In the Light of Evolution: Volume X: Comparative Phylogeography

DETAILS

432 pages | 6 x 9 | HARDBACK
ISBN 978-0-309-44422-4 | DOI: 10.17226/23542

CONTRIBUTORS

John C. Avise and Francisco J. Ayala, Editors; National Academy of Sciences

In the Light of Evolution

Volume X: Comparative Phylogeography

In the Light of Evolution

Volume X: Comparative Phylogeography

JOHN C. AVISE and FRANCISCO J. AYALA, *Editors*

NATIONAL ACADEMY OF SCIENCES

THE NATIONAL ACADEMIES PRESS
Washington, DC
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

This volume is based on the Arthur M. Sackler Colloquium of the National Academy of Sciences, "In the Light of Evolution X: Comparative Phylogeography," held January 8–9, 2016, at the Arnold and Mabel Beckman Center of the National Academies of Sciences, Engineering, and Medicine in Irvine, California. It is the tenth (and final) in a series of annual colloquia under the umbrella title "In the Light of Evolution."

Cataloging-in-Publication Data

In the light of evolution / John C. Avise and Francisco J. Ayala, editors.
p. cm.
Vol. I based on a colloquium of the National Academy of Sciences, held
December 1–2, 2006, in Irvine, California.
Includes bibliographical references.
ISBN-13: 978-0-309-44422-4
ISBN-10: 0-309-44422-5
1. Evolution (Biology)—Congresses. I. Avise, John C, 1948–. II . Ayala, Francisco
José, 1934–III . National Academy of Sciences (U.S.)
QH359.I55 2007
576.8—dc22
2007032455

Digital Object Identifier: 10.17226/23542

Additional copies of this publication are available for sale from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

Cover image: Space (geography) and time (genealogy) are the two axes that constitute the framework of comparative phylogeography. Albert Einstein demonstrated that space and time are connected in astrophysics; phylogeographers are showing that space and time often display relatable patterns in biology also. Drawing by John C. Avise.

Suggested citation: Avise, John C., and Francisco J. Ayala, Eds. 2017. *In the Light of Evolution, Volume X: Comparative Phylogeography*. Washington, DC: The National Academies Press. doi: 10.17226/23542.

Printed in the United States of America

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. C. D. Mote, Jr., is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.national-academies.org**.

Arthur M. Sackler, M.D.
1913–1987

Born in Brooklyn, New York, Arthur M. Sackler was educated in the arts, sciences, and humanities at New York University. These interests remained the focus of his life, as he became widely known as a scientist, art collector, and philanthropist, endowing institutions of learning and culture throughout the world.

He felt that his fundamental role was as a doctor, a vocation he decided upon at the age of four. After completing his internship and service as house physician at Lincoln Hospital in New York City, he became a resident in psychiatry at Creedmoor State Hospital. There, in the 1940s, he started research that resulted in more than 150 papers in neuroendocrinology, psychiatry, and experimental medicine. He considered his scientific research in the metabolic basis of schizophrenia his most significant contribution to science and served as editor of the *Journal of Clinical and Experimental Psychobiology* from 1950 to 1962. In 1960 he started publication of *Medical Tribune*, a weekly medical newspaper that reached over one million readers in 20 countries. He established the Laboratories for Therapeutic Research in 1938, a facility in New York for basic research that he directed until 1983.

As a generous benefactor to the causes of medicine and basic science, Arthur Sackler built and contributed to a wide range of scientific institutions: the Sackler School of Medicine established in 1972 at Tel Aviv University, Tel Aviv, Israel; the Sackler Institute of Graduate Biomedical Science at New York University, founded in 1980; the Arthur M. Sackler Science Center dedicated in 1985 at Clark University, Worcester, Massachusetts; and the Sackler School of Graduate Biomedical Sciences, established in 1980, and the Arthur M. Sackler Center for Health Communications, established in 1986, both at Tufts University, Boston, Massachusetts.

His pre-eminence in the art world is already legendary. According to his wife Jillian, one of his favorite relaxations was to visit museums and art galleries and pick out great pieces others had overlooked. His interest in art is reflected in his philanthropy; he endowed galleries at the Metropolitan Museum of Art and Princeton University, a museum at Harvard University, and the Arthur M. Sackler Gallery of Asian Art in

Washington, DC. True to his oft-stated determination to create bridges between peoples, he offered to build a teaching museum in China, which Jillian made possible after his death, and in 1993 opened the Arthur M. Sackler Museum of Art and Archaeology at Peking University in Beijing.

In a world that often sees science and art as two separate cultures, Arthur Sackler saw them as inextricably related. In a speech given at the State University of New York at Stony Brook, *Some reflections on the arts, sciences and humanities*, a year before his death, he observed: ‘‘Communication is, for me, the *primum movens* of all culture. In the arts . . . I find the emotional component most moving. In science, it is the intellectual content. Both are deeply interlinked in the humanities.’’ The Arthur M. Sackler Colloquia at the National Academy of Sciences pay tribute to this faith in communication as the prime mover of knowledge and culture.

Contents

Preface to the *In the Light of Evolution* Series

Biodiversity—the genetic variety of life—is an exuberant product of the evolutionary past, a vast human-supportive resource (aesthetic, intellectual, and material) of the present, and a rich legacy to cherish and preserve for the future. Two urgent challenges, and opportunities, for 21st-century science are to gain deeper insights into the evolutionary processes that foster biotic diversity, and to translate that understanding into workable solutions for the regional and global crises that biodiversity currently faces. A grasp of evolutionary principles and processes is important in other societal arenas as well, such as education, medicine, sociology, and other applied fields including agriculture, pharmacology, and biotechnology. The ramifications of evolutionary thought also extend into learned realms traditionally reserved for philosophy and religion.

In 1973, Theodosius Dobzhansky penned a short commentary entitled "Nothing in biology makes sense except in the light of evolution." Most scientists agree that evolution provides the unifying framework for interpreting biological phenomena that otherwise can often seem unrelated and perhaps unintelligible. Given the central position of evolutionary thought in biology, it is sadly ironic that evolutionary perspectives outside the sciences have often been neglected, misunderstood, or purposefully misrepresented.

The central goal of the *In the Light of Evolution* (*ILE*) series is to promote the evolutionary sciences through state-of-the-art colloquia—in the series of Arthur M. Sackler colloquia sponsored by the National Academy of Sciences—and their published proceedings. Each installment explores evolutionary perspectives on a particular biological topic that is scientifically intriguing but also has special relevance to contemporary societal issues or challenges. Individually and collectively, the *ILE* series aims to interpret phenomena in various areas of biology through the lens of evolution, address some of the most intellectually engaging as well as pragmatically important societal issues of our times, and foster a greater appreciation of evolutionary biology as a consolidating foundation for the life sciences.

The organizers and founding editors of this effort (Avise and Ayala) are the academic grandson and son, respectively, of Theodosius Dobzhansky, to whose fond memory this *ILE* series is dedicated. May Dobzhansky's words and insights continue to inspire rational scientific inquiry into nature's marvelous operations.

John C. Avise and Francisco J. Ayala
Department of Ecology and Evolutionary Biology,
University of California, Irvine (January 2007)

Preface to *In the Light of Evolution, Volume X: Comparative Phylogeography*

Phylogeography is the study of the spatial arrangement of genealogical lineages, especially within and among conspecific populations and closely related species (Avise, 2000). Ever since its inception in the late 1970s (Avise et al., 1979a,b) and mid-1980s (Avise et al., 1987), the field has sought to extend phylogenetic reasoning to the intraspecific level, and thereby build empirical and conceptual bridges between the formerly separate disciplines of microevolutionary population genetics and macroevolutionary phylogenetics. In the early years, phylogeographers relied on data from restriction-site surveys of mitochondrial (mt) DNA to draw inferences about population structure and historical demography, but stunning improvements in molecular techniques (Emerson et al., 2010; Rocha et al., 2013) and extensions of coalescent theory and other analytical methods (Knowles, 2009b) later broadened the field's scope dramatically (Hickerson et al., 2010). Phylogeographic perspectives have transformed aspects of population biology, biogeography, systematics, ecology, genetics, and biodiversity conservation. One aim of this colloquium was to bring together leading scientists to address the current state of phylogeography as the discipline enters its fourth decade. The broader goal was to update a wide audience on recent developments in phylogeographic research and their relevance to past accomplishments and future research directions.

Many of the advancements in phylogeography have entailed comparative appraisals of one sort or another (Bermingham and Moritz, 1998). Chapters in Part I of this volume emphasize the word "comparative" in a spatial sense, where the phylogeographic assessments entail various species (sometimes generally codistributed) that may be representative of particular kinds of environmental settings, such as oceanic versus continental realms. Part II deals with comparative phylogeography in a genomic sense. Nonrecombining cytoplasmic genomes have been the standard workhorses of genealogical analyses, but in principle the primary library of evolutionary histories is ensconced in nuclear genomes that are increasingly accessible to scrutiny. Part III focuses on comparative phylogeography in a taxonomic sense, emphasizing how phylogeographic findings have impacted ecological and evolutionary thought in a wide diversity of organismal groups. Part IV focuses on comparative phylogeography in a conceptual sense by addressing the place of phylogeography in relation to various allied disciplines in the biodiversity sciences. Several chapters in this book inevitably overlap in their sectional assignments because, for example, disparate taxa tend to inhabit different environments, and because the researchers assembled here had a diversity of phylogeographic objectives, data types, and analytical approaches.

This is the tenth and final edition of the "In the Light of Evolution" series. Previous colloquia in this series dealt with the following topics: I, *Adaptation and Complex Design* (Avise and Ayala, 2007); II, *Biodiversity and Extinction* (Avise et al., 2008); III, *Two Centuries of Darwin* (Avise and Ayala, 2009); IV, *The Human Condition* (Avise and Ayala, 2010); V, *Cooperation and Conflict* (Strassmann et al., 2011); VI, *Brain and Behavior* (Striedter et al., 2012); VII, *The Human Mental Machinery* (Cela-Conde et al., 2013); VIII, *Darwinian Thinking in the Social Sciences* (Skyrms et al., 2014); and IX, *Clonal Reproduction: Alternatives to Sex* (Tibayrenc et al., 2015). We hope that this "In the Light of Evolution" series has indeed enlightened a broad audience about the many manifestations and marvelous biological products of our planet's diverse evolutionary processes.

Part I

COMPARATIVE PHYLOGEOGRAPHY IN A SPATIAL SENSE

Phylogeographers have roamed the planet in their quest to identify patterns that might typify organisms occupying diverse environmental regimes. Understanding the biotic response to past challenges provides a compass for contemporary and future challenges. Brian Bowen and colleagues (Chapter 1) summarize a vast literature on phylogeographic findings for the world's oceans. At face value, the oceanic realm might seem to be featureless and relatively free of impediments to gene flow. Nevertheless, many migrant species in the sea (such as marine turtles, and some cetaceans and pelagic fishes) display phylogeographic patterns implicating physical barriers as promoters of both intraspecific genetic breaks and subsequent allopatric speciation events (properly viewed as stages along a temporal continuum). For marine taxa more generally, a wide variety of phylogeographic outcomes implies the operation of other evolutionary processes as well, such as sympatric or ecological speciation in some cases, environmental effects from historically shifting climates, and the important role of organismal behavior and natural history in shaping current phylogeographic motifs. Empirical phylogeographic patterns in the sea have also illuminated the evolutionary role of biodiversity hotspots and the historical dispersal routes that enhance global biodiversity.

Brett Riddle (Chapter 2) then summarizes an equally large scientific literature for species that occupy terrestrial or semiterrestrial regimes on continental landmasses. The author identifies phylogeographic hotspots, usually ecological transition zones that are focal points for study. Previously, these hotspots were primarily in the Northern Hemisphere and proximal to well-equipped labs, but success and improved technology have inspired a diaspora to geographic settings around the world. Here, as might generally be expected, range contraction-expansion episodes and dispersal from and into Neogene and Quaternary refugia played major roles in shaping the current-day arrangements of phylogeographic lineages for many continental taxa. The classical vicariance-dispersal paradigm, which generated heated debate and much research in the previous century, is transforming into a more nuanced and multifactorial perspective on why lineages are distributed as we observe them today in extant faunas.

In Chapter 3, Ivan Prates and colleagues use a case-history approach to reveal how historical climatic and habitat changes affected three broadly codistributed lizard species in Amazonia and the Atlantic Forest of the Neotropics. This study finds partial congruence in phylogeographic patterns that can be attributable to shared historical influences plus similarities in life history and environmental needs. However, the authors then go much further by embracing the close link between phylogeographic pattern and historical population demography to deduce that these taxa had species-specific population demographies in their recent past. These demographic parameters (in conjunction with ecological niche models related to climate change) forecast very distinct population trajectories over the next 60 years for these three environmentally sensitive lizard species.

Remote oceanic archipelagos offer especially intriguing evolutionary studios for studying phylogeographic processes on isolated chains of islands that originated (typically in sequential temporal order) via hotspot volcanism. These mid-oceanic islands offer a proverbial tabula rasa in which every species has origins elsewhere, and community composition is stilted by the filter of dispersal ability. Several such oceanic archipelagos are scattered around the world's oceans, in most cases with the islands being displaced unidirectionally by plate tectonic movements, as if on an evolutionary conveyor belt. In Chapter 4, Kerry Shaw and Rosemary Gillespie review comparative phylogeographic findings for diverse taxa on six such oceanic archipelagos. They focus especially on the "progression rule," which predicts that older genealogical lineages should map to older islands within an archipelago, with progressively younger lineages mapping to progressively younger islands within each such island chain. Shaw and Gillespie find support for the progression rule in some cases (notably in the Hawaiian archipelago), but less so on some others. They also highlight the precedence effect, whereby the success of initial colonizers makes it more difficult for subsequent colonizers to establish. Both the general trend and particular exceptions to it are highly informative about dispersal and speciational processes in these distinctive environmental settings. As noted here (and by Riddle in Chapter 2), the simplified evolutionary arena of oceanic islands can reveal phylogeographic patterns that are obscured in a more complex continental setting.

