.

Syvänen, A.-C. (2001) Accessing genetic variation: Genotyping single nucleotide polymorphisms. *Nat. Rev. Genet.* **2,** 930–942.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123,** 585–595.

Ueda, H., Howson, J. M. M., Esposito, L., Heward, J., Snook, H., Chamberlain, G., Rainbow, D. B., Hunter, K. M. D., Smith, A. N., Di Genova, G., *et al.* (2003) *Nature* **423,** 506–511.

Wright, S. (1951) Association of the T-cell regulatory gene *CTLA4* with susceptibility to autoimmune disease. *Ann. Eugen.* **15,** 323–354.

16

Genetics and Genomics of *Drosophila* Mating Behavior

TRUDY F. C. MACKAY,* STEFANIE L. HEINSOHN,*
RICHARD F. LYMAN,* AMANDA J. MOEHRING,*†
THEODORE J. MORGAN,* AND
STEPHANIE M. ROLLMANN*

The first steps of animal speciation are thought to be the development of sexual isolating mechanisms. In contrast to recent progress in understanding the genetic basis of postzygotic isolating mechanisms, little is known about the genetic architecture of sexual isolation. Here, we have subjected *Drosophila melanogaster* to 29 generations of replicated divergent artificial selection for mating speed. The phenotypic response to selection was highly asymmetrical in the direction of reduced mating speed, with estimates of realized heritability averaging 7%. The selection response was largely attributable to a reduction in female receptivity. We assessed the whole genome transcriptional response to selection for mating speed using Affymetrix GeneChips and a rigorous statistical analysis. Remarkably, >3,700 probe sets (21% of the array elements) exhibited a divergence in message levels between the Fast and Slow replicate lines. Genes with altered transcriptional abundance in response to selection fell into many different biological process and molecular function Gene Ontology categories, indicating substantial pleiotropy for this complex behavior. Future functional studies are necessary to test the extent to which transcript profiling of divergent selection lines accurately predicts genes that directly affect the selected trait.

*Department of Genetics, North Carolina State University, Box 7614, Raleigh, NC 27695.
†Present address: Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803.

Abbreviations: QTLs, quantitative trait loci; Z, Zimbabwe; GO, Gene Ontology.

> Species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups.
>
> *Recent studies by the students of animal behavior, as well as the revised interpretation of many earlier observations, indicate that behavior differences are among animals the most important factor in restricting random mating between closely related forms.*
>
> E. Mayr (1942)

One of the major challenges facing modern biology is to understand the genetic mechanisms causing speciation. Because sexual isolating mechanisms that act before fertilization [''ethological'' isolating mechanisms (Mayr, 1942)] are thought to precede the evolution of postzygotic isolating mechanisms (inviability and sterility), we need to understand the genetic basis of sexual isolation if we are to gain insight about the early stages of species formation. However, mating behaviors are complex traits, with variation attributable to multiple interacting loci with individually small effects, whose expression depends on the environment. Thus, understanding the genetic architecture of sexual isolation requires that we overcome the twin obstacles of mapping genes causing differences between organisms that, by definition, do not interbreed [Orr, "The Genetic Basis of Reproductive Isolation: Insights from *Drosophila*" (Chapter 2, this volume)] and solving the problem of genetically dissecting complex behavioral traits (Anholt and Mackay, 2004).

DROSOPHILA MATING BEHAVIOR

Drosophila species present an ideal model system in which to investigate the genetic basis of sexual isolation. Several species pairs are only partially reproductively isolated, producing fertile hybrids that can be backcrossed to one of the parental species to generate segregating backcross mapping populations. Furthermore, *Drosophila melanogaster* is a model organism with excellent genetic and genomic resources that are ideal for genetically dissecting complex traits, including the ability to clone chromosomes, replicate genotypes, and rear large numbers of individuals under uniform environmental conditions; publicly available mutations and deficiency stocks useful for mapping; abundant segregating variation in natural populations that can readily be selected in the laboratory to produce divergent phenotypes a complete well annotated genome sequence; and several platforms for whole-genome transcriptional profiling. Courtship behavior of *Drosophila* is composed of sequential actions that exchange auditory, visual, and chemosensory signals between males and females, allowing for individual components of the behavior to be

quantified and separated (Greenspan, 1995; Hall, 1994). Courtship is initiated when the male aligns himself with the female, using visual and olfactory signals for orientation. He then taps the female's abdomen with his foreleg, using pheromonal cues for gender and species recognition, followed by wing vibration to produce a species-specific courtship song. After courtship initiation, the male again uses pheromonal cues by licking the female's genitalia, after which he will attempt to copulate. The female can accept the male or reject him by moving away. Successful copulation is accompanied by the transfer of sperm and seminal fluids that stimulate the release of oocytes by the ovary (Heifetz *et al.*, 2000) and reduce female receptivity to other males (Clark *et al.*, 1995; Wolfner *et al.*, 1997). Components of the seminal fluids are associated with the reduced lifespan of mated females (Chapman *et al.*, 1995), setting up an intersexual conflict [Rice *et al.*, "Inter-Locus Antagonistic Coevolution as an Engine of Speciation: Assessment with Hemiclonal Analysis" (Chapter 3, this volume)].

Given the complexity of *Drosophila* courtship behavior, it is not surprising that mutations in genes affecting multiple biological processes affect mating behavior (Hall *et al.*, 1980; Yamamoto and Nakano, 1998). These include mutations in genes required for normal morphology [*white* (Sturtevant, 1915; Zhang and Odenwald, 1995), *yellow* (Zhang and Odenwald, 1995), and *curved* (Peixoto and Hall, 1998)], as well as genes involved in learning and memory [*Calcium calmodulin kinase II* (Joiner and Griffith, 1997), *dunce* (Kyriacou, 1990; Kyriacou and Hall, 1985), *rutabaga* (Gailey *et al.*, 1984; Kyriacou and Hall, 1984), *turnip* (Gailey *et al.*, 1982, 1984), and *amnesiac* (Ackerman and Seigal, 1986; Kyriacou and Hall, 1984; Siegel and Hall, 1979)], circadian rhythm [*period* (Ewing, 1988; Jackson *et al.*, 1983; Kyriacou, 1990; Kyriacou and Hall, 1994)] and dopamine and serotonin synthesis [*Dopa decarboxylase* (Tempel *et al.*, 1984), *pale* (Buchner, 1991; Neckameyer, 1998), *tan* (Cook, 1980; Tompkins *et al.*, 1982), and *ebony* (Crossley and Zuill, 1970; Kyriacou *et al.*, 1978; Rendel, 1951)], sex determination [*doublesex* (Arthur *et al.*, 1998; Jallon, 1984; Villella and Hall, 1996), *transformer* (Bernstein *et al.*, 1992; Ferveur *et al.*, 1997; Finley *et al.*, 1998; O'Dell and Kaiser, 1995; Taylor *et al.*, 1994; Waterbury *et al.*, 1999), *fruitless* (Baker *et al.*, 2001; Gailey *et al.*, 1991; Ryner *et al.*, 1996; Wheeler *et al.*, 1989), and *sex lethal* (Tompkins and McRobert, 1995)], pheromone production [*desaturase 2* (Takahashi *et al.*, 2001)], and accessory gland-specific peptides (Clark *et al.*, 1995; Fleishmann *et al.*, 1995; Heifetz *et al.*, 2000; Lung *et al.*, 2001; Nakayama *et al.*, 1997; Wolfner *et al.*, 1997).

SEXUAL ISOLATION AMONG SPECIES

Despite the wealth of knowledge regarding genetic mechanisms that affect *Drosophila* courtship behavior, we know virtually nothing of the

genes that cause naturally occurring variation in mating behavior within and among species, their allelic effects, and their interactions. Are the loci that harbor naturally occurring variation a subset of loci identified by mutational analysis, or will the analysis of natural variants reveal novel loci? Is natural variation in mating behavior attributable to a few genes with large effects or many genes with small effects? Do the alleles at different loci interact additively or exhibit epistasis? Do the same genes that affect variation in courtship behavior within species account for sexual isolation between species? Answers to these questions require that we identify the quantitative trait loci (QTLs) affecting sexual isolation between species and variation in mating behavior within species.

Because QTLs often have small effects that are contingent on the environment, they can be mapped only by linkage to markers whose genotype can be scored unambiguously (Mackay, 2001). Before the recent discovery of abundant polymorphic molecular markers, mapping the QTLs affecting sexual isolation between *Drosophila* species was confined to estimates of the effects of each chromosome arm (Coyne, 1989, 1993, 1996a,b; Coyne *et al.*, 2002; Noor, 1997; Zouros, 1981).

Two recent studies addressed the genetic basis of variation in sexual isolation between *Drosophila pseudoobscura* and *Drosophila persimilis* (Noor *et al.*, 2001) and between *Drosophila simulans* and *Drosophila mauritiana* (Moehring *et al.*, 2004) by linkage to molecular markers in large backcross populations. In the first species pair, sexual isolation is attributable to female discrimination against males of the sibling species; males readily court females of either species. QTLs affecting male traits against which *D. pseudoobscura* discriminate are located primarily on the left arm of the X chromosome, with minor contributions from the right arm of the X and second chromosomes. QTLs affecting male traits against which *D. persimilis* discriminate are located on the second chromosome (Noor *et al.*, 2001).

D. mauritiana females rarely mate with *D. simulans* males. At least seven QTLs, mapping to all three chromosomes, affect the discrimination of *D. mauritiana* females against *D. simulans* males; and three QTLs, all on the third chromosome, affect the *D. simulans* male traits against which *D. mauritiana* females discriminate. QTLs for female choice are different from those for the male traits they are choosing against. Although *D. simulans* females mate with *D. mauritiana* males, copulations are abnormally short and often do not result in adequate sperm transfer (Coyne, 1993). At least six autosomal QTLs affect the *D. mauritiana* male traits against which *D. simulans* females discriminate. No epistatic interactions were observed between QTLs affecting prezygotic isolation, in contrast to the genetic architecture of postzygotic isolation [Orr, "The Genetic Basis of Reproductive Isolation: Insights from *Drosophila*" (Chapter 2, this volume)]. Al-

though a few QTLs with moderate effects affect prezygotic reproductive isolation in both of these species pairs, high-resolution recombination mapping will be necessary to identify individual genes.

VARIATION IN MATING BEHAVIOR WITHIN *D. MELANOGASTER*

Genetic variation for incipient sexual isolation has been implicated within populations of *D. melanogaster* by repeated observations that positive assortative mating can evolve as a correlated response to divergent artificial selection for sensory bristle numbers, geotaxis, phototaxis, and locomotor activity (Speith and Ringo, 1983). Presumably, assortative mating evolves because genes affecting the selected traits are closely linked to genes affecting mating behavior or have pleiotropic effects on mating behavior. There is naturally occurring polymorphism for incipient sexual isolation within *D. melanogaster*. Females from populations in Zimbabwe (Z) exhibit strong preference for Z males when given a choice between Z and Cosmopolitan (C) males, but the reciprocal crosses exhibit weaker or no sexual isolation (Wu *et al.*, 1995). Chromosome substitution analyses revealed that QTLs affecting the discrimination of Z females against C males, as well as QTLs affecting the attractiveness of Z males to Z females, reside on all major chromosomes, with the third chromosome having the greatest and the X chromosome the least effect (Hollocher *et al.*, 1997). Recombination mapping of third-chromosome QTLs using visible morphological markers revealed at least four epistatic QTLs affecting Z male mating success and at least two QTLs affecting Z female mating preference (Ting *et al.*, 2001).

Recently, QTLs affecting variation in male mating behavior between Oregon (Ore), a standard wild-type strain, and 2b, a strain selected for reduced male courtship and copulation latency, have been mapped with high resolution by linkage to molecular markers in a panel of 98 recombinant inbred lines derived from these strains (Moehring and Mackay, 2004). The initial genome scan revealed a minimum of one X chromosome and three autosomal QTLs affecting variation in male mating behavior between Ore and 2b. These QTLs mapped to relatively large genomic regions containing on average >600 genes. However, in *D. melanogaster*, one can readily map QTLs to subcM regions using deficiency complementation mapping (Pasyukova *et al.*, 2000) and identify candidate genes corresponding to the QTLs using quantitative complementation tests to mutations at the positional candidate genes (Long *et al.*, 1996; Mackay and Fry, 1996). The three autosomal QTLs fractionated into five QTLs containing 58 genes on average. Complementation tests to all 45 available mutations at the positional candidate genes delimited by deficiency mapping re-

vealed seven novel candidate genes affecting male mating behavior: *eagle*, *18 wheeler*, *Enhancer-of-split*, *Polycomb*, *spermatocyte arrest*, *l (2)05510*, and *l (2)k02006*. These genes are involved in spermatogenesis, chromatin and gene silencing, serotonin neuron fate determination, and nervous system development. None of these genes has been previously implicated in mating behavior, demonstrating that quantitative analysis of subtle variants can reveal novel pleiotropic effects of key developmental genes on behavior (Moehring and Mackay, 2004).

Our ability to map the genes affecting naturally occurring variation in mating behavior within *D. melanogaster* is compromised by two factors. First, the size of the mapping populations determines the minimum QTL effect that can be detected. Increasing the sample size will increase the numbers of mapped QTLs, because linked QTLs can be separated by recombination, and the minimum detectable effect decreases as the sample size increases. Second, any two strains used to map QTLs are limited samples of the existing variation (Mackay, 2001). Recently, there has been great excitement about the utility of whole genome transcriptional profiling to identify candidate genes regulating complex traits by assessing changes in gene expression between lines selected for different phenotypic values of the trait (Toma *et al.*, 2002). Here, we describe the results of 29 generations of replicated selection for increased and decreased mating speed from a large heterogeneous base population and the analysis of the whole genome transcriptional response to artificial selection.