1

Comparative Phylogeography of the Ocean Planet

BRIAN W. BOWEN,*[††] MICHELLE R. GAITHER,[†]
JOSEPH D. DI BATTISTA,[‡] MATTHEW IACCHEI,*[§]
KIMBERLY R. ANDREWS,[‖] W. STEWART GRANT,[#]
ROBERT J. TOONEN,* AND JOHN C. BRIGGS**

Understanding how geography, oceanography, and climate have ultimately shaped marine biodiversity requires aligning the distributions of genetic diversity across multiple taxa. Here, we examine phylogeographic partitions in the sea against a backdrop of biogeographic provinces defined by taxonomy, endemism, and species composition. The taxonomic identities used to define biogeographic provinces are routinely accompanied by diagnostic genetic differences between sister species, indicating interspecific concordance between biogeography and phylogeography. In cases where individual species are distributed across two or more biogeographic provinces, shifts in genotype frequencies often align with biogeographic boundaries, providing intraspecific concordance between biogeography and phylogeography. Here, we provide examples of comparative phylogeography from (i) tropical seas that host the highest marine biodiversity, (ii) temperate seas with high productiv-

---

*Hawai'i Institute of Marine Biology, University of Hawai'i, Kaneohe, HI 96744; [†]School of Biological and Biomedical Sciences, Durham University, Durham DH1 3LE, United Kingdom; [‡]Department of Environment and Agriculture, Curtin University, Perth, WA 6845, Australia; [§]Department of Oceanography, School of Ocean and Earth Science and Technology, University of Hawai'i at Mãnoa, Honolulu, HI 96822; [‖]Department of Fish and Wildlife Sciences, University of Idaho, Moscow, ID 83844; [#]Commercial Fisheries Division, Alaska Department of Fish and Game, Anchorage, AK 99518; and **Department of Fisheries and Wildlife, Oregon State University, Corvallis, OR 97333. [††]To whom correspondence should be addressed. Email: bbowen@hawaii.edu.

ity but volatile coastlines, (iii) migratory marine fauna, and (iv) plankton that are the most abundant eukaryotes on Earth. Tropical and temperate zones both show impacts of glacial cycles, the former primarily through changing sea levels, and the latter through coastal habitat disruption. The general concordance between biogeography and phylogeography indicates that the population-level genetic divergences observed between provinces are a starting point for macroevolutionary divergences between species. However, isolation between provinces does not account for all marine biodiversity; the remainder arises through alternative pathways, such as ecological speciation and parapatric (semi-isolated) divergences within provinces and biodiversity hotspots.

Phylogeography has roots in biogeography, wherein geographic provinces are identified by concordant shifts in species composition. If the partitions defined by taxonomy are regarded as first-order approximations of evolutionary genetic separations, then continuity between biogeography and phylogeography is apparent. Marine biogeography, the study of species' distributions and evolutionary processes in the sea, began in the mid-19th century based on taxonomic distinctions. Dana (1853) divided the surface waters of the world into several temperature zones based on the distributions of corals and crustaceans. Woodward (1851) identified a series of marine provinces based on the distributions of mollusks. Forbes (1859) made three enduring observations: (*i*) each biogeographic province is a center of origin for new species; (*ii*) these new species tend to migrate outward from the center of origin; and (*iii*) provinces, like species, must be traced back to their historical origins to be understood. These three fundamental contributions appeared in the same decade in which Darwin and Wallace (1858) and Darwin (1859) identified geography and natural selection as agents of evolutionary change.

It is remarkable that five essential publications in the 1850s (Woodward, 1851; Dana, 1853; Darwin and Wallace, 1858; Darwin, 1859; Forbes, 1859) set the stage for 150 years of biogeographic research. Subsequent effort was devoted to species descriptions, geographic ranges, and relationships. Evolutionary hypotheses were formulated by examining the morphology and distribution of organisms. However, not until the advent of molecular technologies in the 1970s did biogeography transition through another fundamental change (Avise et al., 1987).

A primary theme emerging from marine biogeography is concordant levels of endemism in very diverse taxa. For example, endemism in Hawai'i is 25 percent for red algae and fishes (Abbott, 1999; Randall, 2007) and 20 percent for mollusks (Kay, 1980). The Caribbean Province has 33 percent endemism for fishes (Floeter et al., 2008), 32 percent for decapod crustaceans (Boschi, 2000), and 37 percent for corals (Veron, 2000). In the Red Sea, endemism is 13 percent for fishes and polychaetes, 8 percent for echinoderms, 17 percent for ascidians, and 5.5 percent for corals (DiBattista et al., 2016b). This concordance across diverse taxonomic groups indicates unifying evolutionary processes.

Here, we demonstrate concordance between biogeographic provinces defined by taxonomy and phylogeographic clusters identified with DNA sequences. At the level of interspecific comparisons, this concordance is obvious; genetic partitions between sister species are expected. However, below this level, at the inception of speciation, it is still unclear how genetic partitions within species (defined by allele-frequency shifts and significant *F*-statistics) translate into species-level divergences (reciprocal monophyly and morphological distinction). Concordance between taxonomy-based biogeography and genetic-based phylogeography would indicate a continuum from population isolation to morphological divergence to evolutionary innovation. In this review, we examine comparative phylogeography, first across biogeographic provinces and second across taxonomic groups with widely divergent life histories.

A second goal is to summarize aspects of comparative phylogeography that illuminate the origins of marine biodiversity. As in terrestrial and freshwater systems, phylogeographic comparisons among species often reveal a diversity of outcomes, attributed to the idiosyncrasies of individual taxa (Toonen et al., 2011; Riddle, Chapter 2, this volume). However, the comparative approach can reveal insights unavailable from any one example (Bermingham and Moritz, 1998), as illustrated by the terrestrial biota of Hawai'i (Shaw and Gillespie, Chapter 4, this volume). Finally, illuminating the origins of new species at biodiversity hotspots and centers of endemism can illustrate conservation priorities for the ocean, the cradle of life on our beleaguered planet.

BIOGEOGRAPHIC PROVINCES

Tropical Oceans

Tropical oceans are characterized by biodiversity hotspots, including the Caribbean and the Coral Triangle (between the Philippines, Indonesia, and New Guinea) (Fig. 1.1*A*) and endemism hotspots, such as Hawai'i and the Red Sea on the periphery of the Indo-Pacific. The evolutionary role of biodiversity hotspots versus endemism hotspots is contentious although biodiversity hotspots are widely recognized as evolutionary incubators producing new species (Briggs, 2003; Bowen et al., 2013).

The Coral Triangle has been a stable reef habitat for tens of millions of years, and this persistence is believed to be key to the production and export of species (Pellissier et al., 2014). Pervasive signals of population

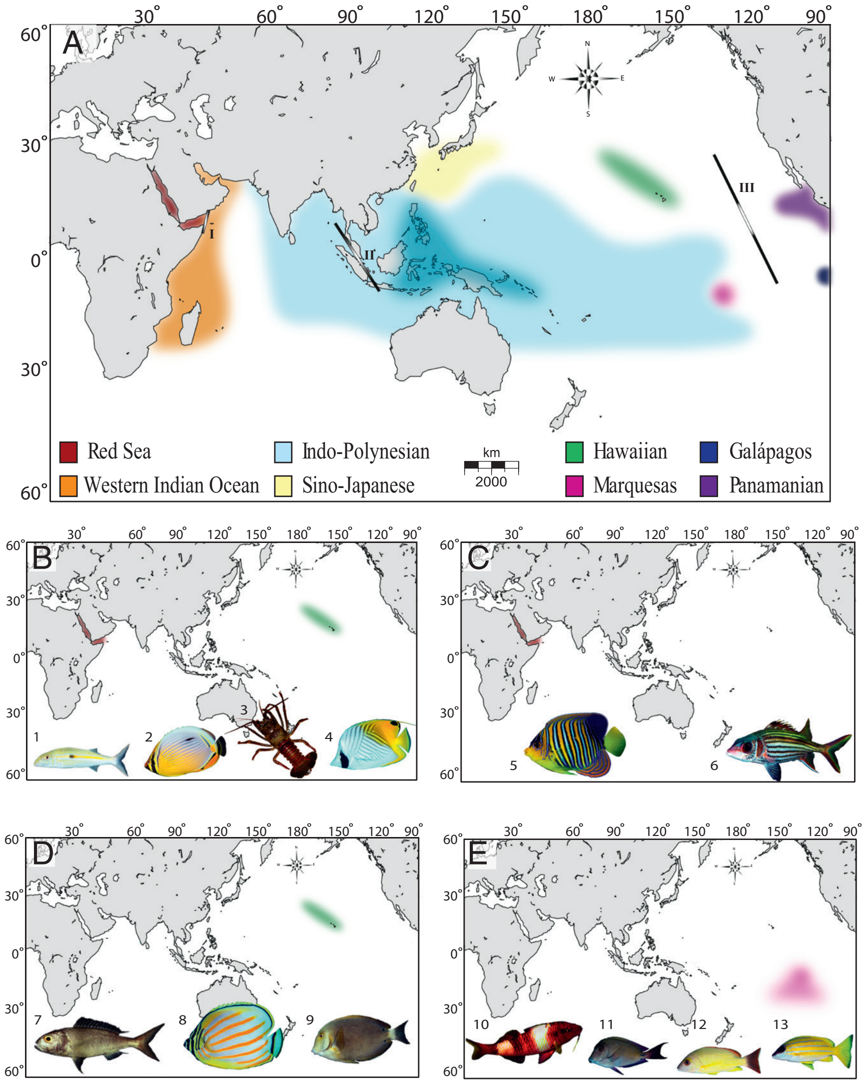


FIGURE 1.1 (*A*) Biogeographic provinces of the tropical Indo-Pacific as defined by >10% endemism (Briggs and Bowen, 2012). Coral Triangle is indicated in dark blue. Primary barriers include (site I) Red Sea Barrier, (site II) Indo-Pacific Barrier, and (site III) East Pacific Barrier. (*B–E*) Minimaps illustrating widespread species with phylogeographic separation (strong allele-frequency shifts and significant *F*-statistics) at peripheral provinces. For each panel, the peripheral region(s) of phylogeographic distinction is highlighted in color, and photos are of the species with genetic evidence for that pattern as follows: (*B*) Hawai'i and the Red Sea [1,

structure indicate that novel species are arising by parapatric means within the Coral Triangle, wherein partial isolation between subregions reinforces isolation along ecological gradients (Timm et al., 2008; Carpenter et al., 2010; Barber et al., 2011; Tornabene et al., 2015). Based on phylogenies of three reef fish families, Cowman and Bellwood (2013) estimate that 60 percent of Indo-Pacific reef fauna have origins in the Coral Triangle. In contrast, peripheral endemism hotspots were previously regarded as evolutionary dead ends (Alison Kay and Palumbi, 1987; Bellemain and Ricklefs, 2008), in which rare colonization events can produce endemic species, but with no further evolutionary radiations. This assumption has been challenged in recent years because phylogeographic studies show that both Hawaiian and Red Sea provinces can export novel biodiversity (DiBattista et al., 2013; Eble et al., 2015).

The dominant feature of tropical marine biogeography is the vast Indo-Polynesian Province (IPP), spanning almost half the planet (Fig. 1.1*A*). Concomitant with this large province are unusually large range sizes, averaging 9 million km^2 for reef fishes, roughly the size of mainland China (Allen, 2008). Genetic surveys of reef organisms are generally consistent with the boundaries of the IPP, showing little genetic structure across broad areas with a few important exceptions (e.g., Indo-Pacific Barrier) (Eble et al., 2015). Schultz et al. (2008) use bathymetry profiles to demonstrate that dispersal across most of this range (Polynesia to Western Australia) requires no deep-water traverse greater than 800 km. Undoubtedly, this continuity of shallow habitat contributes to the cohesiveness of the IPP.

Mulloidichthys flavolineatus (Fernandez-Silva et al., 2015); 2, *Corallochaetodon* species complex (Waldrop et al., 2016); 3, *Panulirus penicillatus* (Iacchei et al., 2016); 4, *Chaetodon auriga* (DiBattista et al., 2015)]; (C) Red Sea only [5, *Pygoplites diacanthus* (Coleman et al., 2016); 6, *Neoniphon sammara* (DiBattista et al., 2013)]; (*D*) Hawai'i only [7, *Pristipomoides filamentosus* (Gaither et al., 2011b); 8, *Chaetodon ornatissimus* (DiBattista et al., 2012); 9, *Acanthurus nigroris* (DiBattista et al., 2011)]; (*E*) Marquesas/French Polynesia [10, *Parupeneus multifasciatus* (Szabó et al., 2014); 11, *Acanthurus nigrofuscus* (Eble et al., 2011); 12, *Lutjanus fulvus* (Gaither et al., 2010); 13, *Lutjanus kasmira* (Gaither et al., 2010)]. Photo credits: J. E. Randall/FishBase (photograph 7); Tane Sinclair-Taylor (all other fish photographs); Matthew Iacchei (photograph of *Panulirus penicillatus*).

At the center of this vast province is an intermittent barrier around the Indo-Malay Archipelago, known as the Indo-Pacific Barrier (Fig. 1.1*A*). In the mid-Miocene (16–8 Ma), the Australian and Eurasian plates collided and reduced water flow between the Pacific and Indian Oceans (Kennett et al., 1985). During Pleistocene glacial cycles, sea level dropped as much as 130 m below present levels, further constricting connections between these ocean basins. Evidence for interruptions of gene flow can be found in the distributions of sister species, coupled with phylogeographic partitions (as defined by reciprocal monophyly or $\Phi_{ST} > 0.10$) in green turtles (Dethmers et al., 2006), dugongs (Blair et al., 2014), and ~80 percent of surveyed reef species (Fig. 1.2) (Gaither et al., 2010; Ahti et al., 2016). Given the cyclic nature of this barrier, phylogeographic partitions driven by Pleistocene glacial fluctuations are expected to be concordant in terms of geography, but not necessarily concordant in terms of chronology.

On the eastern periphery of the enormous IPP are three isolated provinces with high endemism in reef fishes: (*i*) the Hawaiian Islands with 25 percent endemism (Randall, 2007), (*ii*) the Marquesas Islands with 13.7 percent endemism (Delrieu-Trottin et al., 2015), and (*iii*) Easter Island with 21.7 percent endemism (Randall and Cea, 2011). Phylogeographic studies of the first two provinces show strong concordance with biogeographic partitions (Fig. 1.1*B*). In Hawai'i, 11 of 16 fishes surveyed are genetically distinct from conspecifics elsewhere in the Pacific (reviewed in Gaither et al., 2011b). At the Marquesas, three of five studies reveal divergences that range from $F_{ST} \geq 0.24$ at allozyme loci to reciprocal monophyly at mtDNA (Gaither et al., 2010; Szabó et al., 2014), and a RADSeq study reveals strong divergence between a Marquesan surgeonfish and a widespread sister species (Gaither et al., 2015a).