MATERIALS AND METHODS

Drosophila Selection Lines

The base population consisted of 60 isofemale lines collected in Raleigh, NC, in 2002 using fruit baits. The 60 lines were crossed in a round-robin design (♀1×♂2, ♀2×♂3, . . . , ♀60×♂1) in separate culture vials,with three females and three males per vial. After 2 days, one inseminated female from each cross was placed in each of two culture bottles to initiate replicate selection lines. The progeny from each replicate bottle were scored for copulation latency to initiate Generation 1 of selection. A total of 50 pairs of 4- to 7-day-old virgin males and females from each replicate bottle were placed in culture vials, and the time to copulation was scored for each pair, for a total of 3 h. The 20 fastest pairs from each replicate were placed in culture bottles to initiate the two Fast selection lines, and the 20 slowest pairs were placed in culture bottles to initiate the two Slow lines. Control lines were started from the 10 middle-scoring pairs from each line, plus 10 pairs of virgin males and females that were not scored. In the second and subsequent generations, 50 males and females from the

six replicate lines were scored for copulation latency. The Fast lines were maintained by selecting the 20 fastest pairs each generation, and the Slow lines were maintained by selecting the 20 slowest pairs. The control lines were maintained by 20 pairs that were chosen at random with respect to copulation latency. Pairs that did not mate in the 3-h observation period were given a score of 180 min.

Flies were reared on standard cornmeal–molasses–agar medium and maintained in an incubator at 25°C and a 12:12 h light/dark cycle. Mating behavior was assessed for 3 h in the morning, 2 h after lights on.

Quantitative Genetic Analysis of Selection Response

Realized heritability of copulation latency was computed for each replicate from the regression of cumulated response (as a deviation from the control) on cumulated selection differential (Falconer and Mackay, 1996).

Male Mating Behavior

We assessed correlated responses in male mating behavior in response to selection for copulation latency from generations 21–23. Male mating behavior was assessed for 1 h, immediately after the flies were paired. Otherwise, the conditions were identical to those under which copulation latency was scored. Courtship latency is the time to initiate courtship behavior. We scored courtship intensity by observing individual males every minute after initiation of courtship until copulation occurred and recording whether they were engaged in courtship behavior. The measure of courtship intensity was the number of times they were observed courting divided by the total number of observations.

Transcriptional Profiling

At Generation 23, three replicate groups of 50 4- to 7-day-old virgin males and females were collected from the two Fast and two Slow replicate lines (i.e., the same age and mating status as the flies before selection). Total RNA was extracted independently for each of the 24 samples (four lines × two sexes × three replicates) by using the TRIzol reagent (GIBCO/BRL). The samples were treated with DNase and purified on Qiagen (Chatsworth, CA) RNeasy columns. Biotinylated cRNA probes were hybridized to high-density oligonucleotide Affymetrix *Drosophila* GeneChip 2.0 microarrays and visualized with a streptavidin–phycoerythrin conjugate, as described in the Affymetrix GeneChip Expression Analysis Technical Manual (2000), using internal references for quantification.

Micorarray Data Analysis

We normalized the expression data by scaling overall probe set intensity to 300 on each microarray using standard reference probe sets on each GeneChip for the normalization procedure. Every gene on the Affymetrix *Drosophila* GeneChip 2.0 is represented by a probe set consisting of 14 perfect match (PM) and 14 mismatch (MM) probe pairs. The quantitative estimate of expression of each probe set is the Signal (*Sig*) metric. *Sig* is computed by using the one-step Tukey's biweight estimate, which gives the weighted mean of the log(PM – MM) intensities for each probe set (AffymetrixMicroarray Suite, Ver. 5.0). A detection call (present, marginal, or absent) is also given for each probe set. We eliminated probe sets from consideration if over one-half were called absent. In practice, this retained probe sets with sex-specific expression and removed those with low and variable *Sig* values.

We performed two-way fixed-effect ANOVAs of the expression values for all remaining probe sets, according to the model $Y = \mu + S + L + S \times L + E$, where S and L are the crossclassified effects of sex and selection line (Fast replicate 1, Fast replicate 2, Slow replicate 1, and Slow replicate 2), respectively, and E is the variance between replicate arrays. P values were computed from F ratio tests of significance for each of the terms in the ANOVA. Because there are >18,000 probe sets on the array, this poses a huge multiple testing problem for determining the significance threshold using P values. Bonferroni corrections for multiple tests are too conservative, and a conventional 5% significance threshold will yield too many false positives. We used a $Q = 0.001$ false-discovery rate criterion (Storey and Tibshirani, 2003) for the significance of any of the terms in theANOVA model. Unlike the P value, which is the number of false positives expected when truly nothing is significant, the false discovery rate Q value controls the proportion of false positives among all terms declared significant (Storey and Tibshirani, 2003).

Variation in transcript abundance between lines could be attributable to changes in gene frequency due to random drift or to changes in frequency of genes under selection. In the latter case, one would expect common alleles affecting variation in transcript abundance to have the same effect in both selection lines. Therefore, contrast statements were used to assess whether transcript abundance for probe sets with L and/or $S \times L$ terms at or below the $Q = 0.001$ threshold was significantly different between the two Fast lines and the two Slow lines, both pooled over sexes, and for each sex separately.

Statistical analyses were conducted by using SAS software (SAS Institute, Cary, NC). Cytological locations and biological process and

molecular function gene ontologies were given by the NetAffyx (www.affymetrix.com/analysis/index.affx) database, supplemented by information from the FlyBase Consortium (2003), current as of December 31, 2004.

RESULTS

Phenotypic Response to Selection for Copulation Latency

The result of 29 generations of replicated selection for increased and decreased copulation latency is depicted in Fig. 16.1*A*. The selection response is highly asymmetrical in the direction of increased copulation latency. The Fast and Slow replicate lines were significantly diverged from Generation 25. We analyzed the mating speed data from generations 25–29 according to the mixed model ANOVA $Y = \mu + S + G + G \times S + R(S) + G \times R(S) + E$, where μ is the overall mean; S and G are the crossclassified fixed effects of direction of selection (Fast vs. Slow) and generation, respectively; R is the random effect of replicate line; and E is the variance within lines. The effect of direction of selection was highly significant ($F_{1, 2} = 617.71$, $P = 0.0016$).

We computed realized heritabilities (h^2) of mating speed from the regressions of cumulated response on cumulated selection differentials [Falconer and Mackay (1996) and Fig. 16.1*B* and *C*). Estimates of h^2 (±SE of the regression coefficient) were $h^2 = 0.047$ (0.025) and $h^2 = 0.011$ (0.020) for Replicate 1 and 2 Fast lines, respectively; neither estimate is significantly different from zero. Estimates of h^2 for the Replicate 1 and 2 Slow lines, respectively, were $h^2 = 0.059$ (0.015, $P = 0.0006$) and $h^2 = 0.099$ (0.016, $P < 0.0001$). Heritabilities estimated from the divergence were $h^2 = 0.056$ (0.011, $P < 0.0001$) and $h^2 = 0.078$ (0.012, $P < 0.0001$) for Replicates 1 and 2, respectively.

Reduced mating speed could be attributable to reduced male copulation latency, reduced female receptivity, or both. At generations 18, 20, and 21, we assessed copulation latency when Fast females of each replicate were paired with Slow males and when Slow females of each replicate were paired with Fast males. The results of these tests, as well as the responses of the selection lines in these generations, are shown in Fig. 16.1*D*. We analyzed the copulation latency data by the fixed-effects ANOVA model $Y = \mu + C + G + C \times G + E$, where C is cross, G is generation, and E is the variation within each cross and generation. The effect of cross was highly significant ($F_{7, 1176} = 221.95$, $P < 0.0001$). Post hoc Tukey tests revealed there was no significant difference in mating speed between Fast females of either replicate when paired with Fast or Slow

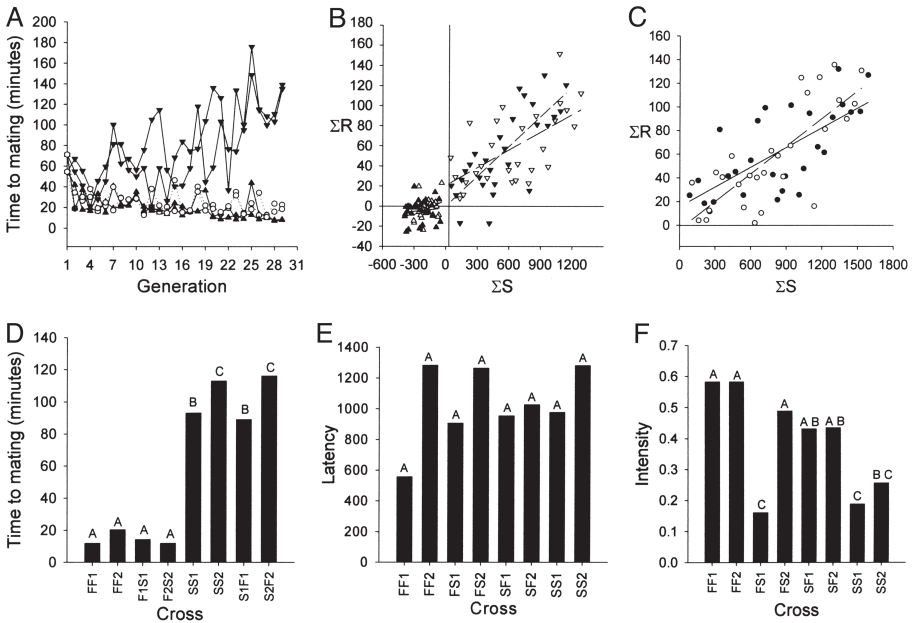


FIGURE 16.1 Phenotypic response to selection for copulation latency. (*A*) Mean mating speed of selection lines. ▲, Fast lines; ▼, Slow lines; o, Control lines. (*B*) Regressions of cumulated response on cumulated selection differential for Fast and Slow selection lines. ▲, Replicate 1, Fast; ∇, Replicate 2, Fast; ▼, Replicate 1, Slow; ∇, Replicate 2, Slow. (*C*) Regressions of cumulated response on cumulated selection differential for divergence between Fast and Slow selection lines. ●, Replicate 1; o, Replicate 2. (*D*) Mating speeds averaged over generations 18, 20, and 21 for Fast females paired with Fast males (FF), Fast females paired with Slow males (FS), Slow females paired with Slow males (SS), and Slow females paired with Fast males (SF). The subscripts denote Replicates 1 and 2, respectively. *A*, *B*, and *C* indicate the results of Tukey tests. Groups with the same letter are not significantly different. (*E*) Male courtship latency. Groups are the same as in *D*. (*F*) Male courtship intensity. Groups are the same as in *D*.

males. However, Slow females were equally slow when paired with Slow or Fast males. Clearly, the rapid evolution of reduced copulation latency is attributable to reduced female receptivity: slow females are picky.

We assessed correlated responses in male behavior by measuring courtship latency and courtship intensity for each of the reciprocal pairs of selection lines (Fast females and Fast males, Fast females and Slow males, Slow females and Slow males, and Slow females and Fast males) for each replicate. The data were analyzed by ANOVA, as described above for copulation latency. There was no detectable difference in courtship latency of males in any of the crosses ($F_{7, 143} = 1.54$, $P = 0.158$; Fig. 1*E*). There were, however, highly significant differences in courtship intensity between the crosses ($F_{7, 142} = 5.92$, $P < 0.0001$; Fig. 1*F*). The courtship intensity of Fast males with Fast females was much greater than that with Slow males and Slow females. The courtship intensity of both replicates of Fast males with Slow females was not significantly different from that of these males with Fast females. However, the courtship intensity of Slow males from Replicate 1 with Fast females was as low as with Slow females, but the courtship intensity of Slow males from Replicate 2 with Fast females was as fast as the Fast males (Fig. 16.1*F*), indicating some divergence between the replicates in correlated male behaviors.

Transcriptional Response to Selection for Copulation Latency

We assessed transcript abundance at the time of selection for the Fast and Slow selection lines, using Affymetrix high-density oligonucleotide whole genome microarrays. Raw expression data are given in Table 4, which is published as supporting information on the PNAS web site. Statistically significant differences in transcript abundance were evaluated by factorial ANOVA (with line and sex the two crossclassified main effects) for each probe set. Using a false discovery rate of $Q = 0.001$ (i.e., one false positive in 1,000 among probe sets declared significant), 10,336 probe sets were significant for the main effect of sex, 4,420 were significant for the main effect of line, and 1,107 were significant for the line $\times$ sex interaction.

We used ANOVA contrast statements to detect probe sets that were up- or down-regulated in both Fast and Slow selection replicates, as would be expected if gene frequencies of the same common alleles changed in both selection lines. Remarkably, a total of 3,727 probe sets met this criterion (Table 5, which is published as supporting information on the PNAS web site). Of these, 836 were male-specific (505 of these probe sets were up-regulated in Fast males, and 331 were up-regulated in Slow males), 1,336 were female-specific (912 were up-regulated in Fast females, and 424 were up-regulated in Slow females), and 1,490 affected both sexes

(575 were up-regulated in Fast lines, and 915 were up-regulated in Slow lines). In addition, transcript abundance for 65 probe sets had sexually antagonistic effects. Of these, 23 were up-regulated in Fast females and down-regulated in Fast males, and 42 were upregulated in Fast males and down-regulated in Fast females. Clearly, there has been a widespread transcriptional response to selection for mating speed. However, the magnitude of the changes of transcript abundance is not great,with the vast majority much less than 2-fold (Fig. 16.2).

We assessed whether probe sets with significantly altered transcript abundance were randomly distributed among the five major chromosome arms. We counted the number of probe sets on each chromosome arm and used a χ^2 goodness-of-fit test to check for departure from the expected number, computed based on the total fraction of the genome on each chromosome arm. We observed a nonrandom distribution of probe sets that were up-regulated in Fast relative to Slow males ($\chi_4^2 = 20.19$; $P = 0.0005$) and for probe sets that were up-regulated in Slow relative to Fast males ($\chi_4^2 = 19.56$; $P = 0.0006$) (Fig. 16.3). In both cases, a deficiency of up-regulated transcripts on the X chromosome contributed to the significant χ^2 statistic. In addition, there was an excess of transcripts up-regulated in Slow relative to Fast males on chromosome 2L. We also assessed whether probe sets were nonrandomly distributed along each chromosome arm, as might be expected if selection caused linkage disequilibrium between selected loci and closely linked genes. We counted the number of probe sets in each major cytological division and used a χ^2 goodness-of-fit test to check for departure from the expected number, based on the total fraction of genes on each chromosome arm per cytological division. Only 5 of the 30 χ^2 statistics were significant at $P < 0.05$ and, of these, only one test statistic was significant based on a Bonferroni correction for multiple tests. This was for probe sets on chromosome 2L that were up-regulated in Fast relative to Slow females ($\chi^2_{19} = 50.638$; $P = 0.0001$), where bands 25, 32, and 35 had fewer up-regulated probe sets than expected, and bands 29 and 31 had more upregulated probe sets than expected. Thus, there was little evidence for nonrandom distribution of probe sets with significantly altered transcript abundance within each chromosome arm.

The probe sets that were up-regulated in each comparison of Fast and Slow selection lines fell into all major biological process and molecular function Gene Ontology (GO) categories. Comparison of the numbers of up-regulated probe sets in each GO category with the number expected based on representation on the microarray revealed that many categories were significantly over- or underrepresented. We hypothesize that GO categories that are overrepresented contain probe sets for which transcript abundance has been altered as a consequence of artificial selection, whereas natural selection opposes artificial selection for probe sets in GO

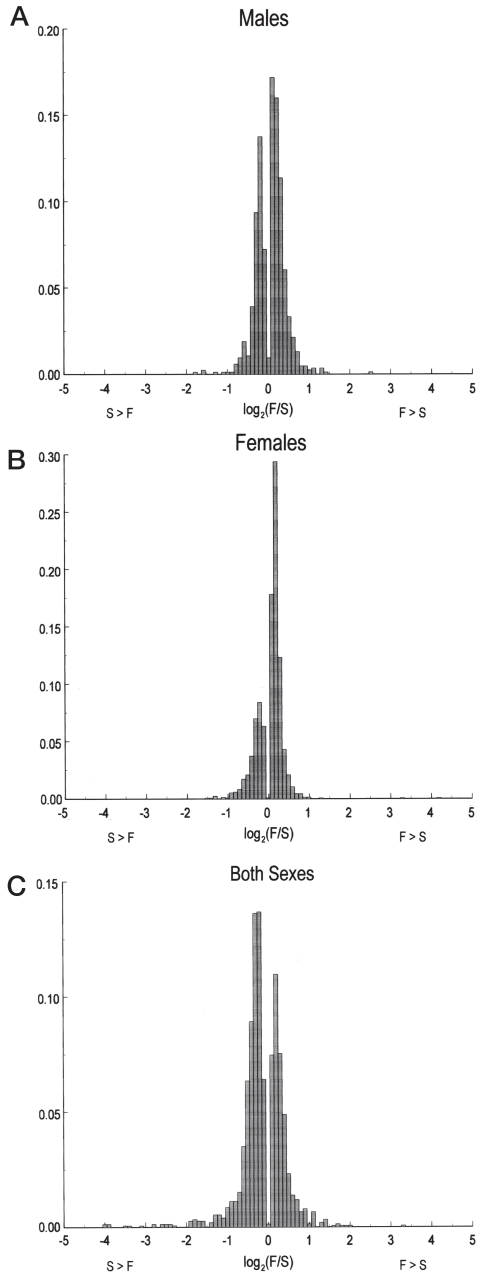


FIGURE 16.2 Relative log2 fold changes in transcript abundance in Fast vs. Slow selection lines. (*A*) Male-specific transcripts. (*B*) Female-specific transcripts. (*C*) Both sexes.

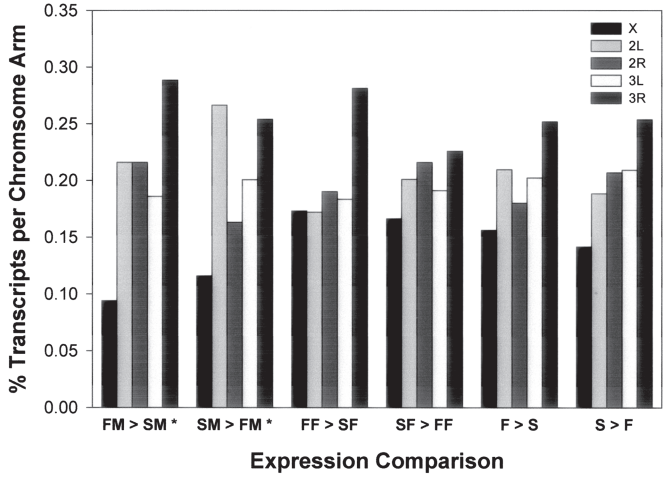


FIGURE 16.3 Chromosomal distribution of transcripts on the major chromosome arms.*, χ_4^2, $P < 0.001$.

categories that are underrepresented. For example, more probe sets than expected that are up-regulated in Fast relative to Slow females fall into the physiological biological process and binding molecular function categories. On the other hand, there are fewer probe sets than expected in the regulation biological process and transcription regulator categories that exhibit significant changes in transcript abundance in multiple comparisons of selection lines (Tables 16.1 and 16.2).

We can begin to build a picture of the transcriptional response to artificial selection by examining GO categories that are overrepresented in the various comparisons of selection lines (Tables 6 and 7, which are published as supporting information on the PNAS web site). Probe sets that are up-regulated in Fast relative to Slow females fall more often than expected in the biological processes categories of cell growth and maintenance ($P = 1.55 \times 10^{-7}$), oocyte maturation ($P = 6.03 \times 10^{-7}$), chromatin silencing ($P = 7.50 \times 10^{-9}$), sexual reproduction ($P = 5.44 \times 10^{-7}$), gene silencing ($P = 2.63 \times 10^{-9}$), RNA metabolism ($P = 2.12 \times 10^{-14}$), DNA metabolism ($P = 1.66 \times 10^{-26}$), and transcription ($P = 1.73 \times 10^{-4}$) and the molecular function categories of histone binding ($P = 4.55 \times 10^{-5}$), DNA replication origin binding ($P = 1.44 \times 10^{-23}$), chromatin binding ($P = 9.45 \times 10^{-14}$), RNA binding ($P = 1.70 \times 10^{-20}$), and helicase activity ($P = 7.91 \times 10^{-8}$). Probe sets involved in neurotransmitter catabolism ($P = 3.53 \times 10^{-13}$) and electron transport ($P = 9.02 \times 10^{-7}$) and that have NADH dehydrogenase activity ($P = 1.87 \times 10^{-7}$) are up-regulated more often than expected in Slow relative to Fast females. Probe sets involved

TABLE 16.1 Biological Process GO Categories

	Male-Specific, *n*		Female-Specific, *n*		Both Sexes, *n*	
GO Category	F > S	S > F	F > S	S > F	F > S	S > F
Behavior	11	6	*6*	11	4	15
	(9.02×10^{-2})	(1.90×10^{-1})	(3.10×10^{-2})	(1.17×10^{-1})	(1.36×10^{-1})	(4.93×10^{-1})
Cellular	*105*	60	293	139	*114*	233
	(2.04×10^{-4})	(8.36×10^{-2})	(2.84×10^{-1})	(9.49×10^{-1})	(2.46×10^{-8})	(1.39×10^{-1})
Development	*42*	22	134	64	*39*	121
	(2.75×10^{-3})	(3.34×10^{-2})	(7.98×10^{-1})	(9.11×10^{-1})	(8.85×10^{-7})	(6.93×10^{-1})
Physiological	217	117	**491**	236	274	423
	(1.21×10^{-1})	(4.78×10^{-1})	(**2.02×10^{-3}**)	(4.27×10^{-1})	(3.27×10^{-1})	(8.58×10^{-1})
Regulation	*20*	*11*	108	*31*	*29*	71
	(8.16×10^{-5})	(6.95×10^{-3})	(6.66×10^{-2})	(2.83×10^{-2})	(3.83×10^{-4})	(1.91×10^{-1})

NOTE: Numbers are the numbers of up-regulated probe sets in each comparison. *P* values (in parentheses) are from χ^2 tests of departure from expected numbers in each GO category, based on the frequency of probe sets in each category on the GeneChip. Italics denote fewer up-regulated probe sets than expected by chance; bold denotes more up-regulated probe sets than expected by chance.

TABLE 16.2 Molecular Function GO Categories

	Male-Specific, *n*		Female-Specific, *n*		Both Sexes, *n*	
GO Category	F > S	S > F	F > S	S > F	F > S	S > F
Transcription regulator activity	*13* (1.24×10^{-4})	*6* (7.31×10^{-3})	73 (1.44×10^{-1})	*16* (6.79×10^{-3})	*25* (1.68×10^{-2})	*41* (5.31×10^{-3})
Enzyme regulator activity	13 (8.49×10^{-1})	4 (2.86×10^{-1})	32 (1.67×10^{-1})	8 (2.07×10^{-1})	9 (1.14×10^{-1})	22 (5.36×10^{-1})
Signal transducer activity	33 (1.72×10^{-1})	10 (2.04×10^{-1})	*41* (6.51×10^{-6})	32 (4.03×10^{-1})	*20* (8.66×10^{-6})	68 (3.96×10^{-1})
Translation regulator activity	2 (4.31×10^{v1})	1 (6.00×10^{-1})	10 (1.32×10^{-1})	1 (2.28×10^{-1})	5 (6.08×10^{-1})	1 (4.25×10^{-1})
Binding	*68* (1.18×10^{-4})	*33* (8.76×10^{-3})	**257** (**1.43×10^{-12}**)	84 (5.99×10^{-1})	*74* (1.13×10^{-6})	*139* (1.26×10^{-4})
Antioxidant activity	2 (4.95×10^{-1})	2 (1.28×10^{-1})	1 (3.64×10^{-1})	1 (9.07×10^{-1})	0 (NA)	5 (1.05×10^{-1})
Catalytic activity	**169** (**1.69×10^{-5}**)	72 (2.36×10^{-1})	250 (9.40×10^{-1})	**146** (**7.63×10^{-4}**)	**224** (**1.71×10^{-14}**)	**278** (**1.86×10^{-3}**)
Structural molecule activity	*15* (7.90×10^{-3})	13 (7.89×10^{-1})	40 (7.43×10^{-2})	31 (2.50×10^{-1})	*9* (5.27×10^{-6})	60 (2.31×10^{-1})
Motor activity	2 (5.54×10^{-1})	1 (6.98×10^{-1})	5 (7.96×10^{-1})	3 (8.44×10^{-1})	2 (4.06×10^{-1})	3 (2.72×10^{-1})
Transporter activity	49 (6.28×10^{-2})	**28** (**1.87×10^{-2}**)	*34* (1.75×10^{-6})	**45** (**4.65×10^{-2}**)	51 (2.44×10^{-1})	81 (9.00×10^{-2})

NOTES: Numbers are the numbers of up-regulated probe sets in each comparison. *P* values (in parentheses) are from χ^2 tests of departure from expected numbers in each GO category, based on the frequency of probe sets in each category on the GeneChip. Italics denote fewer up-regulated probe sets than expected by chance; bold denotes more up-regulated probe sets than expected by chance. NA, not applicable.

in postmating behavior ($P = 6.00 \times 10^{-4}$), spermstorage ($P = 4.67 \times 10^{-7}$), lipid metabolism ($P = 5.77 \times 10^{-5}$), and defense response ($P = 9.71 \times 10^{-3}$), and that have hydrolase activity ($P = 2.53 \times 10^{-4}$) are up-regulated more often than expected in Fast relative to Slow males. Slow males are distinguished from Fast males by overrepresentation of up-regulated transcripts involved in postmating behavior ($P = 1.19 \times 10^{-12}$), insemination ($P = 4.78 \times 10^{-11}$), sperm displacement ($P = 1.11 \times 10^{-12}$), and steroid metabolism ($P = 4.35 \times 10^{-5}$).

Because 21% of the probe sets on the array are implicated in the transcriptional response to selection, one expects the same fraction of loci in any pathway to be represented by chance. Nevertheless, it is gratifying that transcript levels of many genes that have previously been implicated in mating behavior have been altered by selection. These include several male-specific transcripts and accessory gland proteins (*Acp26Aa*, *Acp26Ab*, *Acp29AB*, *Acp36CD*, *Mst35Bb*, *Mst57Da*, *Mst84Dd*, and *Mst89B*) and genes involved in sex determination (*doublesex*, *transformer*, *transformer 2*, and *fruitless*), circadian rhythm/courtship song (*nonA* and *period*), and dopamine metabolism (*ebony*). In addition, transcript abundance of two of the genes identified by mapping QTLs that cause variation in mating behavior between Oregon and 2b and *18 wheeler* and *Enhancer of split* (Moehring and Mackay, 2004) was also altered between these selection lines. Novel candidate genes affecting mating behavior implicated by changes in transcript abundance between selection lines include 15 of the 39 members of the predicted family of odorant binding proteins; genes involved in circadian rhythm, larval locomotion, learning and memory, and olfactory behavior; and genes involved in neurogenesis (Table 16.3).

DISCUSSION

We have shown that *Drosophila* mating speed responds to artificial selection, and that response is largely attributable to an increase in female copulation latency (i.e., a reduction in female receptivity). Thus, there is naturally segregating variation for at least one component of mating behavior. The average divergence in mating speed from generations 25–29 is seemingly large at 113 min but is only 3.5 times the phenotypic standard deviation, which is a rather modest long-term selection response (Falconer and Mackay, 1996). To date, analysis of mating behavior in these lines has been confined to no-choice tests in which each female is paired with a single male. In the future, it will be of considerable interest to conduct choice mating tests to determine whether preference for Slow males has evolved as a correlated response to increased discrimination of Slow females (or vice versa) and to assess correlated responses of fertility, longevity, and other behavioral traits in the selection lines.