On the western side of the IPP lies the Red Sea biogeographic province, an endemism hotspot characterized by a shallow connection to the Indian Ocean and latitudinal gradients in temperature, salinity, and nutrient load (DiBattista et al., 2016a,b). Many Red Sea endemics have sister species in the adjacent Western Indian Ocean (DiBattista et al., 2016a). This interspecific pattern aligns with mtDNA partitions within species ranging from haplotype frequency shifts to reciprocal monophyly in fishes and invertebrates (table 2 in DiBattista et al., 2016a). For example, the Indo-Pacific damselfish (*Dascyllus aruanus*; Liu et al., 2014) and yellowstripe goatfish (*Mulloidichthys flavolineatus*; Fernandez-Silva et al., 2015) both demonstrate similar divisions in mtDNA sequences ($\Phi_{ST} > 0.65$) and microsatellite genotypes ($F_{ST} > 0.03$). In some cases, coalescence analyses reveal that Red Sea lineages are older than those in the Indian Ocean, indicating that the former can export biodiversity to adjacent waters (DiBattista et al., 2013).

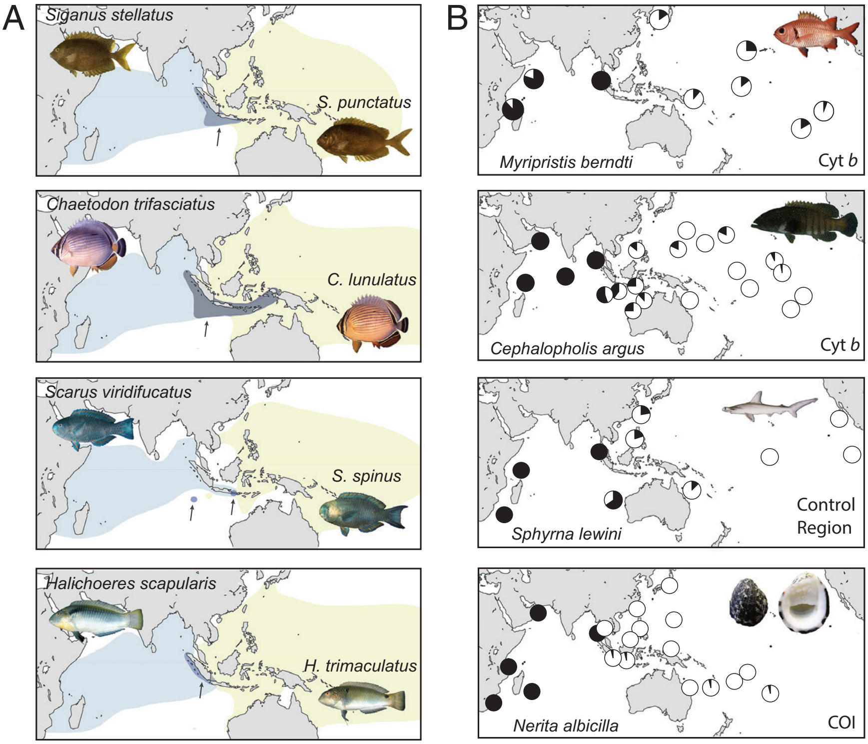


FIGURE 1.2 Evidence of isolation across the Indo-Pacific Barrier. (*A*) Distribution patterns of sister species pairs. Distributions shaded in purple (and indicated by arrows) represent areas of species overlap. (*B*) Phylogeographic studies demonstrating divergent genetic lineages within species. Black and white in pie diagrams indicate distribution of mtDNA phylogroups separated by at least three mutations. In all cases, there is evidence of population expansion with overlap in the Indo-Malay-Philippine biodiversity hotspot (Coral Triangle) (Crandall et al., 2012). *Myripristis berndti* data from Craig et al. (2007), *Cephalopholis argus* data from Gaither et al. (2011a), *Sphyrna lewini* data from Duncan et al. (2006), and *Nerita albicilla* data from Crandall et al. (2008). COI, cytochrome oxidase subunit 1; Cyt *b*, cytochrome *b*. Photo credit: J. E. Randall for fishes, Wikimedia commons/Harry Rose for *Nerita albicilla.*

For widely distributed species, genetic divergences at peripheral locations may be the inception of speciation. The pronghorn spiny lobster, *Panulirus penicillatus*, with a 9-mo pelagic larval duration and a distribution across the entire tropical Indo-Pacific, illustrates genetic diversification at both ends of its range. Iacchei et al. (2016) found fixed differences in mtDNA of East Pacific and Red Sea populations ($\Phi_{ST} = 0.74$), corroborated by morphological differentiation in the East Pacific (George, 2005). Speciation in peripheral provinces is apparent in *Thalassoma* wrasses (Bernardi et al., 2004), *Anampses* wrasses (Hodge et al., 2012), *Acanthurus* surgeonfishes (Gaither et al., 2015a), *Mulloidichthys* goatfishes (Fernandez-Silva et al., 2015), and *Montastraea* corals (Gaither et al., 2016).

The East Pacific Barrier (EPB) limits the distribution of tropical species (Gaither et al., 2016), with few taxa able to maintain population connectivity across the EPB, as evidenced by the lobster *P. penicillatus* (Iacchei et al., 2016), and the coral *Porites lobata* (Baums et al., 2012; Forsman et al., 2015). However, some fishes (Lessios and Robertson, 2006) and the echinoderm *Echinothrix diadema* (Lessios et al., 1998) have low or insignificant Φ_{ST} values across the EPB.

Atlantic and Indo-Pacific Connections

Two geological events isolated the tropical Atlantic from the Indo-Pacific: (*i*) closure of the Tethys Sea ~13 Ma, brought about by the collision of Africa and Eurasia; and (*ii*) the rise of the Isthmus of Panama ~3.5 Ma that separated the Atlantic from the East Pacific Ocean (Lessios, 2008). For the latter, some species diverged well before the final closure, although the timing of partitions remains controversial (Marko et al., 2015) (a fruitful topic for genomic studies). Since the closure of the Tethys Sea, natural dispersal between the Atlantic and Indian Oceans has been limited to the hydrographically complex waters around southern Africa (Teske et al., 2011). A warm-water corridor here was curtailed ~2.5 Ma by the advent of modern glacial cycles and upwelling in the Benguela Current on the Atlantic side (Dwyer et al., 1995). However, the Agulhas Current on the Indian Ocean side occasionally forces warm-water gyres into the Atlantic (Hutchings et al., 2009), a potential route of colonization. Phylogeographic studies confirm sporadic dispersal along this route over the last 2.5 My, primarily from the Indian to Atlantic Ocean (Reece et al., 2010; Gaither et al., 2015b).

Summary

In conclusion: (*i*) Biodiversity hotspots and peripheral centers of endemism both produce and export novel evolutionary lineages. (*ii*) Phylogeographic partitions, as defined by mtDNA monophyly or strong population structure, align well with the biogeographic provinces defined by taxonomy. (*iii*) Sporadic dispersal around southern Africa is the primary avenue of colonization between Indo-Pacific and Atlantic Oceans.

Temperate and Polar Seas

Northern seas experienced greater extremes in temperature over the Pleistocene than tropical seas, and northern near-shore ecosystems were periodically eradicated by glaciers encroaching onto continental shelves, whereas interglacial warming led to colonizations and population expansions. Although phylogeographic structure generally occurs between biogeographic provinces, sub-Arctic shelf fauna have been repeatedly disrupted by glacial cycles (Marko et al., 2010). Therefore, present-day physical barriers to gene flow may not exert the same influence on phylogeographic patterns as observed in more stable tropical seas. The most notable barriers separating biogeographic domains are the large expanses of ocean waters across the North Pacific and North Atlantic.

North Pacific

Species in the temperate regions on both sides of the North Pacific show a range of evolutionary divergences that largely depend on dispersal capabilities, temperature tolerances, and climate history. Taxa at higher latitudes tend to have distributions that span the North Pacific (versus taxa at midlatitudes). For example, cold-tolerant cods (*Gadus*), herring (*Clupea*), and king crabs (*Lithodes, Paralithodes*) occur in both the Northwest and Northeast Pacific. Most of these trans-Pacific species show phylogeographic breaks, centered on the Aleutian Archipelago or eastern Bering Sea, that represent secondary contact zones after repeated isolations (Canino et al., 2010; Liu et al., 2012; Grant et al., 2014). In contrast, temperate fishes, invertebrates, and seaweeds at midlatitudes are generally limited to one side of the North Pacific, with closely related species on the other side. A notable exception are disjunct populations of Pacific sardines (*Sardinops*) in the Northwest and Northeast Pacific (Bowen and Grant, 1997).

North Atlantic

This basin is smaller than the North Pacific and has a U-shaped shoreline with Greenland, Iceland, and the Faroe Islands in midocean. Populations of fishes, invertebrates, and seaweeds show a range of genetic divergences across the North Atlantic (Árnason, 2004; Addison and Hart, 2005; McCusker and Bentzen, 2010). Conspecific populations on either side of the North Atlantic were isolated during glacial episodes, and, in some taxa, the Northwest Atlantic was extirpated and reestablished after the Last Glacial Maximum. Some populations in the Northwest Atlantic show closer genetic affiliations to the North Pacific than to the Northeast Atlantic (seagrass and sea urchins) (Olsen et al., 2004). The Baltic, North Sea, and

Mediterranean biogeographic provinces are isolated to some extent from the Atlantic by narrow straits, which often coincide with phylogeographic transitions (Johannesson and André, 2006; Patarnello et al., 2007).

Arctic Biogeographic Province

The far northern ocean has served as a pathway for dispersal between the North Atlantic and North Pacific (Vermeij, 1991). Phylogeographic and taxonomic studies reveal sister species in the North Atlantic and North Pacific, including several fishes (Grant, 1987), invertebrates (Vermeij, 1991), and seaweeds (Lindstrom, 2001). During ~20 percent of the Pleistocene, high sea levels breached the 50-m sill across the Bering Strait (Miller et al., 2005), allowing interocean dispersal as early as 6.4 Ma and again at 3.5 Ma (Marincovich and Gladenkov, 1999). More recent dispersal events have led to the co-occurrence of conspecific populations in both oceans (Carr et al., 1999).

Antarctic Biogeographic Province

The Antarctic is relatively old, ~25 My, compared with about 2.5 My for the Arctic. The result of this ancient formulation is high endemism: 88 percent in fishes (Eastman, 2005) and 42–56 percent in four invertebrate classes (Griffiths et al., 2009). The high homogeneity of taxa across this vast region is facilitated by the Antarctic Circumpolar Current, which circles the entire continent. Phylogeographic studies are consistent with a highly connected Antarctic Province, showing little (or no) population structure for two decapods (Raupach et al., 2010), one nemertean (Thornhill et al., 2008), and four ice fishes (Janko et al., 2007).

Patterns Within Biogeographic Provinces

Within the shallow-water provinces, species often share genetic breaks at specific geological features or geographical regions. Examples range from the classic study by Avise (1996) on the Carolina Province (Southeast United States), through more recent surveys of the benthic fauna along the coast of New Zealand (Ross et al., 2009), the northeastern Pacific (Kelly and Palumbi, 2010), the Coral Triangle (Carpenter et al., 2010), southern Africa (Teske et al., 2011), and Hawai'i (Toonen et al., 2011). Endemic species confined to a single province tend to show more population structure than widespread species at the same geographic scale (Carpenter et al., 2010; Tenggardjaja et al., 2014, 2016). Species that lack pelagic development generally show strong genetic structure whereas species with pelagic development are less predictable (Riginos et al., 2014; Liggins et al., 2016).

Regardless of developmental mode, ecological niche, or evolutionary relationships, species showing geographic structuring often have concordant genetic breaks, indicating that shared history or physical factors drive the observed pattern (Avise, 1996). Examination of 47 reef-associated species across the Hawaiian Archipelago reveals that multispecies trends in genetic diversity are driven by a combination of both the dominant physical, historical, and ecological features of the seascape, and ecological–genetic feedback within communities (Selkoe et al., 2016).

Species that counter these trends may be particularly informative about the process of evolution. For example, Hawaiian limpets of the genus *Cellana* have diversified within the archipelago along a tidal gradient that indicates ecological speciation (Bird et al., 2011). Certainly, species sharing population structure at unexpected locations within biogeographic provinces (such as Fiji in the tropical Pacific) (Drew and Barber, 2012; DiBattista et al., 2015), or other exceptions to those general trends, will provide evolutionary insights.

Summary

In conclusion: (*i*) Species distributions are fundamentally shaped by physiological tolerances to north–south temperature gradients in the North Pacific and North Atlantic. (*ii*) Glacial cycles impact phylogeography by repeatedly altering species distributions, isolating populations, and creating secondary contact zones. (*iii*) Shifting interactions between ocean–climate, coastal configuration, and bottom topography produce barriers to dispersal between ocean basins. (*iv*) Some biogeographic provinces are genetically homogeneous, with little opportunity for allopatric divergences, whereas others host heterogeneous habitats that can promote speciation along ecological boundaries.

TAXON-SPECIFIC PATTERNS

Migratory ability and historical dispersal define taxa along a continuum of evolutionary divergence. Clusters of closely related species, each confined to a single biogeographic province, are at one end of the continuum, and highly migratory megafuana are at the other end. Oceanic migrants provide special challenges to both phylogeographic studies and conservation strategies, because both must be conducted on a scale that transcends biogeographic provinces and political jurisdictions (Toonen et al., 2013). Species in the center of the continuum include temperate taxa inhabiting disjunct regions, such as antitropical taxa, sister species separated by the tropics. Comparative phylogeography of these groups provides insights into the roles of dispersal and isolation in contributing to biodiversity.