TABLE 16.3 Genes with Altered Transcript Abundance in Lines Selected for Increased and Decreased Copulation Latency

Trait	Gene[a]	Comparison	Fold
Olfactory-binding protein	*Obp8a*	F > S	1.37
	Obp18a	S > F	1.24
	Obp19c	S♀ > F♀	1.51
	Obp44a	S♀ > F♀	1.30
	Obp50b	F♂ > S♂	1.39
	Obp50c	F♂ > S♂	1.39
	Obp51a	F♂ > S♂	1.28
	Obp56a	S♀ > F♀	1.94
	Obp56d	S♀ > F♀, F♂ > S♂	1.13, 1.25
	Obp57a	S > F	1.37
	Obp57b	F > S	1.39
	Obp57c	S♀ > F♀	1.18
	Obp83c	S♀ > F♀	1.75
	Obp99b	S > F	2.07
	Obp99c	F > S	1.09
Circadian rhythm	*Pka-R2*	F♀ > S♀	1.10
	Cry	F♂ > S♂	1.20
	Clk	F♂ > S♂	1.19
	sgg	F♂ > S♂	1.19
	tim	S > F	1.48
	Pdf	S > F	1.28
Larval locomotion	*sbb*	S > F	1.30
	for	S > F	1.07
Learning and memory	*Fas2*	S♀ > F♀	1.18
	Pka-R1	S♀ > F♀	1.11
	pum	F♀ > S♀	1.14
Olfaction	*Van*	F♀ > S♀	1.26
Neurogenesis	*pbl*	F♀ > S♀	1.07
	stc	F♀ > S♀	1.04
	lola	S♀ > F♀	1.15
	Ras85D	F♀ > S♀	1.11
	robo	S♀ > F♀	1.35
	Dl	F♀ > S♀	1.25
	disco	S♀ > F♀	1.73
	ab	S > F	1.31
	aay	S > F	1.06
	dlg1	S > F	1.17
	sktl	S > F	1.20
	dally	S > F	1.23
	pnt	F > S	1.51
	elav	S > F	1.35
	numb	S > F	1.43
	cpo	S > F	1.45
Catecholamine metabolism	*Dat*	S♀ > F♀	1.10
Regulation of insulin receptor pathway	*foxo*	F♂ > S♂	1.28
Hsp90 chaperone, stress response	*Hsp83*	F♂ > S♂	1.14
Protein folding, stress response	*Hsp27*	F♀ > S♀	1.14

TABLE 16.3 Continued

Trait	Gene[a]	Comparison	Fold
Tryptophan synthesis serotonin metabolism	*Hn*	F♂ > S♂	1.12
Tyrosine metabolism, defense response	*Bc*	S♀ > F♀	1.33
Specification of segmental identity	*tsh*	S > F	1.45
Female gonad development	*fz2*	S > F	1.29
Cell proliferation	*l(2)gl*	S > F	1.29

[a]See Manning (1961) for full gene names and descriptions.

The response to selection for mating speed was highly asymmetrical, as is often observed for traits that are major components of fitness (Falconer and Mackay, 1996; Frankham, 1990), including previous studies selecting for divergent mating speed in *Drosophila* (Manning, 1961, 1963; Sherwin, 1975; Spuhler *et al.*, 1978). Asymmetrical responses of fitness traits to selection are generally attributable to directional dominance and/or genetic asymmetry, such that alleles increasing fitness are at high frequency. Because we did not observe inbreeding depression for mating speed, as would be expected if deleterious alleles were recessive, we infer that the most likely cause of asymmetry was the segregation of low-frequency alleles affecting increased female copulation latency in the base population.

The transcriptional response to selection for mating speed was profound, with >3,700 probe sets (≈21% of the total number on the microarray) exhibiting a divergence in message levels between the Fast and Slow replicate lines, at a stringent false discovery rate of 0.001. In contrast, a previous study of transcriptional response to long-term selection for geotaxis behavior (Toma *et al.*, 2002) found divergence in message levels for only 5% of the genes assessed. We speculate that this difference is attributable to a difference in criteria for declaring significance: Toma *et al.* (2002) used a 2-fold change threshold, although we used a statistical test. We found that changes in transcript abundance of 10% or even less were often statistically significant.

The chromosomal locations of genes with male-specific changes in expression were nonrandom: the *Drosophila* X chromosome is depauperate for genes that are up-regulated in males. This is an apparently general phenomenon (Parisi *et al.*, 2003; Rantz *et al.*, 2003). X chromosome demasculinization is perhaps attributable to selection against genes that are advantageous in males but deleterious to females (Parisi *et al.*, 2003). The transcriptional response to selection is attributable to genes that have causally responded to selection and that are coregulated by these genes. Because the transcriptional response to single mutations with subtle phe-

notypic effects can involve >100 coregulated genes (Anholt *et al.*, 2003), the number of selected loci causing the changes in transcript abundance between the selection lines could well be rather modest. It will be necessary to map the QTLs causing divergence between the selection lines to disentangle causal vs. consequential transcriptional responses to selection.

Nevertheless, genes exhibiting parallel changes in transcript abundance between replicate Fast and Slow selected lines are candidate genes affecting mating behavior. Could 21% of the genome really be responsible for regulating mating speed? Recent studies assessing subtle quantitative effects of *P* element insertional mutations on numbers of sensory bristles (Norga *et al.*, 2003) and resistance to starvation stress (Harbison *et al.*, 2004) have concluded that >20% of the genome affects each of these traits. These results imply massive pleiotropy: the same genes affect multiple complex traits. Thus, genes regulating mating behavior are as likely to be genes involved in neurogenesis, metabolism, development, and general cellular processes as genes with specific effects on behavior (Sokolowski, 2001). In fact, the same loci may affect multiple behaviors. *Pigment dispersing factor* (*Pdf*) and *cyrptochrome* (*cry*) were defined based on the involvement in circadian rhythm but were up-regulated in lines selected for positive geotaxis and confirmed to affect geotaxis behavior in functional tests (Toma *et al.*, 2002). We note that *Pdf* and *cry* are also differentially expressed between the Fast and Slow mating speed selection lines, implicating them in mating behavior.

CONCLUSION

In the future, functional studies will be required to test the extent to which transcript profiling of divergent selection lines accurately predicts genes that directly affect the selected trait. One such test is to assess whether mutations at candidate genes implicated by the analysis of differential transcript abundance affect the trait. A complication here is that mutational effects may be subtle, of the order of naturally occurring variation within and between strains. Many *Drosophila* mutant stocks have been generated in segregating genetic backgrounds, often containing multiple mutations. It is thus difficult to ascertain whether any difference in the phenotype of a complex trait and a wild-type control is attributable to the mutation in the candidate gene, ancillary mutations, or QTLs affecting the trait segregating between the mutant stock and the wild-type control. Multiple generations of backcrossing the mutation to a common control stock can abrogate this problem (Toma *et al.*, 2002). A more convincing test is to obtain viable hypomorphic mutations that have been generated in a coisogenic background and compare their effect on the

trait to the coisogenic control strain phenotype (Harbison *et al.*, 2004; Norga *et al*., 2003). A subset of the *P* element insertion lines generated by the Berkeley *Drosophila* Gene Disruption Project (Bellen *et al*., 2004) are in coisogenic backgrounds, as is the Exelixis collection of mutations (Parks *et al*., 2004). These resources will prove invaluable in testing the predictions of the expression analyses. Another functional test is to perform quantitative complementation tests of mutations at the candidate genes with the selection lines, to assess whether coregulation of transcription translates to epistasis at the level of trait phenotype. Mutations are not available for many of the genes with altered transcript abundance in response to selection (e.g., the odorant-binding proteins). In this case, one can use linkage disequilibrium mapping (Risch and Merikangas, 1996) to assess whether molecular polymorphisms in these candidate genes are associated with naturally occurring variations in mating behavior.

ACKNOWLEDGMENTS

We are grateful to M. J. Zanis for his PERL programming skills, which helped greatly in the organization of our statistical analyses. This work was supported by National Institutes of Health Grants F32 GM66603 (to T.J.M.), R01 GM 58260, R01 GM 45146, and P01 45344 (to T.F.C.M.). This is a publication of the W. M. Keck Center for Behavioral Biology.

REFERENCES

Ackerman, S. L. & Seigal, R. W. (1986) Chemically reinforced conditioned courtship in Drosophila: Responses of wild type and the *dunce*, *amnesiac* and *don giovanni* mutants. *J. Neurogenet.* **3,** 111–123.

Anholt, R. R. H. & Mackay, T. F. C. (2004) Genetic analysis of complex behaviors in *Drosophila*. *Nat. Rev. Genet.* **5,** 838–849.

Anholt, R. R. H., Dilda, C. L., Chang, S., Fanara, J. J., Kulkarni, N. H., Ganguly, I., Rollmann, S. M., Kamdar, K. P. & Mackay, T. F. C. (2003) The genetic architecture of odor-guided behavior in *Drosophila*: Epistasis and the transcriptome. *Nat. Genet.* **35,** 180–184.

Arthur, B. I., Jallon, J.-M., Caflisch, B., Choffat, Y. & Nothiger, R. (1998) Sexual behavior in *Drosophila* is irreversibly programmed during a critical period. *Curr. Biol.* **8,** 1187–1190.

Baker, B. S., Taylor, B. J. & Hall, J. C. (2001) Are complex behaviors specified by dedicated regulatory genes? Reasoning from *Drosophila*. *Cell* **105,** 13–24.

Bellen, H. J., Levis, R. W., Liao, G., He, Y., Carlson, J. W., Tsang, G., Evans-Holm, M., Hiesinger, R., Schultze, K., Rubin, G. M., *et al.* (2004) The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. *Genetics* **167,** 761–781.

Bernstein, A. S., Neumann, E. & Hall, J. C. (1992) Temporal analysis of tone pulses within the courtship songs of two sibling *Drosophila* species, their interspecific hybrid, and behavioral mutants of *Drosophila melanogaster* (Diptera: Drosophilidae). *J. Insect. Behav.* **5,** 15–36.

Buchner, E. (1991) Genes expressed in the adult brain of *Drosophila* and effects of their mutations on behavior: a survey of transmitter- and second messenger-related genes. *J. Neurogenet.* **7,** 153–192.

Chapman, T., Liddle, L. F., Kalb, J. M., Wolfner, M. F. & Partridge, L. (1995) Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. *Nature* **373,** 241–244.

Clark, A. G., Aguade, M., Prout, T., Harshman, L. G. & Langley, C. H. (1995) Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. *Genetics* **139,** 189–201.

Cook, R. (1980) The extent of visual control in the courtship tracking of *Drosophila melanogaster*. *Biol. Cybern.* **37,** 41–51.

Coyne, J. A. (1989) The genetics of sexual isolation between two sibling species, *Drosophila simulans* and *Drosophila mauritiana*. *Proc. Natl. Acad. Sci. USA* **86,** 5464–5468.

Coyne, J. A. (1993) The genetics of an isolating mechanism between two sibling species of *Drosophila*. *Evolution (Lawrence, Kans.)* **47,** 778–788.

Coyne, J. A. (1996a) Genetics of differences in pheromonal hydrocarbons between *Drosophila melanogaster* and *D. simulans*. *Genetics* **143,** 353–364.

Coyne, J. A. (1996b) Genetics of sexual isolation in male hybrids of *Drosophila simulans* and *D. mauritiana*. *Genet. Res.* **68,** 211–220.

Coyne, J. A., Kim S. Y., Chang, A. S., Lachaise, D. & Elwyn, S. (2002) Sexual isolation between two sibling species with overlapping ranges: *Drosophila santomea* and *Drosophila yakuba*. *Evol. Int. J. Org. Evol.* **56,** 2424–2434.

Crossley, S. & Zuill, E. (1970) Courtship behaviour of some *Drosophila melanogaster* mutants. *Nature* **225,** 1064–1065.

Ewing, A. W. (1988) Cycles in the courtship song of male *Drosophila melanogaster* have not been detected. *Anim. Behav.* **36,** 1091–1097.

Falconer, D. S. & Mackay, T. F. C. (1996) *Introduction to Quantitative Genetics* (Addison–Wesley–Longman, Harlow, Essex, U.K.), 4th Ed.

Ferveur, J.-F., Savarit, F., O'Kane, C. J., Sureau, G., Greenspan, R. J. & Jallon, J.-M. (1997) Genetic feminization of pheromones and its behavioral consequences in *Drosophila* males. *Science* **276,** 1555–1558.

Finley, K. D., Edeen, P. T. Foss, M., Gross, E., Ghbeish, N., Palmer, R. H., Taylor, B. J. & McKeown, M. (1998) *dissatisfaction* encodes a tailless-like nuclear receptor expressed in a subset of CNS neurons controlling *Drosophila* sexual behavior. *Neuron* **21,** 1363–1374.

Fleishmann, I., Dauwalder, B., Chapman, T., Cotton, B. & Kubli, E. (1995) Analysing the sex-peptide reaction-cascade in *Drosophila melanogaster* using brain mutants. *J. Neurogenet.* **10,** 26–27.

FlyBase Consortium (2003) The FlyBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res.* **31,** 172–175.

Frankham, R. (1990) Are responses to artificial selection for reproductive fitness characters consistently asymmetrical? *Genet. Res.* **56,** 35–42.

Gailey, D. A., Jackson, F. R. & Siegel, R. W. (1982) Male courtship in *Drosophila*: the conditioned response to immature males and its genetic control. *Genetics* **102,** 771–782.

Gailey, D. A., Jackson, F. R. & Siegel, R. W. (1984) Conditioning mutations in *Drosophila melanogaster* affect an experience-dependent behavioral modification in courting males. *Genetics* **106,** 613–623.

Gailey, D. A., Taylor, B. J. & Hall, J. C. (1991) Elements of the *fruitless* locus regulate development of the muscle of Lawrence, a male-specific structure in the abdomen of *Drosophila melanogaster* adults. *Development (Cambridge, U.K.)* **113,** 879–890.

Greenspan, R. J. (1995) Understanding the genetic construction of behavior. *Sci. Am.* **272** (4), 72–78.

Hall, J. C. (1994) The mating of a fly. *Science* **264,** 1702–1714.

Hall, J. C., Siegel, R. W., Tomkins, L. & Kyriacou, C. P. (1980) Neurogenetics of courtship in *Drosophila*. *Stadler Genet. Symp.* **12,** 43–82.