Antitropical Taxa

Species with disjunct distributions on both sides of the tropics provide fascinating subjects for phylogeographic study. Equatorial surface waters are lethal to these cold-adapted species, so how do they cross the tropics, and how often can this crossing be accomplished? Sister taxa of fishes on each side of the equator reveal divergences ranging from populations to distinct lineages, but without a clear pattern. For example, a single species of anchovy (*Engraulis*) occurs in the North Atlantic, southern Africa, and Japan, but three additional species have more restricted ranges (Grant et al., 2005; Silva et al., 2014). In contrast, a single species of sardine (*Sardinops*) extends from southern Africa to Australia to Chile, California, and Japan (Bowen and Grant, 1997).

Overall results show that the ability to traverse the tropics is species specific and that these events have not been limited to particular periods of global cooling. However, one possible point of concordance includes the eastern continental margins of the Atlantic (for anchovies) and the Pacific (for sardines). In both cases, colonizations across the equator have been accomplished recently, as indicated by shared mtDNA haplotypes (Fig. 1.3).

Cetaceans

Patterns of gene flow vary extensively across space and time for cetaceans, driven largely by the wide variety of life history traits (Hoelzel, 1998; Andrews, 2014). Most species exhibit limited gene flow between ocean basins, even in taxa with temperate distributions; but genetic structure within ocean basins varies substantially across species. For Mysticetes (baleen whales), patterns of gene flow are shaped by migratory pathways, with individuals typically exhibiting maternally based site fidelity to tropical breeding and temperate/Arctic feeding areas. This fidelity leads to population genetic separations between ocean basins and among breeding areas, with F_{ST} values of 0.05 to 0.1 for right whales (Carroll et al., 2015), blue whales (Torres-Florez et al., 2014), and humpback whales (Jackson et al., 2014).

In contrast, most Odontocetes (toothed whales) do not undertake large-scale migrations and often exhibit genetic structure over relatively short geographic distances due to site fidelity, resource specialization, and social structure. For example, strong fidelity to narrow ranges can result in genetically divergent populations along continuous coastlines or

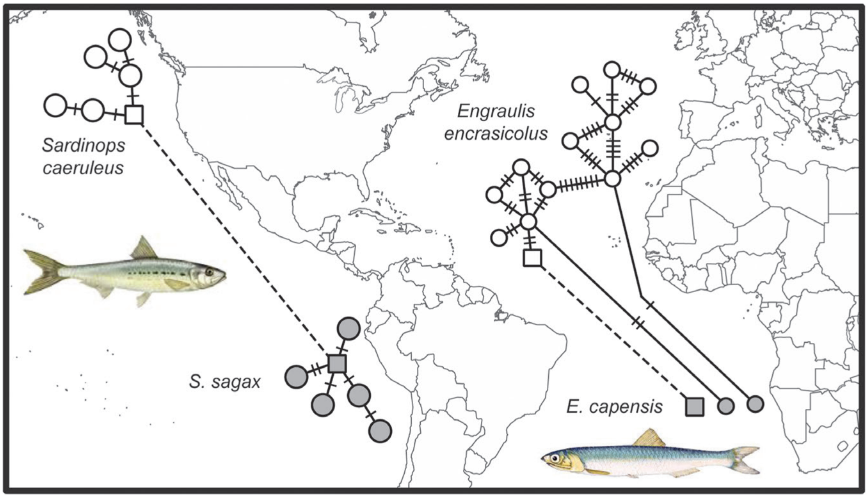


FIGURE 1.3 Sardines (genus *Sardinops*) and anchovies (genus *Engraulis*) are antitropical species that recently surmounted the warm-water barrier between the Northern and Southern Hemispheres, as indicated by mtDNA haplotype networks. For sardines in the East Pacific, transequatorial dispersal is facilitated by a short and steep continental shelf and adjacent deep cold water (Bowen and Grant, 1997). For anchovies in the East Atlantic, transequatorial dispersal is facilitated by upwelling (cold nutrient-rich water) in low latitudes (Grant et al., 2005). Light and dark haplotypes indicate Northern and Southern Hemisphere, respectively. Squares connected by a dashed line indicate haplotypes shared between hemispheres. Note that, in the East Pacific sardine, the haplotype shared between Northern and Southern Hemispheres is internal to both networks, indicating an ancient connection. In contrast, the East Atlantic anchovy has connections across the equator that include both interior and peripheral haplotypes in the network.

between adjacent islands, as is the case for spinner dolphins (Andrews et al., 2010), Hector's dolphins (Hamner et al., 2012), and Indo-Pacific humpback dolphins (Brown et al., 2014). Some Odontocetes have ecologically and behaviorally distinct groups ("ecotypes"), with limited gene flow even in parapatry or sympatry (Hoelzel, 1998). Several dolphin species contain genetically divergent coastal and pelagic ecotypes (Andrews et al., 2013). Killer whales have sympatric ecotypes that differ in prey type, foraging strategy, social structure, and movement (Hoelzel et al., 1998).

Sea Turtles

The seven species of sea turtles show patterns of population structure within ocean basins defined by natal homing, the habit of females (and sometimes males) to return to the vicinity of their natal beach, after decades of growth in ocean and coastal habitats. This behavior is the basis for defining regional management units (Wallace et al., 2010). On a global scale, occasional wandering provides connections between nesting populations and ocean basins. Cold-tolerant species, such as the leatherback turtle, pass freely between ocean basins (Dutton et al., 1999). Tropical species, such as the green turtle and the hawksbill turtle, make rarer connections between the Atlantic and Indo-Pacific via southern Africa (Bourjea et al., 2007; Vargas et al., 2016). Bowen and Karl (2007) note higher genetic divergences between ocean basins in tropical species, providing a signal that allopatric speciation may predominate in this group.

Pelagic Fishes

A primary phylogeographic pattern for these oceanic migrants is low to no genetic structure within ocean basins, and strong genetic structure between the Atlantic and Indo-Pacific. Some pelagic species seem to cross the Benguela Barrier (southern Africa) often enough to preclude the development of evolutionary partitions, including albacore tuna (Vinas et al., 2004; Montes et al., 2012), wahoo (Theisen et al., 2008), and the common dolphinfish (Diaz-Jaimes et al., 2010). However, these species are likely exceptions, with many large, vagile species demonstrating structured populations across this barrier, including the scalloped hammerhead shark (Duncan et al., 2006), whale shark (Castro et al., 2007), and blue marlin (Buonaccorsi et al., 2001). For tunas in particular, a recurring pattern is two mtDNA lineages: one confined to the Atlantic and an Indo-Pacific lineage that is also found in the Atlantic (table 6 in Theisen et al., 2008). This pattern indicates extended periods of isolation, punctuated by dispersal around southern Africa.

Plankton

In the oceanic pelagic zone, where all life stages are planktonic, species' ranges are both extensive and dynamic because adult distributions are not tied to a particular benthic habitat. In turn, biogeographic provinces for the pelagic zone are based on physical and chemical properties (biogeochemical provinces) (Longhurst, 1995) rather than endemism or species assemblages. Longhurst (2007) identified ~55 biogeochemical provinces (BGCPs), nested within four biomes (Polar, Westerly Winds, Trade Winds, Coastal) across four ocean basins (Atlantic, Pacific, Indian, Southern). Like the species they harbor, the boundaries of the BGCPs fluctuate on both seasonal and annual timescales in accordance with changing environmental conditions (Reygondeau et al., 2013). Our understanding of pelagic community composition is still nascent, but recent studies have shown concordance between BGCPs and community composition in taxa ranging from viruses (Brum et al., 2015) to phytoplankton (Alvain et al., 2005) to fishes (Reygondeau et al., 2012).

Cosmopolitan distributions in the pelagic zone initially prompted the conclusion of little to no population structure in the open ocean, a position that has eroded in recent decades (Miya and Nishida, 1997; Norris, 2000). Phylogeographic studies reveal that many cosmopolitan taxa are composed of multiple cryptic species (Miyamoto et al., 2012; Hirai et al., 2015), including some that are sympatric over part of their ranges (Andrews et al., 2014b). Populations of these cosmopolitan species are subdivided in two ways concordant with the BGCP framework: (*i*) by continental land masses separating ocean basins, and (*ii*) by habitat discontinuities in the equatorial region between subtropical gyres in the Northern and Southern Hemispheres (Goetze, 2011; Norton and Goetze, 2013; Andrews et al., 2014b). The few global-scale phylogeographic studies have been restricted to copepods, but evidence from a diversity of other taxa sampled at ocean basin scales indicates that lineages have diverged both in allopatry and sympatry at much smaller geographic distances than anticipated, with examples drawn from chaetognaths (Peijnenburg et al., 2006), euphausiids (Bucklin et al., 2007), and mollusks (Burridge et al., 2015).

These combined results indicate that population discontinuities of pelagic species are determined not by the temporal and spatial scales of dispersal, but by habitat characteristics enabling species to maintain viable populations (Norris, 2000; Peijnenburg et al., 2006). Habitat selection, rather than physical barriers, may be a primary force driving speciation in the pelagic zone (Peijnenburg and Goetze, 2013). Therefore, a biogeographic framework based on water properties is concordant with genetic partitions within species.

Summary

In conclusion: (*i*) Several temperate species show disjunct distributions across the tropics, indicating historical dispersals across warm-water barriers. (*ii*) The deepest phylogeographic separations for oceanic migrants indicate patterns of allopatric isolation between ocean basins, especially for fishes. (*iii*) Migratory sea turtles and cetaceans show population structure based on reproductive site fidelity. (*iv*) An ecological component to speciation is indicated by isolation along behavioral barriers in cetaceans, and by the presence of sympatric sister species in the plankton.

(*v*) Planktonic biogeographic provinces are defined by water masses that can change size and position based on oceanographic conditions. (*vi*) Initial plankton studies indicate concordance between biogeochemical provinces and phylogeographic partitions, particularly at the equatorial break between northern and southern subtropics.

TERRESTRIAL VS. MARINE PHYLOGEOGRAPHY

Life began in the oceans, but the field of phylogeography began with continental biota (Avise et al., 1987; Riddle, Chapter 2, this volume), and many of the insights reviewed here have precedents in terrestrial cases. The biogeographic settings have parallels between land and sea, particularly with latitudinal gradients in biodiversity and concordance between biogeographic provinces and phylogeographic partitions (Riddle, Chapter 2, this volume; Schluter, 2016). Glacial habitat disruptions in northern seas have a strong parallel in continental faunas (Hewitt, 1996; Bernatchez and Wilson, 1998). Biodiversity hotspots in Indo-Pacific reefs, forests of northern Australia, and Neotropical plant communities are all distinguished by periods of stability, habitat heterogeneity, and the ability to export species (Moritz et al., 2013; Pellissier et al., 2014; Antonelli et al., 2015). A primary difference between marine and terrestrial phylogeography is greater dispersal potential and fewer barriers in the oceans. Although a squirrel in Central Park (New York) cannot deposit progeny in Hyde Park (London), a squirrelfish is capable of dispersing on this scale (Craig et al., 2007). This difference in evolutionary processes is clear in the Hawaiian Archipelago, where rare terrestrial colonists have proliferated into dozens and hundreds of species (Shaw and Gillespie, Chapter 4, this volume) whereas marine colonists produce one or a few species (Bird et al., 2011). Therefore, the evolutionary dramas above and below the waterline have the same ingredients (isolation, selection, adaptation, speciation), but markedly different tempos and outcomes (Bowen, 2016).

CONCLUSION

Marine phylogeography encompasses half-billion-year separations and the largest habitat on the planet. Given this diversity, generalizations are few, but some are especially robust. First, phylogeography is the new incarnation of spatial biogeography (Arbogast and Kenagy, 2001). The alignment of population genetic separations and taxonomic distributions reveals that these are part of a continuum. Evolutionary partitions that could previously be described only with taxonomy are now evaluated with the genomic footprints of isolation, selection, and speciation. Second, the model of allopatric speciation that previously dominated evolutionary thought is an incomplete fit to the dispersive aquatic medium. Phylogeography of oceanic migrants indicates a strong role for allopatric speciation, whereas heterogeneous coastal habitats provide more opportunity for sympatric/ecological divergences. Phylogeography in high latitudes is defined by shifting habitats in response to glaciation. Finally, both biodiversity hotspots and endemism hotspots are important in producing novel evolutionary lineages and may work in synergy to enhance biodiversity on the ocean planet.

ACKNOWLEDGMENTS

We thank John C. Avise and Francisco J. Ayala, co-organizers of *In the Light of Evolution X: Comparative Phylogeography*, an Arthur M. Sackler Colloquium of the National Academy of Sciences. For stimulating discussions, advice, and editorial prowess, we thank E. A. Hanni, P. Marko, S. A. Karl, J. Eble, L. Rocha, G. Bernardi, M. Berumen, and the B.B./R.J.T. (ToBo) laboratory. We thank T. Sinclair-Taylor, T. Lilley, and A. Cros for assistance with illustrations and editor John Avise and two anonymous reviewers for comments that improved the manuscript. The authors' research reported here was funded by the National Science Foundation, the Seaver Institute, the University of Hawai'i Sea Grant Program, the National Oceanic and Atmospheric Administration, the North Pacific Research Board, and the Saltonstall–Kennedy Grant Program.

2

Comparative Phylogeography Clarifies the Complexity and Problems of Continental Distribution That Drove A. R. Wallace to Favor Islands

BRETT R. RIDDLE

Deciphering the geographic context of diversification and distributional dynamics in continental biotas has long been an interest of biogeographers, ecologists, and evolutionary biologists. Thirty years ago the approach now known as comparative phylogeography was introduced in a landmark study of a continental biota. Here, I use a set of 455 studies to explore the current scope of continental comparative phylogeography, including geographic, conceptual, temporal, ecological, and genomic attributes. Geographically, studies are more frequent in the Northern Hemisphere, but the south is catching up. Most studies focus on a Quaternary time frame, but the Neogene is well represented. As such, explanations for geographic structure and history include geological and climatic events in Earth history, and responses include vicariance, dispersal, and range contraction-expansion into and out of refugia. Focal taxa are biased toward terrestrial or semiterrestrial vertebrates, although plants and invertebrates are well represented in some regions. The use of various kinds of nuclear DNA markers is increasing, as are multiple-locus studies, but use of organelle DNA is not decreasing. Species distribution models are not yet widely incorporated into studies. In the future, continental comparative phylogeographers will continue to contribute to erosion of the simple vicariance vs. dispersal paradigm, including exposure of the widespread nature of temporal pseudocongruence and

---

School of Life Sciences, University of Nevada, Las Vegas, NV 89154. Email: brett.riddle@unlv.edu.

its implications for models of diversification; provide new templates for addressing a variety of ecological and evolutionary traits; and develop closer working relationships with earth scientists and biologists in a variety of disciplines.