Harbison, S. T., Yamamoto, A. H., Fanara, J. J., Norga, K. K. & Mackay, T. F. C. (2004) Quantitative trait loci affecting starvation resistance in *Drosophila melanogaster*. *Genetics* **166,** 1807–1823.

Heifetz, Y., Lung, O., Frongillo, E. A. & Wolfner, M. F. (2000) The *Drosophila* seminal fluid protein Acp26Aa stimulates release of oocytes by the ovary. *Curr. Biol.* **10,** 99–102.

Hollocher, H., Ting, C.-T., Wu, M.-L. & Wu, C.-I. (1997) Incipient speciation by sexual isolation in *Drosophila melanogaster*: Extensive genetic divergence without reinforcement. *Genetics* **147,** 1191–1201.

Jackson, F. R., Gailey, D. A. & Siegel, R. W. (1983) Biological rhythm mutants affect an experience-dependent modification of male courtship behavior in *Drosophila*. *J. Comp. Physiol.* **151,** 545–552.

Jallon, J.-M. (1984) A few chemical words exchanged by *Drosophila* during courtship and mating. *Behav. Genet.* **14,** 441–477.

Joiner, M. A. & Griffith, L. C. (1997) CaM kinase II and visual input modulate memory formation in the neuronal circuit controlling courtship conditioning. *J. Neurosci.* **17,** 9384–9391.

Kyriacou, C. P. (1990) The molecular ethology of the *period* gene in *Drosophila*. *Behav. Genet.* **20,** 191–211.

Kyriacou, C. P. & Hall, J. C. (1984) Learning and memory mutations impair acoustic priming of mating behaviour in *Drosophila*. *Nature* **308,** 62–65.

Kyriacou, C. P. & Hall, J. C. (1985) Action potentials stop a biological clock in *Drosophila*. *Nature* **314,** 171–173.

Kyriacou, C. P. & Hall, J. C. (1994) Genetic and molecular analysis of *Drosophila* behavior. *Adv. Genet.* **31,** 139–186.

Kyriacou, C. P., Burnet, B. & Connolly, K. (1978) The behavioural basis of overdominance in competitive mating success at the *ebony* locus of *Drosophila melanogaster*. *Anim. Behav.* **26,** 1195–1206.

Long, A. D., Mullaney, S. L., Mackay, T. F. C. & Langley, C. H. (1996) Genetic interactions between naturally occurring alleles at quantitative trait loci and mutant alleles at candidate loci affecting bristle number in *Drosophila melanogaster*. *Genetics* **144,** 1497–1518.

Lung, Y. O. & Wolfner, M. F. (2001) Identification and characterization of the major *Drosophila melanogaster* mating plug protein. *Insect. Biochem. Mol. Biol.* **31,** 543–551.

Mackay, T. F. C. (2001) The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35,** 303–339.

Mackay, T. F. C. & Fry, J. D. (1996) Polygenic mutation in *Drosophila melanogaster*: Genetic interactions between selection lines and candidate quantitative trait loci. *Genetics* **144,** 671–688.

Manning, A. (1961) The effects of artificial selection for mating speed in *Drosophila melanogaster*. *Anim. Behav.* **9,** 82–92.

Manning, A. (1963) Selection for mating speed in *Drosophila melanogaster* based on the behaviour of one sex. *Anim. Behav.* **11,** 116–120.

Mayr, E. (1942) *Systematics and the Origin of Species* (Columbia Univ. Press, New York).

Moehring, A. J. & Mackay, T. F. C. (2004) The quantitative genetic basis of male mating behavior in *Drosophila melanogaster*. *Genetics* **167,** 1249–1263.

Moehring, A. J., Li, J., Schug, M. D., Smith, S. G., DeAngelis M., Mackay, T. F. C. & Coyne, J. A. (2004) Quantitative trait loci for sexual isolation between *Drosophila simulans* and *D. mauritiana*. *Genetics* **167,** 1265–1274.

Nakayama, S., Kaiser, K. & Aigaki, T. (1997) Ectopic expression of sex-peptide in a variety of tissues in *Drosophila* females using the *P[GAL4]* enhancer trap system. *Mol. Gen. Genet.* **254,** 449–455.

Neckameyer, W. S. (1998) Dopamine modulates female sexual receptiveness in *Drosophila melanogaster*. *J. Neurogenet.* **12,** 101–114.

Noor, M. A. F. (1997) Genetics of sexual isolation and courtship dysfunction in male hybrids of *Drosophila pseudoobscura* and *D. persimilis*. *Evolution (Lawrence, Kans.)* **51,** 809–815.

Noor, M. A. F., Grams, K. L., Bertucci, L. A., Almendarez, Y., Reiland, J. & Smith, K. R. (2001) The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. *Evolution (Lawrence, Kans.)* **55,** 512–521.

Norga, K. K., Gurganus, M. C., Dilda, C. L., Yamamoto, A., Lyman, R. F., Patel, P. H., Rubin, G. M., Hoskins, R. A., Mackay, T. F. C. & Bellen, H. J. (2003) Quantitative analysis of bristle number in *Drosophila* mutants identifies genes involved in neural development. *Curr. Biol.* **13,** 1388–1397.

O'Dell, K. M. & Kaiser, K. (1995) Functional dissection of the *Drosophila* mushroom bodies by selective feminization of genetically defined subcompartments. *Neuron* **15,** 55–61.

Parisi, M., Nuttall, R., Naiman, D., Bouffard, G., Malley, J., Andrews, J., Eastman, S. & Oliver, B. (2003) Paucity of genes on the *Drosophila X* chromosome showing male-biased expression. *Science* **299,** 697–700.

Parks, A. L., Cook, K. R., Belvin, M., Dompe, N. A., Fawcett, R., Huppert, K., Tan, L. R., Winter, C. G., Bogart, K. P., Deal, J. E., *et al.* (2004) Systematic generation of high-resolution deletion coverage of the *Drosophila melanogaster* genome. *Nat. Genet.* **36,** 288–292.

Pasyukova, E. G., Vieira, C. & Mackay, T. F. C. (2000) Deficiency mapping of quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Genetics* **156,** 1129–1146.

Peixoto, A. & Hall, J. C. (1998) Analysis of temperature sensitive mutants reveals new genes involved in the courtship song of *Drosophila*. *Genetics* **148,** 827–838.

Rantz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D. & Hartl, D. L. (2003) Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300,** 1742–1745.

Rendel, J. M. (1951) Mating of *ebony*, *vestigial* and wild type *Drosophila melanogaster* in light and dark. *Evolution (Lawrence, Kans.)* **5,** 226–230.

Risch, N. & Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science* **273,** 1516–1517.

Ryner, L. C., Goodwin, S. F., Castrillon, D. H., Anand, A., Villella, A., Baker, B. S., Hall, J. C., Taylor, B. J. & Wasserman, S. A. (1996) Control of male sexual behavior and sexual orientation in *Drosophila* by the *fruitless* gene. *Cell* **87,** 1079–1089.

Sherwin, R. N. (1975) Selection for mating activity in two chromosomal arrangements of *Drosophila pseudoobscura*. *Evolution (Lawrence, Kans.)* **29,** 519–530.

Siegel, R. W. & Hall, J. C. (1979) Conditioned responses in courtship behavior of normal and mutant *Drosophila*. *Proc. Natl. Acad. Sci. USA* **76,** 3430–3434.

Sokolowski, M. B. (2001) *Drosophila*: Genetics meets behaviour. *Nat. Rev. Genet.* **2,** 879–890.

Speith, H. T. & Ringo, J. M. (1983). Mating behavior and sexual isolation in *Drosophila*. In *The Genetics and Biology of Drosophila*, eds. Ashburner, M., Carson, H. L. & Thompson, J. N. (Academic, London), Vol. 3c, pp. 223–284.

Spuhler, K. P., Crumpacker, D. W., Williams J. S. & Bradley, B. P. (1978) Response to selection for mating speed and changes in gene arrangement frequencies in descendants from a single population of *Drosophila pseudoobscura*. *Genetics* **89,** 729–749.

Storey, J. D. & Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100,** 9440–9445.

Sturtevant, A. H. (1915) Experiments in sexual recognition and the problems of sexual selection in *Drosophila*. *J. Anim. Behav.* **5,** 351–366.

Takahashi, A., Tsaur, S.-C., Coyne, J. A. & Wu, C.-I. (2001) The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **98,** 3920–3925.

Taylor, B. J., Villella, A., Ryner, L. C., Baker, B. S. & Hall, J. C. (1994) Behavioral and neurobiological implications of sex-determining factors in *Drosophila*. *Dev. Genet.* **15,** 275–296.

Tempel, B. L., Livingstone, M. S. & Quinn, W. G. (1984) Mutations in the *Dopa decarboxylase* gene affect learning in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **81,** 3577–3581.

Ting, C.-T., Takahashi, A. & Wu, C.-I. (2001) Incipient speciation by sexual isolation in *Drosophila*: Concurrent evolution at multiple loci. *Proc. Natl. Acad. Sci. USA* **98,** 6709–6713.

Toma, D. P., White, K. P., Hirsch, J. & Greenspan, R. J. (2002) Identification of genes involved in *Drosophila melanogaster* geotaxis, a complex behavioral trait. *Nat. Genet.* **31,** 349–353.

Tompkins, L. & McRobert, S. P. (1995) Behavioral and pheromonal phenotypes associated with expression of loss-of-function mutations in the *Sex-lethal* gene of *Drosophila melanogaster*. *J. Neurogenet.* **9,** 219–226.

Tompkins, L., Gross, A. C., Hall, J. C., Gailey, D. A. & Siegel, R. W. (1982) The role of female movement in the sexual behavior of *Drosophila melanogaster*. *Behav. Genet.* **12,** 295–307.

Villella, A. & Hall, J. C. (1996) Courtship anomalies caused by the *doublesex* mutations in *Drosophila melanogaster*. *Genetics* **143,** 331–344.

Waterbury, J. A., Jackson, L. L. & Schedl, P. (1999) Analysis of the *doublesex* female protein in *Drosophila melanogaster*: Role in sexual differentiation and behavior and dependence on *intersex*. *Genetics* **152,** 1653–1667.

Wheeler, D. A., Kulkarni, S. J., Gailey, D. A. & Hall, J. C. (1989) Spectral analysis of courtship songs in behavioral mutants of *Drosophila melanogaster*. *Behav. Genet.* **19,** 503–528.

Wolfner, M. F., Harada, H. A., Bertram, M. J., Stelick, T. J., Kraus, K. W., Kalb, J. M., Lung, Y. O., Neubaum, D. M., Park, M. & Tram, U. (1997) New genes for male accessory gland proteins in *Drosophila melanogaster*. *Insect. Biochem. Mol. Biol.* **27,** 825–834.

Wu, C.-I., Hollocher, H., Begun, D. J., Aquadro, C. F., Xu, Y. & Wu, M.-L. (1995) Sexual isolation in *Drosophila melanogaster*: A possible case for incipient speciation. *Proc. Natl. Acad. Sci. USA* **92,** 2519–2523.

Yamamoto, D. & Nakano, Y. (1998) Genes for sexual behavior. *Biochem. Biophys. Res. Commun.* **246,** 1–6.

Zhang, S. D. & Odenwald, W. F. (1995) Misexpression of the *white* (*w*) gene triggers male-male courtship in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **92,** 5525–5529.

Zouros, E. (1981) The chromosomal basis of sexual isolation in two sibling species of *Drosophila*: *D. arizonensis* and *D. mojavensis*. *Genetics* **97,** 703–718.

17

Genomes, Phylogeny, and Evolutionary Systems Biology

MÓNICA MEDINA*

With the completion of the human genome and the growing number of diverse genomes being sequenced, a new age of evolutionary research is currently taking shape. The myriad of technological breakthroughs in biology that are leading to the unification of broad scientific fields such as molecular biology, biochemistry, physics, mathematics, and computer science are now known as systems biology. Here, I present an overview, with an emphasis on eukaryotes, of how the postgenomics era is adopting comparative approaches that go beyond comparisons among model organisms to shape the nascent field of evolutionary systems biology.

Systems biology is in the eye of the beholder.

Leroy Hood

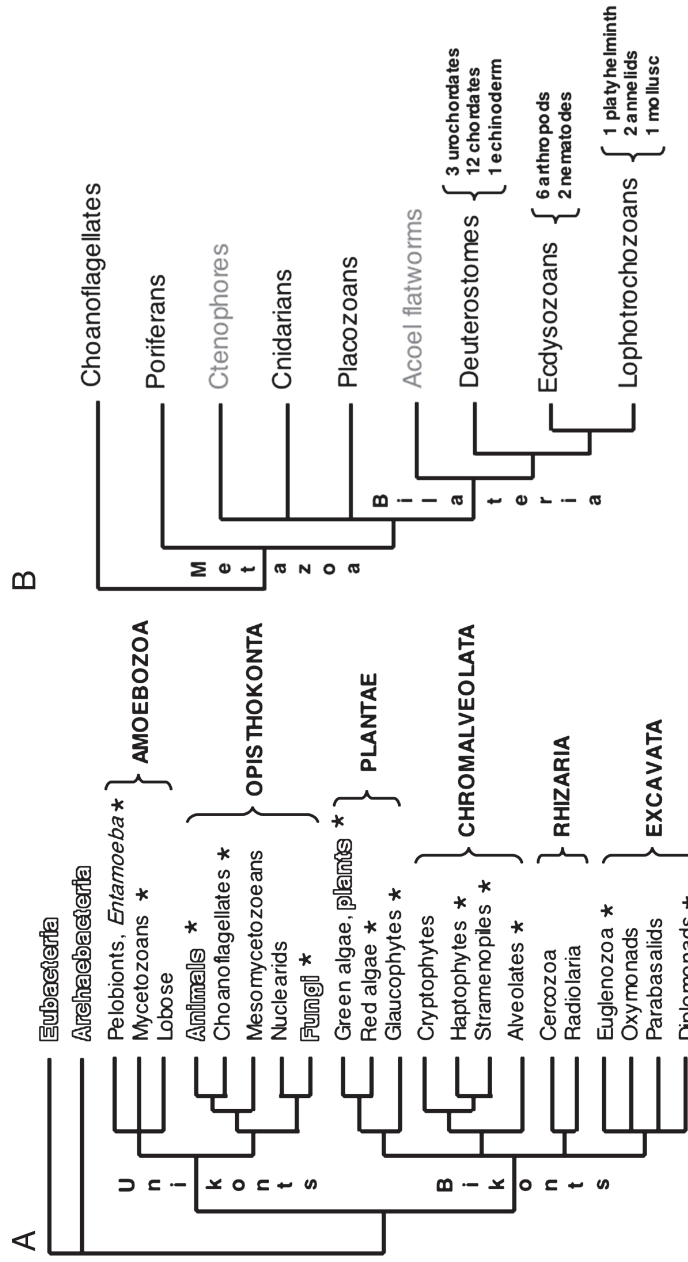
Only in the last decade have we had access to nearly complete genomes of a diversity of organisms allowing for large-scale comparative analysis. The access to this immense amount of data

*Department of Evolutionary Genomics, Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598. Present address: School of Natural Sciences, University of California, P.O. Box 2039, Merced, CA 95344. E-mail: mmedina@ucmerced.edu.

is providing profound insight into the tree of life at all levels of divergence (Fig. 17.1*A*). It is thus not surprising that understanding phylogenetic relationships is a prevalent research goal among not only evolutionary biologists but also all scientists interested in the organization and function of the genome. New genome sequences and analysis methods are helping improve our understanding of phylogeny, and at the same time improved phylogenies and phylogenetic theory are generating a better understanding of genome evolution. Currently however, the level of genome sequencing for different branches of the tree of life is far from equivalent. Prokaryotic genome projects are abundant, mainly due to their small genome sizes, with >200 genomes already published and at least 500 currently in progress (www. genomesonline.org). In contrast, <300 eukaryotic genomes are either finished or in progress (www.genomes online.org). Nevertheless, these data are starting to have a major impact on our understanding of eukaryotic evolution.