A landmark study 30 years ago (Bermingham and Avise, 1986) used mitochondrial DNA (mtDNA) variation to reveal a suite of "cryptic," geographically structured evolutionary lineages embedded within four codistributed species of freshwater fishes in the southeastern United States and attributed interspecific similarities in lineage distributions to shared histories of isolation and divergence, that is, vicariance, across putative geographic barriers. Comparative phylogeography as applied to continental biotas subsequently has expanded in terms of the variety of questions addressed, theory and methods applied, number of regions on different continents explored, and variety of taxa and genomes used. However, of course, biogeographers had long before understood the importance of exploring geographic patterns across multiple codistributed taxa.

CONTINENTAL BIOGEOGRAPHY BEFORE COMPARATIVE PHYLOGEOGRAPHY

In the late 19th century, Alfred Russel Wallace mapped distributions of multiple vertebrate and invertebrate taxa to develop the system of terrestrial zoogeographic regions and subregions used more or less intact to this day (Wallace, 1876), although modifications continue to be proposed using new datasets, concepts, and methods (Morrone, 2015) (Fig. 2.1*A*). However, Wallace was interested in more than pattern, and in particular how Earth features and history (geological and climatic) might be responsible for generating geographic patterns of distribution and diversification. In his consideration of continental biotas, he speculated on the influence of barriers: for example, the uniqueness of "zoological districts" on opposite sides of the Rocky Mountains and Andes (Wallace, 1876) and the influence of large Neotropical rivers on the distributional limits of several species of monkeys (Wallace, 1852). He associated distributional changes and extinction of animals in Europe and North America primarily with the warm and cold climate cycles of "Glacial Epochs," but was largely wrong about the role of geological history, dismissing the continental drift model of Earth history (Wallace, 1880) nearly a century before the plate tectonics framework was introduced. He did, however, predict that geologists eventually would catch up to biologists in reconstructions of Earth history at focal spots such as the biogeographic transition zone between the Neotropical and Nearctic regions (Wallace, 1852).

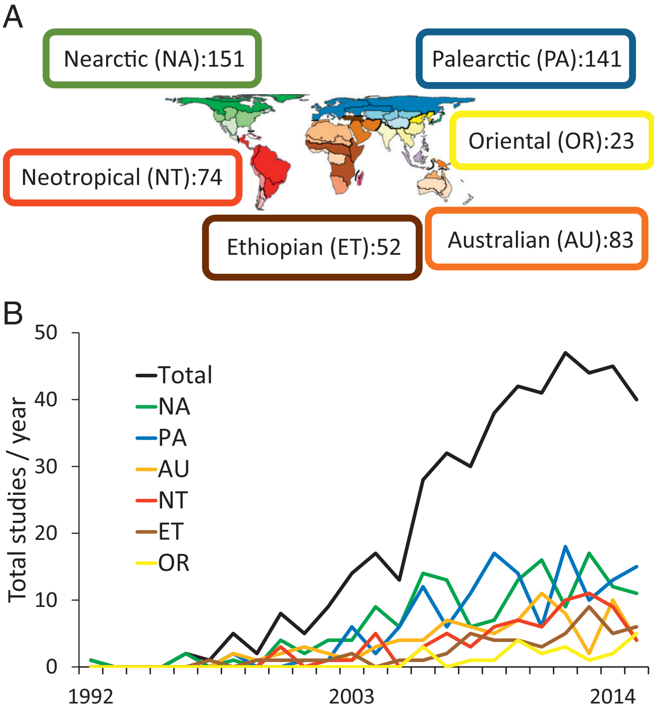


FIGURE 2.1 (*A*) Six terrestrial biogeographic regions of the world, basically dating to Wallace (1876), but this depiction (map modified with permission from Kreft and Jetz, 2010) with boundaries modified modestly and with proposed subregions (see original paper for additional terminology). Numbers refer to distribution of comparative phylogeographic studies of continental biotas using references database described further in text. (*B*) Growth of continental comparative phylogeographic studies 1992–2015, total and tabulated by biogeographic region using references database described further in text (numbers for 2015 not complete because database was downloaded from Web of Science in December 2015).

In his last grand statement on biogeography in his masterful 1880 book *Island Life*, Wallace turned his primary focus away from continents and to islands:

> In islands we have the facts of distribution often presented to us in their simplest forms, along with others which become gradually more and more complex; and we are therefore able to proceed step by step in the solution of the problems they present. . . . *It is not too much to say that*

> *when we have mastered the difficulties presented by the peculiarities of island life we shall find it comparatively easy to deal with the more complex and less clearly defined problems of continental distribution.* (Wallace, 1880, p. 234, italic added)

Islands, in his view, with their simplified (less diverse) biotas and discreet boundaries, could provide insights into the development of continental biotas that would be difficult to extract from the latter alone.

Nevertheless, biogeographers continued to use a "multiple codistributed taxa" approach as applied to continental biotas in the century between Wallace's *Island Life* and the comparative phylogeography introduced by Bermingham and Avise (1986). Jordan (1905) and Grinnell (1914) used comparative evidence when examining the role of geographic isolation across barriers in the production of new species; indeed, Grinnell (1914, p. 253) remarked that "By such a study, of *comparative distribution*, it seems possible that the ranges of birds and mammals may become subject to satisfactory explanation." Wallace's keen interest in the influence of Glacial Epochs was expanded by Adams (1905) into an examination of postglacial distributional shifts in five geographically separated North American biotas out of glacial age "biotic preserves" (in modern terms, refugia). Haffer (1969) proposed a Pleistocene species pump model of avian diversification in the Amazonian region resulting from repeated cycles of forest expansion and contraction. Remington (1968) proposed 13 "suture zones" between major biotic assemblages in North America and suggested some level of hybridization might be occurring in secondary contact between previously isolated species and subspecies. Rosen (1978) used North and Middle American biotas, in conjunction with paleogeographic reconstructions of the region, to outline newly forming theory in "vicariance biogeography." Interactions between biotas across biogeographic regions included a classic study of biotic interchange by Simpson (1950), who focused on the transition zone between Neotropical and Nearctic regions with an accounting of the historical assembly of the South American mammal fauna that included a relatively recent wave of immigration from North America. Darlington (1959) took a more global view than Simpson on dispersal among biogeographic regions, developing perhaps the last preplate tectonics and distinctly nonvicariance model, wherein waves of "dominant" animals would disperse away from an Old World tropics center of origin, displacing inferior species as they invaded northern temperate and southern regions. Investigators increasingly discovered value in use of molecular markers to resolve geographic variation within species and among higher taxa (Avise, 2000), and perhaps the best example of examining codistributed species before the advent of mitochondrial DNA-based phylogeography was a 1984 study that used microcomplement fixation of serum albumin to place geographic isolation and divergence of several Australian frogs within a pre-Quaternary time frame (Maxson and Roberts, 1984).

My main objective here is to address several of the contributions of comparative phylogeography to developing a better understanding of diversification and distributional dynamics in continental biotas. To do so, I first provide an overview of the growth, both conceptually and geographically, of the scope of continental comparative phylogeography over the 30 years since its beginnings (Bermingham and Avise, 1986). I end with a few predictions regarding future directions for a comparative phylogeography of continental biotas and its potential role in the advancement of related disciplines.

THE CONCEPTUAL AND GEOGRAPHIC SCOPE OF CONTINENTAL COMPARATIVE PHYLOGEOGRAPHY

In this section, I explore the context and content of continental comparative phylogeography using a database of hundreds of titles and abstracts of publications within this arena, beginning chronologically with the paper that established firmly the conceptual framework and potential of comparative phylogeography (Avise, 1992).

Building the Database

I began with three topics searches of the Web of Science online database (last search: 2 December 2015) using the following terms: "comparative phylogeograph*"; codistributed not "comparative phylogeograph*"; and codistributed not "comparative phylogeograph*." The basic search returned nearly 900 citations, from which my first pass sought to delete those that dealt with geographies that were not concerned with continental biotas (e.g., islands, marine). Following this pass, 704 citations remained, but many included comparative phylogeograph* only in the Keywords Plus field, which adds keywords according to titles of papers cited in a paper. I therefore made a second pass through the list with the goal of removing citations that did not provide indications of being conducted within a comparative framework, which resulted in 487 citations remaining. Finally, with subsequent searches of the database to harvest information reported here, I eliminated an additional 32 citations that were not concerned with continental biotas but I had not detected as such earlier, resulting in a final database of 455 citations distributed across biogeographic regions as summarized in Fig. 2.1*A*.

One caveat and one further explanation are warranted. First, I emphasize that this collection represents a sampling and not an exhaustive list of citations that focus on comparative phylogeography of continental biotas. As an example illustrating this, two of my own studies (Riddle et al., 2000; Riddle and Hafner, 2006) would not have been included had I not inserted them manually after the initial search. Nevertheless, I would argue that this collection of 455 citations is broadly representative both conceptually and geographically. Second, many studies that are included superficially could be construed to be single-taxon phylogeographic studies. However, one of my guidelines in determining which citations to keep and which to eliminate required evidence that what appeared to be a single-taxon study was embedded within a context of knowledge about the phylogeographies of codistributed taxa from prior studies, either explicitly as a motivator for choosing a focal taxon to study or in the context of inference about the resulting geographic pattern.

Studies Are Becoming Globally Representative

Prior reviews, most recently in 2010 (Hickerson and Meyer, 2008), have pointed to a Northern Hemisphere bias in the distribution of phylogeographic studies. Here, continental comparative phylogeographic studies can be summarized, approximately, into Northern, NA + PA + OR = 315, vs. Southern, AU + ET + NT = 209 (see Fig. 2.1*A* for abbreviations), representing a more equitable 60:40 ratio. Indeed, total numbers of studies have increased over time in both hemispheres, although less rapidly in each of the Southern and in the Oriental biogeographic regions (Fig. 2.1*B*).

Hotspots of Activity Are Increasing and Geographically Widespread

I recognize a continental comparative phylogeography hotspot as a definable geographic area and suite of taxa that either has become, or appears to be emerging, as one that attracts repeated studies, ordinarily wherein more recent studies explicitly reference known phylogeographic structure (Fig. 2.2). Although not all studies are dependent on such systems, they arguably offer added value for purposes of not only building a rich depiction of the pervasiveness of events in Earth history on a complex history of biological diversification and distributional dynamics, but they also have potential to generate novel insights into connections between biogeographic structure and related questions in ecology, evolutionary biology, and biodiversity conservation.

Here, I identify two advances since a 2010 review paper (Hickerson and Meyer, 2008) listed continental comparative phylogeography hotspots including Europe, Pacific Northwest, California, Neotropical rain forests, Baja California, East Africa, Southeast North America, and Australian Wet Tropics. First, I recognize a global proliferation of recognizable or emerg-

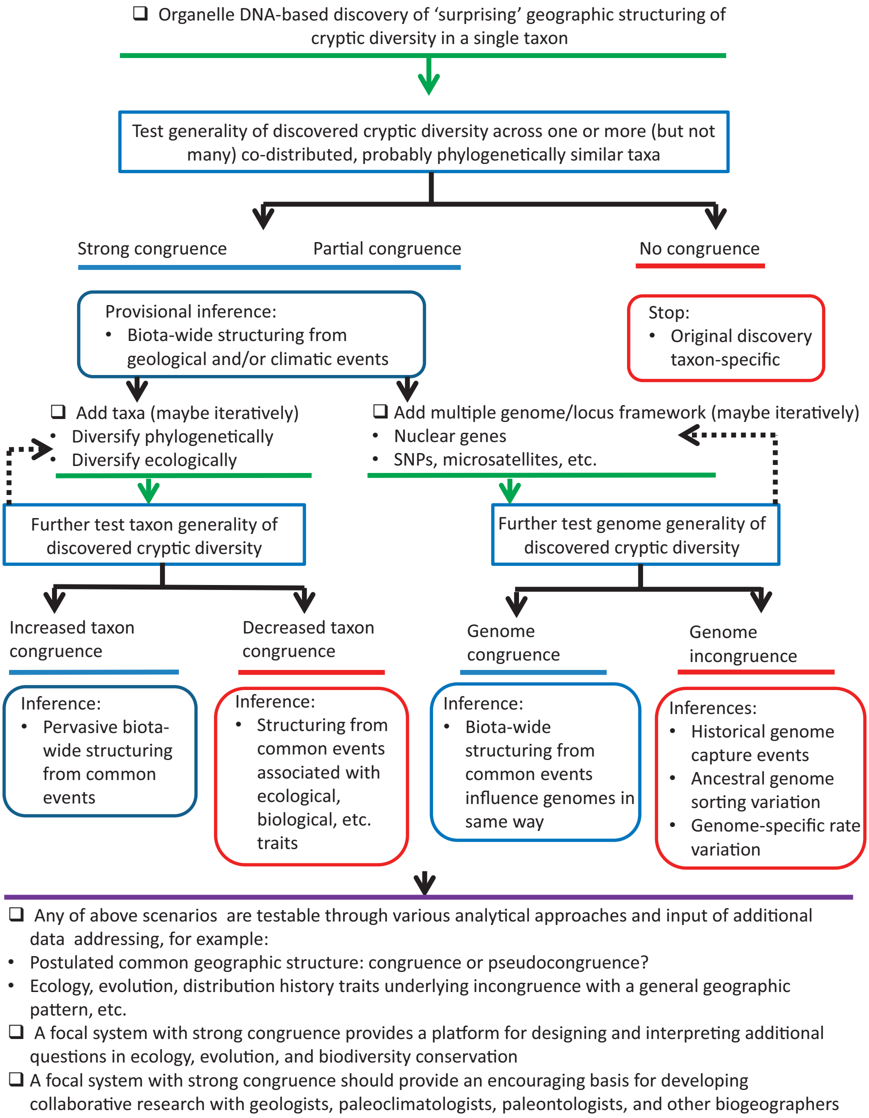


FIGURE 2.2 Anatomy and life history of an idealized focal system in continental comparative phylogeography. A focal system is defined here as one that generates repeated studies, each time adding new taxa and/or new genomes/genetic markers.