These new genomic data have informed our understanding of phylogenetic relationships, and the emerging consensus topologies are adding new insight to the small subunit ribosomal RNA phylogenies. For example, the topology of the ribosomal eukaryotic tree has been recently redrawn with the use of genomic signatures that place the root of all eukaryotic life between two newly uncovered major clades, Unikonts and Bikonts (Fig. 17.1*A*). Unikonts, which contain the heterotrophic groups Opisthokonta and the Amebozoa, share a derived three-gene fusion of enzymeencoding genes in the pyrimidine synthesis pathway (Stechmann and Cavalier-Smith, 2003), whereas Bikonts, which contain the remaining eukaryotic clades, share another derived gene fusion between dihydro folate reductase and thymidine synthase (Stechmann and Cavalier-Smith, 2002). All photosynthetic groups of primary and secondary plastid symbiotic origins are now thought to be within the Bikonts. Although the animal, fungal, and plant lineages are the most widely represented in terms of genome initiatives (Fig. 17.1*B–D*), it is significant that multiple protistan genome projects have also been initiated by the interest of diverse scientific communities, including parasitologists (Gardner *et al.*, 2002), plant pathologists (Waugh *et al.*, 2000), oceanographers (Armbrust *et al.*, 2004), and evolutionary biologists (www.biology.uiowa.edu/workshop).

As more whole-genome projects are being completed, postgenomic biology is also providing insight into the function of biological systems by the use of new high-throughput bioanalytical methods, information technology, and computational modeling. This new revolution in biology has become known as systems biology (Hood, 2003). In addition to shifting approaches to biological research from reductionist strategies to pathway- and system-level strategies (Hartwell *et al.*, 1999), another paradigm



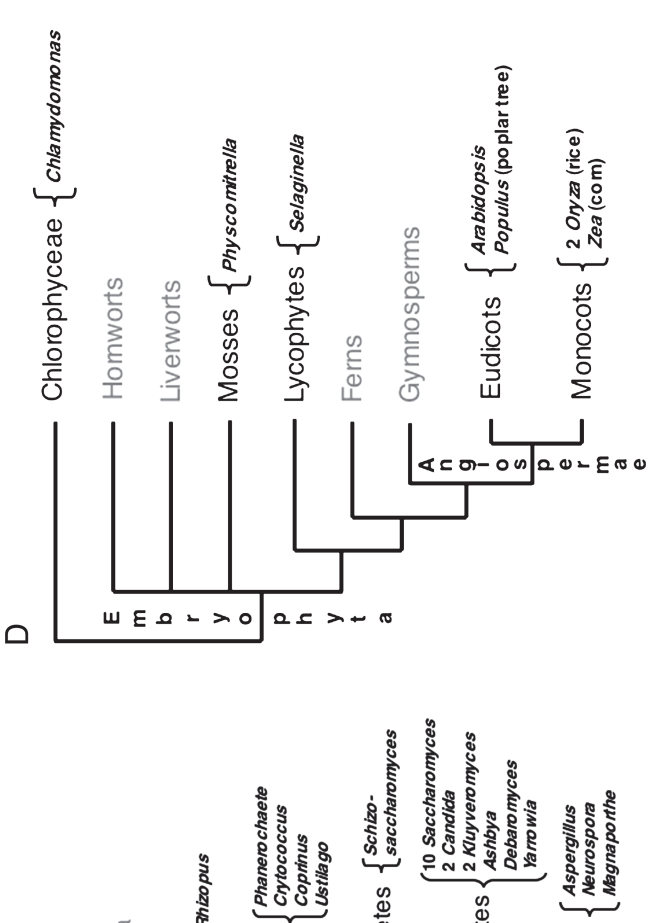


FIGURE 17.1 Current consensus eukaryotic tree. (*A*) The large subclades within Unikonts and Bikonts are recovered by a combination of multiple gene phylogenies, EST data, and genomic level characters (Stechmann and Cavalier-Smith, 2003; Simpson and Roger, 2004; Bhattacharya *et al.*, 2004). Six major eukaryotic groups are now recognized although resolution within them is still lacking. The placement of the root is based on two gene fusion events (Stechmann and Cavalier-Smith, 2002, 2003). Lineages where whole-genome projects are in progress are marked with asterisks. Lineages being studied by large postgenomic initiatives are shadowed. (*B*) Metazoan consensus phylogeny of major branches (Adoutte *et al.*, 1999; Medina *et al.*, 2001; Ruiz-Trillo *et al.*, 2002) and a conservative estimate of finished and ongoing genome projects (highlighted in black). (*C*) Fungal consensus phylogeny (Berbee and Taylor, 1993; Hedges, 2002) and estimate of ongoing genome projects (www.broad.mit.edu/annotation/fungi/fgi) (highlighted in black). (*D*) Consensus phylogeny of green plants (Hedges, 2002; Pryer *et al.*, 2002) and estimate of ongoing genome projects (highlighted in black).

is rapidly emerging, namely the use of phylogenetically based inference in systems biology. Before the genomic revolution, research questions were typically addressed within a single model organism, with only occasional comparative studies when similar information was available for another organism. These comparisons were made between distantly related taxa, and the evolutionary implications were rarely mentioned or taken into account. The increasing importance of comparative analysis is evident in the growing proportion of new prokaryotic genome projects that have been chosen primarily because of their phylogenetic relationship to model organisms, such as *Escherichia coli* and *Bacillus subtilis* and their corresponding related taxa. This same trend is occurring for eukaryotes. Some prominent examples are the multiple *Saccharomyces* genome projects and those of other ascomycote fungi, the several *Plasmodium* projects and other genome initiatives for apicomplexan taxa, the numerous *Caenorhabditis* and other nematode genome projects, the multiple *Drosophila* and arthropod genome projects, and the large number of primate and mammalian genome projects.

GENOMES AND PHYLOGENY OF HIGHER EUKARYOTES

Metazoa

The sampling of the metazoan tree, and in particular of the chordate branch, was undertaken primarily due to the usefulness of the genomes in understanding human biology. However, this larger genomic dataset is already providing a powerful tool for comparative analysis and more accurate evolutionary inference. Deeper divergences in the Metazoan tree have become the target of major scrutiny due to the interest in comparative developmental genetics (Fig. 17.1*B*). Based on molecular phylogenies, the bilaterian phyla have been rearranged into three large clades, deuterostomes, lophotrochozoans, and ecdysozoans, these last two being sister taxa inside the protostome clade. At present, there is still debate regarding the placement of nematodes in the tree (i.e., the Ecdysozoa vs. Coelomata hypotheses) because analysis of genomic data currently challenges the placement of *Caenorhabditis elegans* as an ecdysozoan (Dopazo *et al.*, 2004; Wolf *et al.*, 2004).

In addition to the traditional developmental model organisms, genomes from unrepresented protostome (Annelida, Platyhelmintha, and Mollusca) and basal phyla are now being sequenced (Porifera, Placozoa, and Cnidaria) (www.jgi.doe.gov/sequencing/cspseqplans.html). Finally, another node in the tree of life that has gained recent interest is that of the choanoflagellates, a unicellular sister group to metazoans (King, 2004).

Ribosomal phylogenies suggest that choanoflagellates are the most likely unicellular lineage to have shared common ancestry with the multicellular animals (Medina *et al.*, 2003), but there are a few other unicellular protists that also fall out in this part of the tree in other gene phylogenies (Ruiz-Trillo *et al.*, 2004). A choanoflagellate genome project is now in progress, and multiple EST initiatives for unicellular opisthokont protists are also in place.

In summary, postgenomic research on the metazoans is advancing rapidly because of the large number of model organisms, e.g., *C. elegans* (nematode), *Drosophila melanogaster* (fruitfly), *Danio rerio* (zebrafish), *Mus musculus* (mouse), *Rattus norvegicus* (rat), and *Homo sapiens* (human). On the other hand, sequencing metazoan genomes is a major technical challenge, because of higher level of complexity associated with multicellularity and tissue compartmentalization. These challenges are giving a leading role to the yeast and other unicellular systems described in the next section.

Fungi

The initial driving force behind the choice of genome projects in fungi was the prime status of yeast (*Saccharomyces cerevisiae*) as a model organism. Additionally, the relatively small genome size in other fungi has facilitated the explosion of numerous large scale sequencing projects (www.broad.mit.edu/annotation/fungi/fgi). Consensus phylogenies of fungi place the Chitridiomycota as the most basal lineage, followed by the Zygomycota, with Ascomycota and Basidiomycota as sister crown clades (Berbee and Taylor, 1993; Hedges, 2002). Ribosomal phylogenies suggest that the Nucleárid amoeba are the likely unicellular sister group to Fungi (Amaral Zettler *et al.*, 2001; Medina *et al.*, 2003).

After the completion of *S. cerevisiae*, subsequent fungal genome projects were chosen within the Ascomycota (Fig. 17.1*C*) mainly based on phylogenetic proximity (within the Hemiascomycetes) (Dietrich *et al.*, 2004; Dujon *et al.*, 2004; Kellis *et al.*, 2004), although now more distantly related taxa including additional model organisms such as *Neurospora crassa* and *Aspergillus nidulans* have also been sequenced. Basidiomycete genomes have been sequenced (Martinez *et al.*, 2004) or are in progress as well (www.broad.mit.edu/annotation/fungi/fgi). The combination of both *S. cerevisiae* as the best characterized unicellular eukaryote and the thorough comparative genomics allowed by the numerous fungal genome projects have made this branch of the eukaryotic tree an ideal target for validation and improvement of postgenomic approaches.

Plantae

Most of the species diversity of plants is represented in the crown group, the angiosperms, which encompasses the Monocots and the Eudicots. Consensus phylogenies place paraphyletic gymnosperms basal to angiosperms and ferns as a sister group to this clade (Pryer *et al.*, 2002) (Fig. 17.1*D*). The placement of some of the basal groups in the Embryophyta (hornworts, liverworts, and mosses) is still unresolved, although lycophytes are now considered the sister group to the clade containing ferns, gymnosperms, and angiosperms (Hedges, 2002; Pryer *et al.*, 2002) (Fig. 17.1*D*). Finally, multiple sources of evidence point to the green algae as the unicellular sister group to plants (reviewed in Archibald and Keeling, 2002) (Fig. 17.1*D*).

Genome projects for green plants have been hampered by the larger genome sizes of most members of this group. Nonetheless, the first draft Plantae genome published was from *Arabidopsis thaliana*, a flowering plant model organism (*Arabidopsis* Genome Initiative, 2000). Genome drafts of two different rice strains (*Oryza sativa*) have been recently published (Goff *et al.*, 2002; Yu *et al.*, 2002). This effort is now complemented by the completion of a unicellular alga (*Chlamydomonas reinhardtii*), the poplar tree (*Populus trichocarpa*), and partial genome data from corn (*Zea mays*), whereas two basal lineages, the moss *Physcomitrella patens* and the lycophyte *Selaginella moellendorffii*, will be sequenced this year (www.jgi.doe.gov/sequencing/cspseqplans.html). Thus, although more sparse than for metazoans and fungi, the Plantae branch of the eukaryotic tree is rapidly expanding in terms of genomic data. Agricultural interests will likely drive future choice of Plantae genomes to some degree, but decisions will also be influenced by phylogenetic implications as reflected in the recent choice of the *P. patens* and *S. moellendorffii* for genome sequencing.