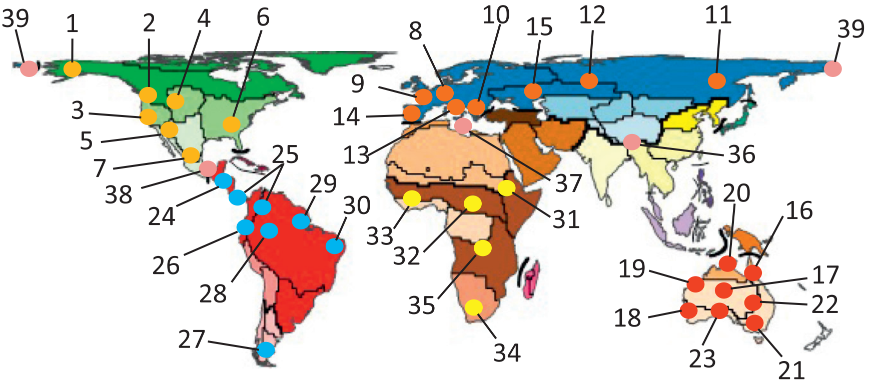


FIGURE 2.3 Continental comparative phylogeography hotspots or emerging hotspots (replicate studies using cumulatively increasing numbers of taxa) as inferred from scan of references database used here. Each set of colored dots represents a biogeographic region, except for those numbers that represent transition zones between regions: 36, OR-PA, Eastern Himalayas/Qunghai-Tibet Plateau; 37, ET-PA, Mediterranean Basin; 38, NT-NE, Middle America; 39, PA-NA, Beringea. All other localities identified as follows: 1, Alaska/East Beringea; 2, Pacific Northwest; 3, California; 4, Rocky Mountains; 5, Southwestern Aridlands; 6, Southeastern/Eastern; 7, Trans-Mexican Volcanic Belt; 8, European Alps/Carpathians; 9, Western Palearctic; 10, Balkan Peninsula; 11, Eastern Asia; 12, Eurasia; 13, Italian Peninsula; 14, Iberian Peninsula; 15, Central Asia/Ponto-Caspian; 16, Australian Wet Tropics; 17, Interior Deserts; 18, Southwestern Australia; 19, Pilbara Region; 20, Monsoonal Tropics; 21, Souteastern Australia/Tallaganda; 22, Endorheic Basins; 23, Southern Australia; 24a, Lower Central American lowlands; 24b, Lower Central American highlands; 25, Trans-Andean; 26, Andes; 27, Patagonia; 28, Amazon Rain Forest; 29, Caatinga/Cerrado; 30, Atlantic Rain Forest; 31, Eastern Arc (Afromontane) Mountains; 32, Central African Forest; 33, West African Forests/Savannas; 34, Cape Region; 35, Tropical African Forests/Savannas. Map modified with permission from Kreft and Jetz (2010).

ing hotspots distributed across biogeographic regions (Fig. 2.3), although none were detected in India or the Middle East, and the "hottest" area of activity in the Oriental region is transitional between it and the Palearctic in the vicinity of the Tibetan Plateau and Eastern Himalayas. Second, as with overall growth of studies in the Southern Hemisphere described above, my assessment indicates a more even distribution of established or emerging hotspots between hemispheres. Transition zones between biogeographic regions also are represented (Lower Central America between Neotropical and Nearctic; Mediterranean between Ethiopian and Palearctic; Qinghai-Tibetan Plateau between Oriental and Palearctic; and Beringia between Palearctic and Nearctic), signaling an emergent comparative phylogeographic approach to revealing the history of biotic interchange between regions that so captivated Simpson (1950), but with the added realization that such transition zones often cross active tectonic boundaries between continental plates.

The Topics Addressed Are Diverse

Here, I explore the frequencies of terminology generally associated with concepts in evolution, ecology, biogeography, and biodiversity conservation to assess their relative importance in continental comparative phylogeography, summarized under six categories designated a priori (Fig. 2.4): speciation, contact (between once isolated lineages), evolution/ecology (other than speciation), population level, geography, and response (to drivers such as geological and climatic changes). Very frequently used terms are Divergen* under speciation, Endemi* (somewhat surprisingly) under geography, and Refug* under response. However, a summary of the frequency of the six main categories (Fig. 2.4*G*) suggests that, although topics under speciation and response are most frequently mentioned, the distribution becomes more even among speciation, evolution/ecology, population level, and response if the overwhelming influences of Divergen*, Endemi*, and Refug* are removed (Fig. 2.4*H*). As with phylogeography generally, continental comparative phylogeography addresses a wide range of core concepts in modern biology in a geographic context.

Focal Taxa Are by Definition Geographically Codistributed, but Might Also Have an Ecological or Biological Relationship

The majority of studies in continental comparative phylogeography use taxa that are geographically codistributed without reference to additional criteria such as specified ecological relationships, for example, contrasting habitat specializations, but several studies of plant comparative phylogeography in the European Alps provide a good example of doing so, with plants identified as specialists on either calcareous or silicicolous substrates (Tribsch and Schonswetter, 2003). One interesting but not yet common approach is to investigate the historical assembly of taxa associated ecologically into trophic interactions and food webs (Garrick et al., 2012). A reasonable subset of studies (NA = 8; PA = 16; AU = 1; NT = 4; ET = 5) goes further and investigates phylogeographic congruence, or lack thereof, between species that have an obligatory relationship such as hosts

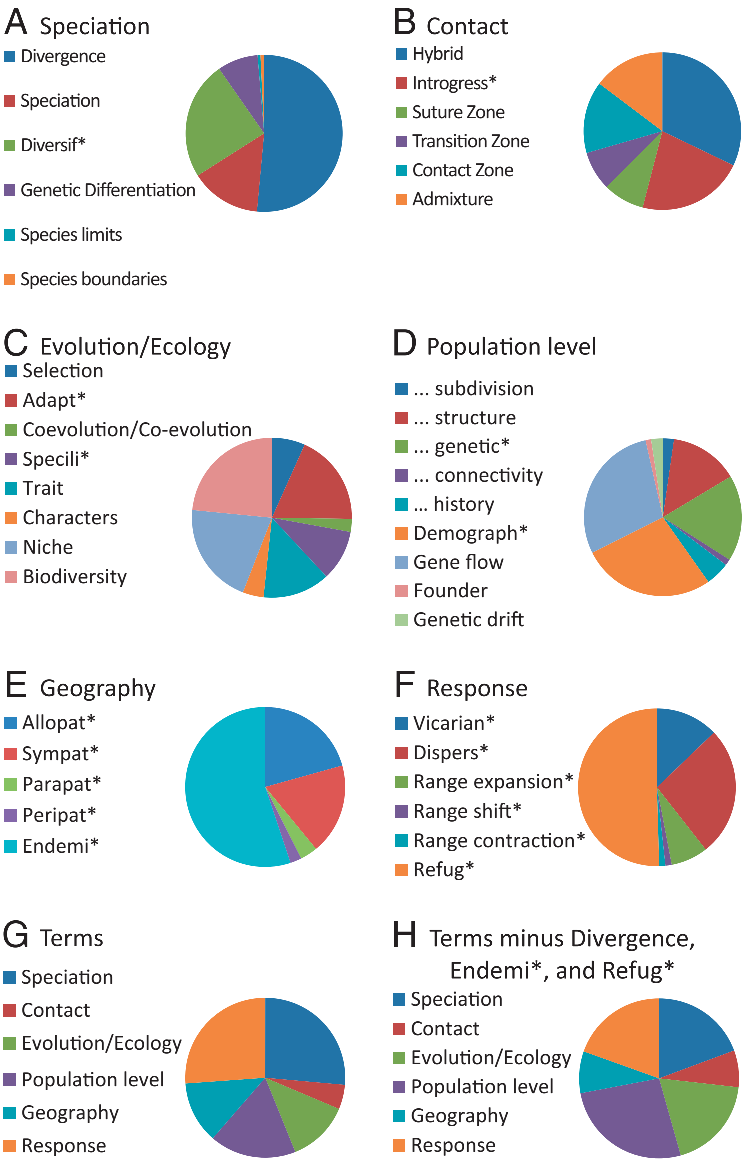


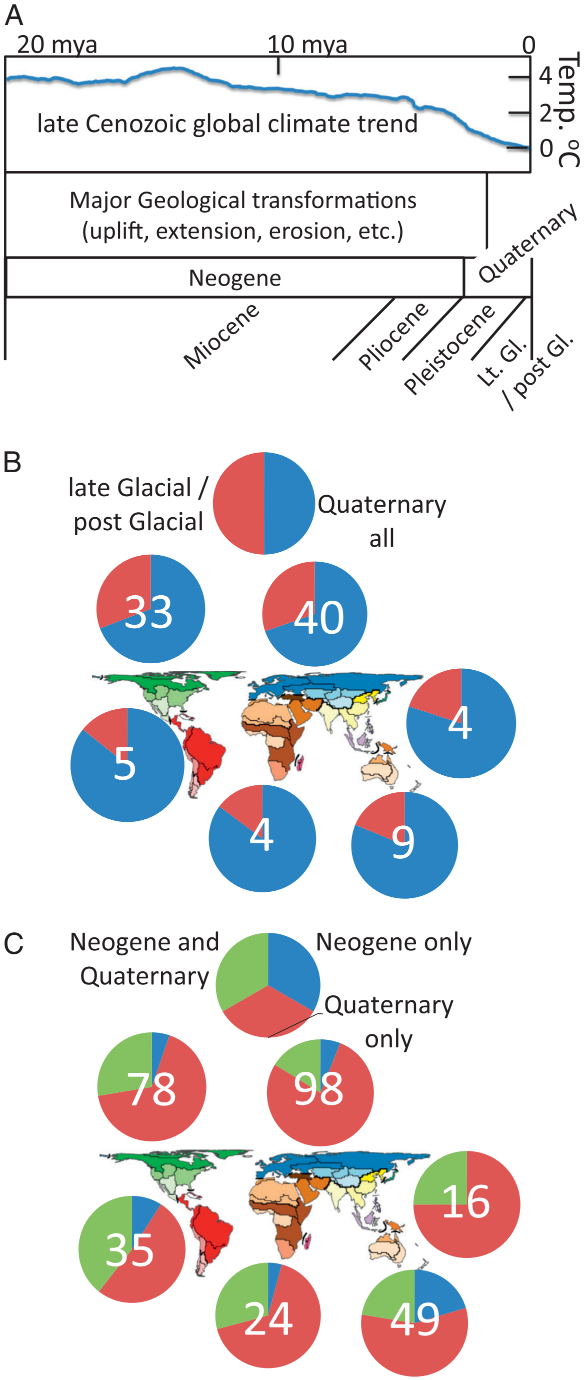
FIGURE 2.4 Frequency of terms generated through searches of references database used here, embedded within six categories (*A*–*F*) considered here to represent a range of primary foci of continental comparative phylogeography. Those followed by an asterisk denote root terms that find any variation (e.g., allopat* will find allopatric, allopatry, allopatrically). The total frequencies of the six primary foci (*G* and *H*) differ because the latter was generated after removing three terms of overriding frequency, showing a more even distribution across the primary foci.

and parasites or predators (Tavares et al., 2015), or plants and pollinators (Espíndola and Alvarez, 2011): a particularly well-studied system involves oak gall wasps and their parasitoids in the Palearctic region (Lohse et al., 2012).

Temporal Scope Spans Neogene and Quaternary Periods

In general, the Neogene and early Quaternary periods are regarded as times of major geological activity on a number of continents, including uplift, extension, erosion, incision of large sections of continents, and episodes of marine transgressions (Fig. 2.5*A*). Major episodes of geological transformation are more localized to the Neogene in, for example, the western Nearctic and Neotropical regions, but extend into the Quaternary in the Oriental-Palearctic transition area (Yue et al., 2012). Beginning in the later Neogene and accelerating into and throughout the Quaternary, global temperatures decreased and the oscillations between glacial and interglacial climatic regimes increased in magnitude (Zachos et al., 2001). As such, continental comparative phylogeography studies that reference processes that could be driving divergence and range dynamics during the Neogene and early Quaternary are likely to consider a combination of geological and climatic processes, whereas those that reference the later Quaternary, and more specifically the late glacial (e.g., Last Glacial Maximum or LGM) into postglacial time frames are more likely to emphasize climate changes alone, considering the geological template to be for the most part already established by this time (Fig. 2.5*B*).

Here, I find that a number of studies reference a late glacial into postglacial time frame specifically (Fig. 2.5*C*), but although these represent an appreciable portion of all studies referencing the Quaternary, there are many additional studies centered in a Quaternary time frame that do not adhere to this relatively recent and temporally narrow period. I also find that out of the larger number of studies that span either Neogene, or Quaternary, or both (Fig. 2.5*C*), those limited to the Quaternary represent the highest frequency in all biogeographic regions, but each region also contains roughly a quarter (more in AU and NT, fewer in PA) studies that reference either the Neogene or both Neogene and Quaternary of time frames. These results support a search indicating that, although most studies suggest that taxa are responding to climatic change as a driver, a number of them also reference geology as a potential driver (Climat* = 638 total hits; Geol* = 103 total hits), corroborating the generality of the signature of Neogene divergence within "intraspecific and closely related species" foreseen decades earlier by, for example, Maxson and Roberts (1984).