EVOLUTIONARY SYSTEMS BIOLOGY

With whole-genome data allowing reconstruction of more robust phylogenies for the major eukaryotic groups, new biological questions can now be addressed. Genomic and postgenomic data offer a new ''global'' view of the function of living systems across the tree of life. These new data suggest that biological systems (e.g., a cell) are composed of discrete ''modules'' of interacting components with different functions, and in turn these modules form biological networks that carry out the myriad functions of living systems (Hartwell *et al.*, 1999). Multiple metabolic and regulatory networks are now being characterized in diverse organisms for which reasonably annotated genomes are available. Metabolites, be-

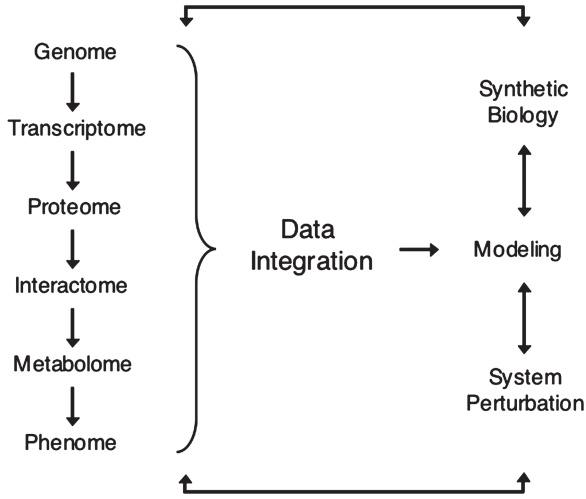


FIGURE 17.2 Overview of systems biology. Hierarchical information from the genome (DNA) to the phenome (phenotype) is integrated to predict mathematical models. These models can then be tested by ''synthetic biology'' (*de novo* design of biological modules) and/or by system perturbations that generate a cycle of hypothesis-driven science (Ge *et al.*, 2003; Ideker *et al.*, 2001; Kitano, 2002).

ing the end products of cellular regulatory networks, are one of the most directly accessible windows into the cell's dynamic phenotype (Fiehn, 2002).

Systems biology is a rapidly expanding field that integrates widely diverse areas of science such as physics, engineering, computer science, mathematics, and biology, toward the goal of elucidating the hierarchy of metabolic and regulatory systems in the cell, and ultimately leading to a predictive understanding of the cellular response to perturbations (Ideker *et al.*, 2001; Kitano, 2002) (Fig. 17.2). As the theoretical and experimental tools of systems biology rapidly advance, multiple fields are embracing systems biology approaches as a mainstream method of research. Because postgenomics research is taking place throughout the tree of life, comparative approaches are a way to combine data from many organisms to understand the evolution and function of biological systems from the gene to the organismal level. Therefore, systems biology can build on decades of theoretical work in evolutionary biology, and at the same time evolutionary biology can use systems approaches to go in new uncharted directions. For instance, although comparative genomics has benefited from a long tradition of theoretical work by molecular evolutionists (Wolfe and Li, 2003), new datasets being provided by systems biology are allow-

ing theoreticians new ways to study evolutionary processes (Barabasi and Oltvai, 2004).

Comparative studies can give insight into even the highest-level principles of life. For example, revolutionary findings in network theory have in part come from genomic data from a wide range of organisms, leading scientists to propose laws that seem to govern biological networks (Jeong *et al.*, 2000; Ravasz *et al.*, 2002). Different types of cellular networks (e.g., protein interaction and metabolic networks) seem to share properties with other complex abiotic networks such as their ''scale-free'' nature and ''small world'' organization. In scale-free networks, a few nodes (hubs) have the largest number of connections to other nodes, whereas most of the nodes have just a few connections. This property is reflected in a power-law distribution. In practical terms, this relationship means that, in a protein interaction network, most proteins interact with a couple of others whereas a few proteins (hubs) interact with a large number, and that, in a metabolic network, a few molecules (hubs) participate in most reactions whereas the rest participate in one or two. The ''small world'' concept refers to the property of such spoke-and-hub networks that there is a small path length between nodes, just as in modern air travel where only a few flights connect any two cities in the world. This property means that a path of just a few interactions or reactions will connect almost any pair of molecules in the cell (Barabasi and Oltvai, 2004).

Additional levels in the hierarchy of biological networks and the interactions between them are now being characterized that will allow for integration of data and new theoretical predictions (Ge *et al.*, 2003). Processes widely studied by evolutionary biologists such as selection, gene duplication, and neutral evolution are being examined in the context of network models as opposed to at the level of individual genes or molecules (Hahn *et al.*, 2004; van Noort *et al.*, 2004; Wagner, 2003a,b; Wuchty, 2004).

EVOLUTION OF BIOLOGICAL NETWORKS

Transcriptional Networks

High-throughput global gene expression approaches such as EST sequencing and microarrays are now common practice for functional assessment of the genome. The extensive microarray gene expression datasets available for model and non-model organisms are starting to be incorporated into a comparative approach to study transcriptome evolution at multiple levels of divergence. At lower levels of divergence, studies in organisms including fish (Olesiak *et al.*, 2002), fruitfly (Meiklejohn *et al.*, 2003; Ranz *et al.*, 2003), and yeast (Townsend *et al.*, 2003) have now

shown that extensive variation exists in the transcriptome in natural populations and that this variation is likely to be an important factor in organismal evolution. Transcriptome comparisons across several primate and mouse species, however, suggest that the majority of gene expression differences within and between species evolve in a selectively neutral or nearly neutral fashion (Khaitovich *et al.*, 2004). At intermediate levels of divergence, less information is available at present due to lack of genomic data. Although analytically challenging, the use of gene expression profiling by heterologous hybridization to a single species cDNA microarray is starting to be explored, potentially opening the door to comparative analyses of taxa as divergent as 200 mega-annum (Ma) (Renn *et al.*, 2004). This application would be of great significance for the comparative study of non-model organisms that are only distantly related to an already sequenced species. At deep levels of divergence, coexpression of large aggregates of functionally related genes seems to be conserved across evolution. Two recent comparisons of the transcriptomes of several of the model organisms [*S. cerevisiae*, *D. melanogaster*, *C. elegans*, and *H. sapiens* in one case (Stuart *et al.*, 2003), and these four plus *A. thaliana* and *E. coli* in the second case (Bergmann *et al.*, 2004)] support the hypothesis that coexpression networks can be split into multiple components enriched for genes involved in similar functional processes. Some of these identified components can be unique to a certain clade, such as the signaling pathway and neuronal function components present only in metazoans in the four-species comparison (Stuart *et al.*, 2003). These cross-species comparisons promise to provide more information about coexpression network evolution as the transcriptomes of additional diverse lineages becomes available (Zhou and Gibson, 2004).

Central to postgenomic analysis is the accuracy of genome annotation. The degree of accuracy in which genomes are annotated is affected by the quality of sequence assembly, gene prediction, and functional annotation by both bioinformatics and experimental data. This relationship is particularly critical in genome projects of non-model organisms where little genetic work has been performed in the past.All these factors, combined with the lack of network information outside the model organisms, point to the tradeoff between a comprehensive systems analysis of a particular network within a well-studied organism, versus the historical perspective introduced by evolutionary conservation or divergence of systems through time in phylogenetic comparisons. Therefore, although only partial inference is possible at present, studies have already shown that the comparative approach to coexpression not only is giving insight into the universal rules that govern biological systems but also has practical implications by helping improve functional annotations of both model and non-model organisms (Bergmann *et al.*, 2004; Stuart *et al.*, 2003). Be-

cause comparative analyses of coexpression data from several model organisms have shown high levels of conservation between such divergent taxa as prokaryotes (*E. coli* and *B. subtilis*) (Snel *et al.*, 2004), opisthokont eukaryotes (Stuart *et al.*, 2003), and even prokaryotes and eukaryotes (Bergmann *et al.*, 2004), some efforts are now targeting the coupled evolution of regulatory networks and the transcriptome.

Regulatory Networks

The characterization of the transcriptome is only a fraction of the information needed to understand global cellular processes because gene expression is driven by the spatio-temporal localization of regulatory networks and details of specific protein–DNA and protein–protein interactions. Genomewide efforts to characterize transcriptional regulatory networks have already been fruitful in model organisms like yeast (Lee *et al.*, 2002) and *E. coli* (Shen-Orr *et al.*, 2002). In multicellular organisms, fractions of the regulatory networks are being characterized for sea urchins (Davidson, 2001), *Drosophila* (Berman *et al.*, 2004), and mammals (ENCODE Project Consortium, 2004).

Transcription factors are regulatory proteins that influence the expression of specific genes. They work by binding to cis-regulatory elements (short and often degenerate sequence motifs frequently located upstream of genes) where they interact with the transcription apparatus to either enhance or repress gene expression. Even though identifying cis-regulatory elements in new genomes is an inherently difficult task due to their short sequence length and as yet unknown syntax, comparative approaches have been helpful. By aligning orthologous regions flanking a gene from multiple species, conserved noncoding sequence motifs can be distinguished. These evolutionary conserved motifs are then hypothesized to be potential functional elements. This method, called phylogenetic footprinting (Tagle *et al.*, 1988), has successfully been used to identify a limited number of regulatory regions in vertebrates (Dermitzakis *et al.*, 2003; Gumucio *et al.*, 1992) and plants (Hong *et al.*, 2003; Kaplinsky *et al.*, 2002). More sophisticated comparative approaches are starting to combine computational prediction and laboratory validation of regulatory networks. Coexpression data and known cisregulatory elements from *S. cerevisiae* were used in a multispecies comparison of 13 published ascomycete genomes, finding multiple cases of regulatory conservation but also some cases of regulatory diversification (Gasch *et al.*, 2004). It has become apparent, however, that sequence conservation alone will not help identify all cis-regulatory elements by phylogenetic footprinting, and additional data and experimental approaches have to be integrated (Richards *et al.*, 2005).

Gene expression can be regulated not only at transcriptional initiation but also at other levels, such as during mRNA editing, transport, or translation, and characterizing these interactions and their evolution is one of the many future challenges of systems biology (Wei *et al.*, 2004). For example, comparative work on populations of yeast and fruitfly has recently shown that protein–protein interactions are negatively associated with evolutionary variation in gene expression (Lemos *et al.*, 2004). A comparative analysis of the *E. coli* and yeast regulatory networks has demonstrated that gene duplication has a key role in network evolution both in eukaryotes and prokaryotes (Teichmann and Babu, 2004). Finally, introducing concepts of network dynamics has revealed new topological changes in the regulatory network in yeast (Luscombe *et al.*, 2004), an approach that, incorporated into a comparative framework, will eventually provide answers to the evolution of morphological divergence in multicellular taxa (Howard and Davidson, 2004).

Protein Networks

The proteome for several of the model organisms is now characterized, and this global scale information has been used to predict protein–protein interaction networks (interactomes) for *D. melanogaster* (Giot *et al.*, 2003), *C. elegans* (Li *et al.*, 2004), and *S. cerevisiae* (Uetz *et al.*, 2000). Assuming some degree of evolutionary conservation, these data can also be used to transfer interactome annotations to genomes that have not been characterized experimentally. Comparisons across multiple species have shown conserved protein interactions that allow for initial drafts of protein–protein interaction maps of human (Lehner and Fraser, 2004) and *A. thaliana* (Yu *et al.*, 2004). When formulating evolutionary hypotheses, however, attention to the phylogenetic relationships is necessary. For example, some of the conclusions from the analysis of the *C. elegans* interactome (Luscombe *et al.*, 2004) are weakened by the incorrect assumption that plants (*A. thaliana*) and animals (*C. elegans*, *D. melanogaster*) are more closely related to each other than to yeast. Current phylogenies show that multicellularity has occurred independently in metazoans, fungi, and plants (Fig. 17.1*A*), and that unicellularity in yeasts is a derived rather than ancestral state (Fig. 17.1*C*).

Metabolic Networks and "Ome" Data Integration

The metabolome is made up of all of the low-molecular weight molecules (metabolites) present in a cell at a particular time point, and their levels can be regarded as the functional response of biological systems to genetic or environmental stimuli (Fiehn, 2002). Challenges faced in the

global study of metabolites, such as their dynamic behavior and chemistry, are being addressed by emerging technologies such as liquid and gas chromatography mass spectrometry and NMR (Stitt and Fernie, 2003). Plant biologists have led in the application of these advances (Oksman-Caldentey *et al.*, 2004), and soon there will likely be large datasets for multiple plant and other eukaryotic species. Although high-throughput metabolome projects are just now being initiated, comparative analysis of 43 known metabolic networks has already shown that they seem to follow a power-law distribution (Barabasi and Oltvai, 2004; Jeong *et al.*, 2000).

The integration of data from the different levels of cellular networks (transcriptome, regulome, interactome, and metabolome) is the next obvious step to identify patterns of network interactions in individual species and in multispecies comparisons (Castrillo and Oliver, 2004; Ge *et al.*, 2003; Papin *et al.*, 2004). This integrative approach has already been fruitful in model organisms such as *C. elegans* (Walhout *et al.*, 2002) and *S. cerevisiae* (Ge *et al.*, 2001).

It is clear that producing a large scale comparative systems biology analysis will have to involve the work of many research groups and that many challenges will need to be overcome. For example, rigorous standards will need to be established to facilitate the comparison of results from high-throughput ''omic'' analyses before we can make conclusive evolutionary inferences (Levesque and Benfey, 2004). A pioneer example is the ENCODE initiative, which aims to identify all functional elements in the human genome by using coordinated computational and experimental efforts in a multispecies framework (ENCODE Project Consortium, 2004).

Although we can already find global patterns of network evolution, in the future we should be able to look at trends and patterns in the evolution of biological systems within phylogenies. For instance, we should be able to look at how many of the biological network similarities are due to homoplasy as opposed to phylogenetic constraints due to common ancestry. Thus, by using the theoretical framework developed for the comparative method, phylogenetic information can allow only for improvement of evolutionary inference at the systems level. Finally, to bring evolutionary systems biology to the highest level of biological organization, ecosystem-level factors have to be taken into consideration. To this end, the use of highthroughput approaches for the study of interactions among organisms and between organisms and their natural environments is engaging the interest of ecologists (Benfey, 2004; Thomas and Klaper, 2004).

HISTORICAL PERSPECTIVE

Darwin's theory of natural selection and, later on, the integrative nature of the modern synthesis consolidated the study of evolution as a solid discipline to address fundamental questions in biology. The scientific advances that allowed for the discovery of the structure of DNA and the development of molecular biology eventually led to large-scale whole-genome initiatives. This unfolding was a revolutionary moment in the scientific mentality of 20th century researchers, because it generated the integrative approaches of systems biology, which will most likely become the standard of 21st century biology. Organismal biologists have been thinking along these lines for the past few decades, advocating integrative and multidisciplinary approaches to evolutionary questions (Wake, 2003). Thus, bridging knowledge between evolutionary theory and systems biology will be only a natural process. Together, these approaches offer the promise to solve two of the ultimate questions in biology: the function of biological systems and an understanding of the evolution of life's diversity.