A
20 mya
10 mya
0
4
2
0
Temp. °C
late Cenozoic global climate trend
Major Geological transformations
(uplift, extension, erosion, etc.)
Neogene
Quaternary
Miocene
Pliocene
Pleistocene
Lt. Gl.
/ post Gl.
B
late Glacial /
post Glacial
Quaternary
all
33
40
5
4
4
9
C
Neogene and
Quaternary
Neogene only
Quaternary
only
78
98
35
16
24
49

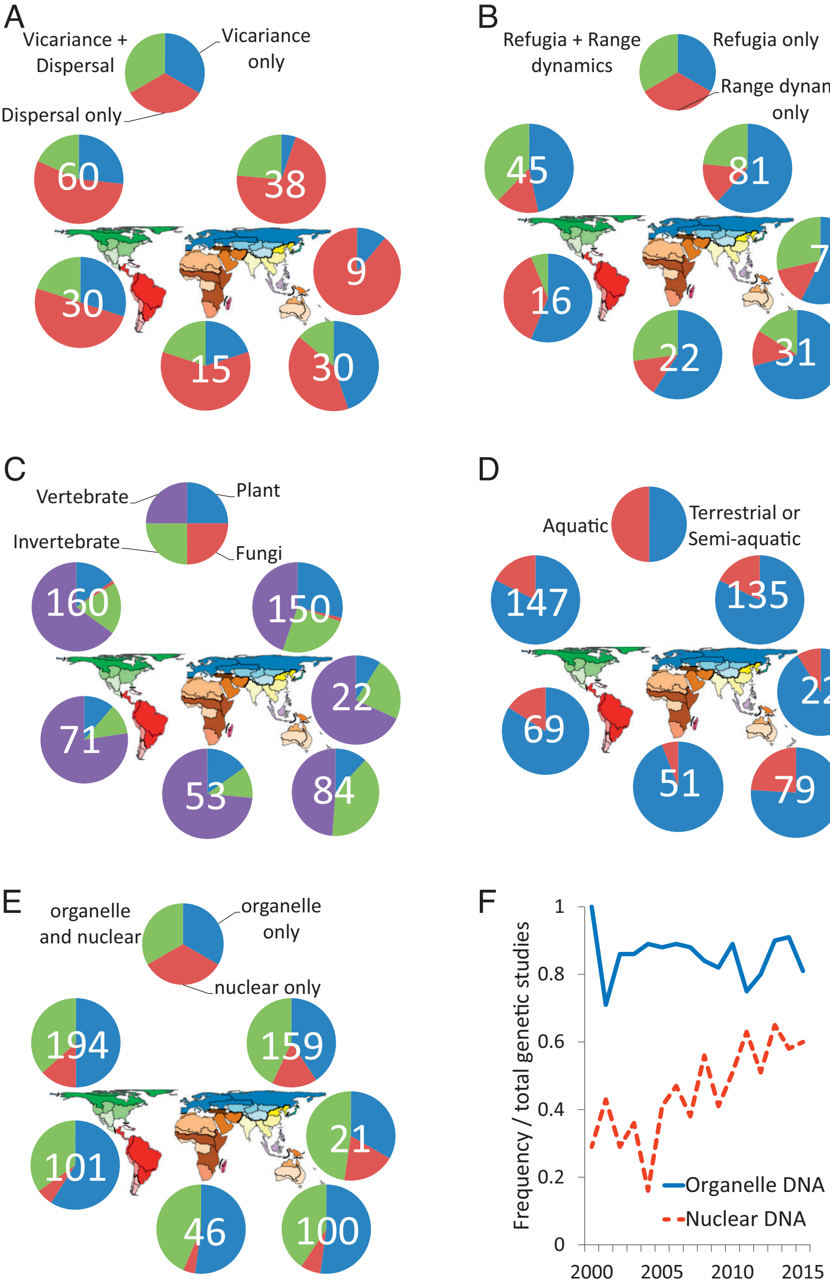
By Spanning Neogene and Quaternary Periods, an Array of Biotic Responses to Geological and Climatic Events in Earth History Are Addressed

Another way to examine the role of Earth history on continental comparative phylogeographic structure is to use search terms directly referencing the full range of possible biotic responses originally envisioned within separate research paradigms, including vicariance and dispersal, e.g., Rosen (1978) and Darlington (1959), and refugia and range dynamics (expansion, contraction, shifting), e.g., Adams (1905) and Haffer (1969). First, vicariance or dispersal or both vary in frequency between regions (Fig. 2.6*A*); notably, vicariance appears most often in studies of Australian biotas and less in the Palearctic than in all other regions except the Oriental (although the sample size there is small). Within a late Quaternary time frame, we should expect the concepts of glacial refugia and postglacial range expansion to be addressed together in a number of studies, and out of a total tabulation of papers that address one or the other or both of these concepts that connection holds true in each region, primarily in the Nearctic but not as frequently in the Neotropical region (Fig. 2.6*B*).

Taxa Are Biased Toward Vertebrates and Terrestrial or Semiaquatic Organisms

Although continental comparative phylogeography began (Bermingham and Avise, 1986) with an investigation of four codistributed freshwater fish taxa, to date far more studies have emphasized terrestrial or semiterrestrial taxa, and this is true across regions (Fig. 2.6*C*). The regional distribution of taxa from five categories (vertebrates, inver-

FIGURE 2.5 The time frames explored empirically by continental comparative phylogeography. Terms used to describe geological time frames differ among studies (*A*) but are subsumed for purposes of searching the reference database as Neogene (= Miocene + Pliocene) and Quaternary (= Pleistocene + Late Glacial–Postglacial + LGM + Holocene). Late Cenozoic global climate trend traced from an online summary (www.columbia.edu/~mhs119/Sensitivity+SL+CO2) of a well-known figure (Zachos et al., 2001). Studies using terms describing time frames (total numbers tallied for each region overlaid on pie charts) summarized as (*B*) late Glacial–Postglacial vs. Quaternary all, and (*C*) Neogene only vs. Quaternary only vs. Neogene and Quaternary. Map modified with permission from Kreft and Jetz (2010).


A
Vicariance + Dispersal
Vicariance only
Dispersal only
60
38
30
9
15
30
B
Refugia + Range dynamics
Refugia only
Range dynamics only
45
81
16
7
22
31
C
Vertebrate
Plant
Invertebrate
Fungi
160
150
71
22
53
84
D
Aquatic
Terrestrial or Semi-aquatic
147
135
69
22
51
79
E
organelle and nuclear
organelle only
nuclear only
194
159
101
21
46
100
F
Frequency / total genetic studies
1
0.8
0.6
0.4
0.2
0
2000
2005
2010
2015
Organelle DNA
Nuclear DNA

tebrates, plants, fungi, microbes) indicates an appreciable bias in each region toward vertebrates (Fig. 2.6*D*), although plants and invertebrates are also relatively well represented in the Palearctic and invertebrates in the Australian. Very few studies have focused on fungi, and studies of microbes appear to be nonexistent or nearly so.

The Use of Nuclear DNA and Multilocus Data Is Increasing, but Organelle DNA Is Not Going Away

Controversy arose a few years ago with regard to the continued efficacy and importance of organelle (mitochondrial, chloroplast) genomes in phylogeography as theoretical and practical considerations argue for greater use of multilocus data from the nuclear genome (Zink and Barrowclough, 2008; Barrowclough and Zink, 2009; Edwards and Bensch, 2009). A recent argument (Bowen et al., 2014) was made for the continuing importance of mtDNA in marine phylogeography even as multilocus studies grow in frequency and importance. Here, I find that a relatively small number of continental comparative phylogeographers are yet to use nuclear DNA to the exclusion of organelle DNA (Fig. 2.6*E*), particularly in the more southern Neotropical, Ethiopian, and Australian regions.

FIGURE 2.6 (*A* and *B*) Categories referenced frequently in continental comparative phylogeography of biotic responses to geological and climatic events (total numbers tallied for each region overlaid on pie charts). Studies divided to contrast terms (*A*) vicariance only vs. dispersal only vs. vicariance + dispersal, and (*B*) refugia only vs. range dynamics only vs. refugia + range dynamics. Range dynamics includes references to range expansion, range contraction, and range shift and variants of each. (*C* and *D*) Frequencies of organisms, summarized into taxonomic groups and medium inhabited, studied in continental comparative phylogeography according to search of the database used in this study. Taxa contrasted (*C*) as vertebrate vs. invertebrate vs. plants vs. fungi; and medium (*D*) contrasted as terrestrial or semiaquatic vs. aquatic organisms. (*E* and *F*) Frequencies of genomes used in continental comparative phylogeography studies. Organelle DNA (including mitochondrial DNA and chloroplast DNA, also searched as mtDNA and cpDNA) contrasted (*E*) with nuclear DNA (including all types of markers) as organelle only vs. nuclear only vs. organelle and nuclear. The frequency of studies (*F*) using markers from each genome plotted through time (numbers for 2015 not complete because database was downloaded from Web of Science in December 2015). Map modified with permission from Kreft and Jetz (2010).

However, although a higher frequency of studies use some combination of both nuclear and organelle data, a (perhaps) surprisingly high number still rely on organelle genome data alone. Indeed, the use of organelle DNA does not appear to have decreased into current times as use of nuclear DNA has clearly increased rapidly over the last decade or more (Fig. 2.6*F*).

There likely is a combination of reasons for the ongoing popularity of organelle DNA in continental comparative phylogeographic studies. First, ease and cost of data generation probably remain central drivers, particularly when a principal objective remains the goal of sampling as many geographically distributed populations and individuals as possible across multiple codistributed taxa. Second, organelle DNA, particularly mitochondrial DNA in animals, still likely delivers a very strong and heuristically valuable first approximation of geographic genetic architecture, providing a reasonable baseline for hypothesis testing when a greater diversity of genetic data from across genomes and independently sampled portions of the nuclear genome become available. Third, for purposes of contrasting and comparing signals of geographic structure across divergent taxa, perhaps the simplicity and strong geographic signal generated by organelle DNA are still something of an advantage, considering for example the number of such studies that continue to be subjected to analysis within hierarchical approximate Bayesian computation (hABC) frameworks (Dolman and Joseph, 2012; Smith et al., 2014b).

Species Distribution and Ecological Niche Models Are Promising, but Still Not Widely Used in Continental Comparative Phylogeography

Species distribution modeling or ecological niche modeling was beginning to get introduced into phylogeography nearly a decade ago (Richards et al., 2007; Waltari et al., 2007). However, although being used commonly across phylogeography (a Web of Science search returned 419 hits for phylogeograph* and “species distribution model*” or “ecological niche model*”), that use still is somewhat limited in continental comparative phylogeography (NA = 15; PA = 8; AU = 4; NT = 7; ET = 1; OR = 2). I would expect its use in continental comparative phylogeography to grow with appreciation for developing independent, testable hypotheses for sorting alternative explanations of predicted responses (both historical and future) across codistributed taxa; potential to connect climatic and other habitat variables of specific ecological importance to target taxa; and potential to accurately model distributional responses under past and future climate models (Reilly et al., 2015).

INTO THE FUTURE

Comparative phylogeography has enjoyed a large measure of the success that eluded Wallace in 1880 when trying to understand biogeographic patterns and processes in continental biotas. Uncovering a previously unknown and unpredicted level of cryptic diversity in animals and plants and applying sophisticated analytical approaches to mapping phylogenetic and population affinities onto geography have made inroads into the questions posed by the 20th century biogeographers, ecologists, and evolutionary biologists. One ongoing challenge and promise is to continue developing connections to related disciplines that usually lie at different spatial, temporal, and pattern/process scales (Fig. 2.7). In doing so, continental comparative phylogeography will continue to be transformed by ongoing incorporation of, for example, species distribution modeling,

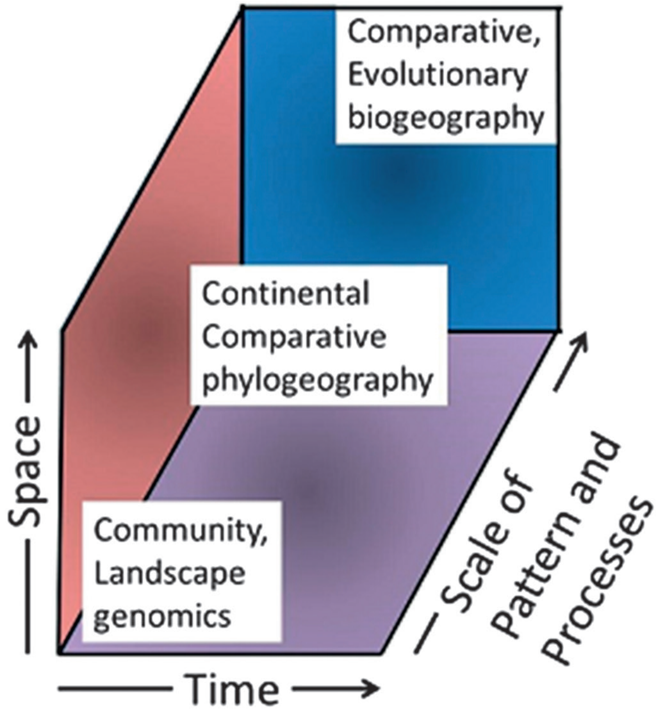


FIGURE 2.7 Positioning of continental comparative phylogeography within a depiction of related disciplines across spatial, temporal, and scale of pattern/process hierarchies.

the genomics revolution, geospatial statistical platforms, and an explicit hypothesis testing context (Dawson, 2014). Here are a few predictions of some future directions and developments.

Continental Comparative Phylogeography Will Continue Contributing to Erosion of the Simple Vicariance vs. Dispersal Paradigm

Wallace was not wrong about the complexity of continental biotas, and comparative phylogeography continues to reveal that this is true in a temporal and spatial context. Pseudocongruence is a recognized pattern in biogeography (Donoghue and Moore, 2003), and in a temporal context, describes the phenomenon of a suite of taxa that superficially appear to have responded in concert to a putative barrier (e.g., a single, simultaneous vicariant event) or via mass dispersal to the opening of a dispersal route between previously isolated regions (also known by the useful but underused term "geodispersal"), but in reality were either isolated or dispersed in geographically similar fashions but at different times. One of the more detailed examples of pseudocongruence to date (Smith et al., 2014b) studied 27 lineages of Neotropical birds that are codistributed with isolated populations on either side of the Andean mountains. Mitochondrial DNA variation was analyzed in all individuals, and a subset was further analyzed for multilocus variation across more than 100 loci. In both datasets, hABC analyses demonstrated strong signals of pseudocongruence across the Andes, the Isthmus of Panama, and three large Neotropical rivers (Amazon, Madeira, and Negro). Trans-Andean divergence was estimated to have happened between 9 and 29 times during the Pleistocene and was attributed to dispersal rather than vicariance from earlier Andean uplift. Ecology played a role as well, in which forest canopy species, considered more capable of long-distance dispersal than understory birds, showed generally more recent times of separation of lineages across barriers.

Studies arguing for or providing evidence of pseudocongruence are becoming widespread across biogeographic regions, but several that can be highlighted in addition to the above include the following: North American southwestern aridlands (Riddle and Hafner, 2006; Leaché et al., 2007), southeastern North America (Soltis et al., 2006), lower Central America (Bagley and Johnson, 2014), Beringian crossing between the Palearctic and Nearctic (Hope et al., 2013), and Australian wet tropics (Moreau et al., 2015).