ACKNOWLEDGMENTS

I thank Francisco Ayala, Jody Hey, and Walter Fitch for the invitation to participate in the Mayr colloquium. Pilar Francino and Paramvir Dehal provided thoughtful insight for both the seminar and the manuscript. I also thank Mike Colvin, Benoît Dayrat, Jodi Schwarz, Rick Baker, Kevin Helfenbein, and an anonymous reviewer for helpful comments on previous versions of the manuscript. Benoît Dayrat and Peter Brokstein helped with figure design. I thank Caturro Mejía for introducing me to Mayr's writings many years ago. This work was performed under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research, University of California, Lawrence Berkeley National Laboratory under contract DE-AC03-76SF00098. I also acknowledge support by National Science Foundation Grant OCE 0313708.

REFERENCES

Adoutte, A., Balavoine, G., Lartillot, N. & de Rosa, R. (1999) Animal evolution. The end of intermediate taxa? *Trends Genet*. **15,** 104–108.

Amaral Zettler, L. A., Nerad, T. A., O'Kelly, C. J. & Sogin, M. L. (2001) The nucleariid amoebae: More protists at the animal-fungal boundary. *J. Eukaryotic Microbiol.* **48,** 293–297.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408,** 796–815.

Archibald, J. M. & Keeling, P. J. (2002) Recycled plastids: a "green movement" in eukaryotic evolution. *Trends Genet*. **18,** 577–584.

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S., Allen, A. E., Apt, K. E., Bechner, M., *et al*. (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306,** 79–86.

Barabasi, A. L. & Oltvai, Z. N. (2004) Network biology: Understanding the cell's functional organization. *Nat*. *Rev*. *Genet*. **5,** 101–113.

Benfey, P. N. (2004) Development and ecology in the time of systems biology. *Dev*. *Cell* **7,** 329–330.

Berbee, M. L. & Taylor, J. W. (1993) Dating the evolutionary radiations of the true fungi. *Can*. *J*. *Bot*. **71,** 1114–1127.

Bergmann, S., Ihmels, J. & Barkai, N. (2004) Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol*. **2,** E9.

Berman, B. P., Pfeiffer, B. D., Laverty, T. R., Salzberg, S. L., Rubin, G. M., Eisen, M. B. & Celniker, S. E. (2004) Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Genome Biol*. **5,** R61.

Bhattacharya, D., Yoon, H. S. & Hackett, J. D. (2004) Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *BioEssays* **26,** 50–60.

Castrillo, J. I. & Oliver, S. G. (2004) Yeast as a touchstone in post-genomic research: strategies for integrative analysis in functional genomics. *J*. *Biochem*. *Mol*. *Biol*. **37,** 93–106.

Davidson, E. H. (2001) *Genomic Regulatory Systems: Development and Evolution* (Academic, San Diego).

Dermitzakis, E. T., Reymond, A., Scamuffa, N., Ucla, C., Kirkness, E., Rossier, C. & Antonarakis, S. E. (2003) Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). *Science* **302,** 1033–1035.

Dietrich, F. S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pohlmann, R., Luedi, P., Choi, S., *et al*. (2004) The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304,** 304–307.

Dopazo, H., Santoyo, J. & Dopazo, J. (2004) Phylogenomics and the number of characters required for obtaining an accurate phylogeny of eukaryote model species. *Bioinformatics* **20,** Suppl. 1, I116–I121.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuveglise, C., Talla, E., *et al*. (2004) Genome evolution in yeasts. *Nature* **430,** 35–44.

ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306,** 636–640.

Fiehn, O. (2002) Metabolomics—the link between genotypes and phenotypes. *Plant Mol*. *Biol*. **48,** 155–171.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., *et al*. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419,** 498–511.

Gasch, A. P., Moses, A. M., Chiang, D. Y., Fraser, H. B., Berardini, M. & Eisen, M. B. (2004) Conservation and evolution of cis-regulatory systems in ascomycete fungi. *PLoS Biol*. **2,** e398.

Ge, H., Liu, Z., Church, G. M. & Vidal, M. (2001) Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat*. *Genet*. **29,** 482–486.

Ge, H., Walhout, A. J. & Vidal, M. (2003) Integrating "omic" information: a bridge between genomics and systems biology. *Trends Genet*. **19,** 551–560.

Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E., *et al*. (2003) A protein interaction map of *Drosophila melanogaster*. *Science* **302,** 1727–1736.

Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., *et al*. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296,** 92–100.

Gumucio, D. L., Heilstedt-Williamson, H., Gray, T. A., Tarle, S. A., Shelton, D. A., Tagle, D. A., Slightom, J. L., Goodman, M. & Collins, F. S. (1992) Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human gamma and epsilon globin genes. *Mol. Cell. Biol.* **12,** 4919–4929.

Hahn, M. W., Conant, G. C. & Wagner, A. (2004) Molecular evolution in large genetic networks: Does connectivity equal constraint? *J. Mol. Evol.* **58,** 203–211.

Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. (1999) From molecular to modular cell biology. *Nature* **402,** C47–C52.

Hedges, S. B. (2002) The origin and evolution of model organisms. *Nat. Rev. Genet.* **3,** 838–849.

Hong, R. L., Hamaguchi, L., Busch, M. A. & Weigel, D. (2003) Regulatory elements of the floral homeotic gene AGAMOUS identified by phylogenetic footprinting and shadowing. *Plant Cell* **15,** 1296–1309.

Hood, L. (2003) Systems biology: Integrating technology, biology, and computation. *Mech. Ageing Dev.* **124,** 9–16.

Howard, M. L. & Davidson, E. H. (2004) cis-Regulatory control circuits in development. *Dev. Biol.* **271,** 109–118.

Ideker, T., Galitski, T. & Hood, L. (2001) A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2,** 343–372.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. (2000) The large-scale organization of metabolic networks. *Nature* **407,** 651–654.

Kaplinsky, N. J., Braun, D. M., Penterman, J., Goff, S. A. & Freeling, M. (2002) Utility and distribution of conserved noncoding sequences in the grasses. *Proc. Natl. Acad. Sci. USA* **99,** 6147–6151.

Kellis, M., Birren, B. W. & Lander, E. S. (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428,** 617–624.

Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansorge, W. & Paabo, S. (2004) A neutral model of transcriptome evolution. *PLoS Biol.* **2,** E132.

King, N. (2004) The unicellular ancestry of animal development. *Dev. Cell* **7,** 313–325.

Kitano, H. (2002) Systems biology: A brief overview. *Science* **295,** 1662–1664.

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., *et al*. (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298,** 799–804.

Lehner, B. & Fraser, A. G. (2004) A first-draft human protein-interaction map. *Genome Biol.* **5,** R63.

Lemos, B., Meiklejohn, C. D. & Hartl, D. L. (2004) Regulatory evolution across the protein interaction network. *Nat. Genet.* **36,** 1059–1060.

Levesque, M. P. & Benfey, P. N. (2004) Systems biology. *Curr. Biol.* **14,** R179–80.

Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T., *et al*. (2004) A map of the interactome network of the metazoan *C. elegans*. *Science* **303,** 540–543.

Luscombe, N. M., Babu, M. M., Yu, H., Snyder, M., Teichmann, S. A. & Gerstein, M. (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431,** 308–312.

Martinez, D., Larrondo, L. F., Putnam, N., Gelpke, M. D., Huang, K., Chapman, J., Helfenbein, K. G., Ramaiya, P., Detter, J. C., Larimer, F., *et al*. (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat. Biotechnol.* **22,** 695–700.

Medina, M., Collins, A. G., Silberman, J. D. & Sogin, M. L. (2001) Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. *Proc. Natl. Acad. Sci. USA* **98,** 9707–9712.

Medina, M., Collins, A. G., Taylor, J. W., Valentine, J. W., Lipps, J. H., Amaral Zettler, L. A. & Sogin, M. L. (2003) Phylogeny of Opisthokonta and the evolution of multicellularity and complexity in Fungi and Metazoa. *Int. J. Astrobiol.* **2,** 203–211.

Meiklejohn, C. D., Parsch, J., Ranz, J. M. & Hartl, D. L. (2003) Rapid evolution of male-biased gene expression in Drosophila. *Proc. Natl. Acad. Sci. USA* **100,** 9894–9899.

Oksman-Caldentey, K. M., Inze, D. & Oresic, M. (2004) Connecting genes to metabolites by a systems biology approach. *Proc. Natl. Acad. Sci. USA* **101,** 9949–9950.

Olesiak, M. J., Churcill, G. A. & Crawford, D. L. (2002) Variation in gene expression within and among natural populations. *Nat. Genet.* **32,** 261–266.

Papin, J. A., Reed, J. L. & Palsson, B. O. (2004) Hierarchical thinking in network biology: the unbiased modularization of biochemical networks. *Trends Biochem. Sci.* **29,** 641–647.

Pryer, K. M., Schneider, H., Zimmer, E. A. & Ann Banks, J. (2002) Deciding among green plants for whole genome studies. *Trends Plant Sci.* **7,** 550–554.

Ranz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D. & Hartl, D. L. (2003) Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300,** 1742–1745.

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabasi, A. L. (2002) Hierarchical organization of modularity in metabolic networks. *Science* **297,** 1551–1555.

Renn, S. C., Aubin-Horth, N. & Hofmann, H. A. (2004) Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. *BMC Genomics* **5,** 42.

Richards, S., Liu, Y., Bettencourt, B. R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M. J., Chen, R., Meisel, R. P., *et al.* (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res.* **15,** 1–18.

Ruiz-Trillo, I., Inagaki, Y., Davis, L. A., Sperstad, S., Landfald, B. & Roger, A. J. (2004) *Capsaspora owczarzaki* is an independent opisthokont lineage. *Curr. Biol.* **14,** R946–R947.

Ruiz-Trillo, I., Paps, J., Loukota, M., Ribera, C., Jondelius, U., Baguña, J. & Ruitort, M. (2002) A phylogenetic analysis of myosin heavy chain type II sequences corroborates that Acoela and Nemertodermatida are basal bilaterians. *Proc. Natl. Acad. Sci. USA* **99,** 11246–11251.

Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31,** 64–68.

Simpson, A. G. & Roger, A. J. (2004) The real "kingdoms" of eukaryotes. *Curr. Biol.* **14,** R693–R696.

Snel, B., van Noort, V. & Huynen, M. A. (2004) Gene co-regulation is highly conserved in the evolution of eukaryotes and prokaryotes. *Nucleic Acids Res.* **32,** 4725–4731.

Stechmann, A. & Cavalier-Smith, T. (2002) Rooting the eukaryote tree by using a derived gene fusion. *Science* **297,** 89–91.

Stechmann, A. & Cavalier-Smith, T. (2003) The root of the eukaryote tree pinpointed. *Curr. Biol.* **13,** R665–R666.

Stitt, M. & Fernie, A. R. (2003) From measurements of metabolites to metabolomics: an "on the fly" perspective illustrated by recent studies of carbon-nitrogen interactions. *Curr. Opin. Biotechnol.* **14,** 136–144.

Stuart, J. M., Segal, E., Koller, D. & Kim, S. K. (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302,** 249–255.

Tagle, D. A., Koop, B. F., Goodman, M., Slightom, J. L., Hess, D. L. & Jones, R. T. (1988) Embryonic epsilon and gamma globin genes of a prosimian primate (*Galago crassicaudatus*). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J. Mol. Biol.* **203,** 439–455.

Teichmann, S. A. & Babu, M. M. (2004) Gene regulatory network growth by duplication. *Nat. Genet.* **36,** 492–496.

Thomas, M. A. & Klaper, R. (2004) Genomics for the ecological toolbox. *Trends Ecol. Evol.* **19,** 439–445.

Townsend, J. P., Cavalieri, D. & Hartl, D. L. (2003) Population genetic variation in genome-wide gene expression. *Mol. Biol. Evol.* **20,** 955–963.

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., *et al.* (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403,** 623–627.

van Noort, V., Snel, B. & Huynen, M. A. (2004) The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model. *EMBO Rep.* **5,** 280–284.

Wagner, A. (2003b) How the global structure of protein interaction networks evolves. *Proc. R. Soc. London Ser. B Biol. Sci.* **270,** 457–466.

Wagner, A. (September 30, 2003a) Does selection mold molecular networks? *Sci. STKE*, 10.1126/stke.2003.202.pe41.

Wake, M. L. (2003) What is "Integrative Biology"? *Integr. Comp. Biol.* **43,** 239–241.

Walhout, A. J., Reboul, J., Shtanko, O., Bertin, N., Vaglio, P., Ge, H., Lee, H., Doucette-Stamm, L., Gunsalus, K. C., Schetter, A. J., *et al.* (2002) Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. *Curr. Biol.* **12,** 1952–1958.

Waugh, M., Hraber, P., Weller, J., Wu, Y., Chen, G., Inman, J., Kiphart, D. & Sobral, B. (2000) The *Phytophthora* genome initiative database: informatics and analysis for distributed pathogenomic research. *Nucleic Acids Res.* **28,** 87–90.

Wei, G. H., Liu, D. P. & Liang, C. C. (2004) Charting gene regulatory networks: strategies, challenges and perspectives. *Biochem. J.* **381,** 1–12.

Wolf, Y. I., Rogozin, I. B. & Koonin, E. V. (2004) Coelomata and not Ecdysozoa: evidence from genome-wide phylogenetic analysis. *Genome Res.* **14,** 29–36.

Wolfe, K. H. & Li, W. H. (2003) Molecular evolution meets the genomics revolution. *Nat. Genet.* **33,** Suppl., 255–265.

Wuchty, S. (2004) Evolution and topology in the yeast protein interaction network. *Genome Res.* **14,** 1310–1314.

Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296,** 79–92.

Yu, H., Luscombe, N. M., Lu, H. X., Zhu, X., Xia, Y., Han, J. D., Bertin, N., Chung, S., Vidal, M. & Gerstein, M. (2004) Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res.* **14,** 1107–1118.

Zhou, X. J. & Gibson, G. (2004) Cross-species comparison of genome-wide expression patterns. *Genome Biol.* **5,** 232.

Index

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z