Implications of this growing number of examples of pseudocongruence are several, should the historically default model of simple vicariance eventually be realized to be a relatively rare event in the history of continental biotas. First, given enough data (large number of taxa and enough genetic data), it should be possible in many cases to resolve ecological trait differences that are predictive of the isolation and dispersal histories of suites of taxa within rigorous geological and climatic contexts (Smith et al., 2014b). Second, investigators might turn attention to the notion that much of diversification history on continents might be explained through a model of temporally expanded "species pumps" across geographically stable barriers. In this case, a compelling question will be to address the biotic and abiotic features that combine to keep earlier isolated taxon pairs still isolated as more recent rounds of barrier formation and erosion close and open opportunities for cross-barrier dispersal, as might happen, for example, through repeated cycles of Quaternary glacial–interglacial climate oscillations (Pyron and Burbrink, 2007).

More Ecological, Evolutionary, Life History, and Physiological Traits Will Be Incorporated into Continental Comparative Phylogeography

Although some of these properties have focused the attention of continental comparative phylogeographers, perhaps from the beginning of the approach, recent studies are taking advantage of, for example, distribution modeling and massively high throughput sequencing to refine the level of resolution. A recent study (Jezkova et al., 2015) took advantage of ecological differences between two codistributed species of kangaroo rat (genus *Dipodomys*) in western North American deserts to ask whether degree of substrate specialization led to a predictable difference in demographic parameters along a postglacial range expansion front. A habitat specialist, *Dipodomys deserti*, was predicted to lose more genetic diversity than the generalist, *Dipodomys merriami*, from south [where species distribution models (SDMs) supported locations of late glacial refugia in both species] to north. These predictions were supported by both mtDNA and double-digest restriction site associated DNA (ddRADseq) SNP datasets. Other small desert vertebrates and invertebrates have also expanded, as modeled under SDMs and in some cases shown with genetic data (Graham et al., 2013; Jezkova et al., 2016) along this same expansion front, providing the intriguing potential to begin to sort among other ecological (e.g., substrate specialists) and physiological (e.g., endotherms vs. ectotherms; homeothermic vs. heterothermic endotherms) traits that are predictive with regard to distributional and demographic properties.

Note that the example above also illustrates the earlier point regarding the value of designing studies within well-studied comparative phylogeography hotspots (Fig. 2.2). Reinforcing this observation within the same system, one recent study (Mott et al., 2015) used the backbone general phylogeographic pattern to gain understanding into the historical development of a complex life history structure between two species of ants in the genus *Pogonomyrmex*. Another (Schield et al., 2015) took advantage of combining mtDNA and ddRADseq SNPs to reveal both a deep phylogeographic break (mtDNA) and recent gene flow (SNPs) across a transition zone that is coincident with at least 14 additional cryptic taxon pairs of mammals, birds, and herps (Riddle and Hafner, 2006; Pyron and Burbrink, 2007), suggesting the prospect of studying ongoing speciation processes across a transition zone (that almost certainly contains a rich history of temporal pseudocongruence) at a community scale using similar genomic tools, dramatically building on Remington's (1968) concept of suture zones.

Continental Comparative Phylogeographers Will Work More Closely with Geologists, Paleoclimatologists, and Paleontologists

Revolutions in geology have historically signaled revolutions in biogeography: most notably, the vicariance biogeography paradigm that followed closely on the widespread acceptance of continental drift under the plate tectonics framework (Lomolino et al., 2010). However, Earth scientists and biogeographers often have worked independently rather than in a concerted fashion to address the inextricable links between Earth and biotic histories. Continental comparative phylogeographers have the potential to integrate more explicitly with Earth scientists to productive ends.

A good example comes from the robust phylogeographic signal of cryptic divergence between lineages distributed in the southern vs. northern parts of the Baja California Peninsula (Riddle and Hafner, 2006). A series of mid-peninsular seaways have been postulated as isolating barriers to account for these phylogeographic splits, even though geological evidence supporting existence of seaways had been largely absent. In this case, the comparative phylogeographic signal motivated geologists to investigate the plausibility of mid-peninsular seaways, concluding that regional geology is consistent with Miocene but not Plio-Pleistocene seaways (Dolby et al., 2015).

When a rich fossil record is available, it should be possible to use comparative phylogeographic and fossil information to test diversification hypotheses. For example, the Neogene fossil record for rodent taxa in western North America has been used to establish a relationship between topographic complexity and diversity (Badgley et al., 2014): phylogeographic evidence can provide an independent approach to testing a variety of hypotheses available to account for this pattern given the high standing diversity of rodent lineages in extant biotas covering this region.

More Patterns and Inferred Processes Derived Through Continental Comparative Phylogeographic Analyses Will Be Incorporated into Biodiversity Conservation Plans

How do we try to maximize retention of biodiversity and the ecological and evolutionary processes that underlie the origination and sustainability of that diversity in a rapidly changing world? Comparative phylogeography has been envisioned as having an important role in biodiversity conservation nearly since its beginnings (Moritz and Faith, 1998) and continues to contribute valuable information for conservation purposes. The breadth of potential applications is enormous, for example, deriving quantitative measures of phylogenetic diversity and phylogenetic endemism in biodiversity hotspots on either end of an entire continent (Laity et al., 2015). However, value on much smaller spatial scales is also possible, illustrated by the use of phylogeographic data from 12 species of mammals, herps, and invertebrates to derive estimates of biodiversity hotspots (genetic diversity and genetic divergence) within the Mojave and Sonoran deserts in southwestern North America (Wood et al., 2013). The maps derived from this exercise should become important tools for resource managers and planners making daily decisions in the face of immediate threats to biodiversity from solar farm development, expanding urban developments, and off-road vehicle recreationists.

Continental Comparative Phylogeography Will Have an Important Role in Development of a Broadly Construed, Highly Integrative Evolutionary Biogeography

Comparative phylogeography has been used to illustrate a multifaceted, stepwise research protocol (Riddle and Hafner, 2006), and it will continue to become more firmly embedded within a highly integrative approach to reconstructing the diversification and distributional dynamics of continental biotas that draws from the broad spectrum of analytical resources available in biogeography, systematics, population genetics, and ecology to create new bridges between, for example, ecology and evolutionary biology (Marske et al., 2013). An "evolutionary biogeography" has been imagined (Morrone, 2009) that, rather than continuing a long history of divisiveness between practitioners of different approaches to biogeography, placed phylogeography into a multifaceted, stepwise protocol designed to create an emergent "geobiotic scenario." This approach identifies natural biogeographic units and analyzes the biotic relationships among them; they are next analyzed for the full suite of taxon histories, including a teasing apart of the natural pseudocongruence that is a likely result of a long history of episodes of dispersal and vicariance. A fully realized program of this nature will require innovation in how comparative phylogeography bridges and links traditionally independent research programs in a phylogenetics-based historical biogeography with a population genetics-based landscape and community genetics (Fig. 2.7).

A full suite of additional exciting prospects for the advancement of comparative phylogeography can be found in other contributions to this volume. Perhaps if Alfred Russel Wallace had access to the concepts and methods enjoyed by modern comparative phylogeographers, he would not have turned his focus quite as energetically toward islands to gain greater understanding of the geographic structure and history of continental biotas! Another review a decade from now will undoubtedly report on a vastly transformed continental comparative phylogeography.

ACKNOWLEDGMENTS

I thank Profs. J. C. Avise, B. W. Bowen, and F. J. Ayala for organizing and inviting me to participate in the extraordinary colloquium that generated the basic ideas incorporated into this chapter. I thank Prof. H. Kreft who provided encouragement for me to use the figure of biogeographic regions that he and Prof. W. Jetz published 6 years ago.

3

Inferring Responses to Climate Dynamics from Historical Demography in Neotropical Forest Lizards

IVAN PRATES,[*†**] ALEXANDER T. XUE,[*†]
JASON L. BROWN,[*‡] DIEGO F. ALVARADO-SERRANO,[*§]
MIGUEL T. RODRIGUES,[‖] MICHAEL J. HICKERSON,[*†#]
AND ANA C. CARNAVAL[*†]

We apply a comparative framework to test for concerted demographic changes in response to climate shifts in the Neotropical lowland forests, learning from the past to inform projections of the future. Using reduced genomic (SNP) data from three lizard species codistributed in Amazonia and the Atlantic Forest (*Anolis punctatus*, *Anolis ortonii*, and *Polychrus marmoratus*), we first reconstruct former population history and test for assemblage-level responses to cycles of moisture transport recently implicated in changes of forest distribution during the Late Quaternary. We find support for population shifts within the time frame of inferred precipitation fluctuations (the last 250,000 years) but detect idiosyncratic responses across species and uniformity of within-species responses across forest regions. These results are incongruent with expectations of concerted population expansion in response to increased rainfall and fail to detect out-of-phase demographic syndromes (expansions vs. contractions) across forest regions. Using reduced genomic data to infer

---

[*]Department of Biology, City College of New York, New York, NY 10031; [†]Department of Biology, Graduate Center, City University of New York, New York, NY 10016; [‡]Cooperative Wildlife Research Laboratory, Department of Zoology, Southern Illinois University, Carbondale, IL 62901; [§]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109; [‖]Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP 05508-090, Brazil; and [#]Division of Invertebrate Zoology, American Museum of Natural History, New York, NY 10024. [**]To whom correspondence should be addressed. Email: ivanprates@gmail.com.

species-specific demographical parameters, we then model the plausible spatial distribution of genetic diversity in the Atlantic Forest into future climates (2080) under a medium carbon emission trajectory. The models forecast very distinct trajectories for the lizard species, reflecting unique estimated population densities and dispersal abilities. Ecological and demographic constraints seemingly lead to distinct and asynchronous responses to climatic regimes in the tropics, even among similarly distributed taxa. Incorporating such constraints is key to improve modeling of the distribution of biodiversity in the past and future.

Nearly three decades have passed since Avise (Avise et al., 1987; Avise, 2000) first inspired tropical biologists to study community-level responses to regional changes through the lens of population genetics. However, despite the high species richness, endemism, and genetic structure observed in tropical regions (Moritz et al., 2000), comparative phylogeographic approaches are relatively unexplored in the tropics. To date, most available phylogeographic comparisons report spatial patterns of organelle DNA variation, primarily mitochondrial, across a handful of tropical taxa (Schneider et al., 1998; Pastorini et al., 2003; Nicolas et al., 2008; Carnaval et al., 2009; Moritz et al., 2009; Avila-Pires et al., 2012; Dolman and Joseph, 2012). Although addressing different regions and organisms, these studies have centered on two major topics: the role of climatic discontinuities (Schneider et al., 1998; Dolman and Joseph, 2012) or rivers (Pastorini et al., 2003; Avila-Pires et al., 2012) as barriers to dispersal and the influence of glacial–interglacial temperature oscillations on species ranges, a topic originally explored in studies of Northern Hemisphere taxa (Hewitt, 2000) but widely embraced by tropical phylogeographers worldwide (Nicolas et al., 2008; Carnaval et al., 2009; Moritz et al., 2009). This first generation of comparative studies indicated pervasive geographic structure of mitochondrial lineages (Carnaval et al., 2009; Moritz et al., 2009; Avila-Pires et al., 2012) and either congruent spatial patterns of genetic diversity across species of broadly similar ecologies (Schneider et al., 1998; Carnaval et al., 2009) or differences in ways largely expected, given their life history traits (Crawford et al., 2007; Nicolas et al., 2008). However, as tropical phylogeographers began to sample loci (Lorenzen et al., 2012; d'Horta et al., 2013; Demos et al., 2014) and taxa (Moritz et al., 2009; Dolman and Joseph, 2012; Bagley and Johnson, 2014) more widely, reports of asynchronous lineage diversification and distinct demographic syndromes among codistributed taxa (such as population expansions and contractions) became commonplace.

If idiosyncratic responses to common landscape changes characterize ecological communities in the tropics (Colwell and Rangel, 2010; Smith et al., 2014b) and beyond (Papadopoulou and Knowles, Chapter 8, this volume; Zamudio et al., Chapter 11, this volume), their incorporation in phylogeographic methods can significantly improve our understanding of the impacts of former environmental shifts on regional species pools. Methods that build upon coalescent theory to account for historical heterogeneity across taxa under a single statistical framework (Hickerson et al., 2007; Xue and Hickerson, 2015) provide increased power to test concerted demographic responses at the level of ecological assemblages. These highly flexible approaches have profited from novel bench and analytical methods that build on genomic or reduced genome data, resulting in improved estimates of demographic parameters and histories (Pinto et al., 2014; Streicher et al., 2014; Rittmeyer and Austin, 2015). Although still rare, studies that combine coalescent-based demographic simulations and broader genomic sampling across taxa are unquestionably deepening our understanding of how tropical diversity came to be distributed (Smith et al., 2014b).

Improved knowledge of species responses to climate and landscape shifts will also strengthen forecasts of community responses to future global changes. Going beyond predictions of geographic ranges under future climates (Hijmans and Graham, 2006; Sinclair et al., 2010), molecular data can guide the parameterization of plausible species-specific models of the distribution of genetic diversity, and hence evolutionary potential, under predicted scenarios of global change (Brown et al., 2016). This as yet underexplored application of comparative phylogeography is especially promising and timely given today's threats to tropical ecosystems (Moritz et al., 2000).

Using reduced genome data of codistributed Neotropical forest taxa, we apply a comparative phylogeographic framework to address two major goals. First, we ask whether present-day patterns of genomic diversity reflect congruent, multitaxa responses to historical environmental shifts documented within and across rainforest systems (Cheng et al., 2013). Then, using the distribution of suitable habitats and the molecular data to gain insight about species-specific demographical parameters (dispersal rates and population densities over time), we develop plausible models of the spatial distribution of genetic diversity under future climates. To do so, we analyze SNP data from *Anolis punctatus* and *Anolis ortonii* (Dactyloidae) and from *Polychrus marmoratus* (Polychrotidae). These codistributed arboreal lizard species occur in both Amazonia and the coastal Atlantic Forest of South America but are absent from the intervening open and drier scrublands and savannahs of the Caatinga and Cerrado domains.

Instead of examining the effects of temperature fluctuations associated with glacial–interglacial cycles commonly addressed by phylogeographic studies, we focus our historical study on the influence of a recently hypothesized spatial dynamics of moisture transport in tropical South

America over the last 250,000 years (Cheng et al., 2013). Paleoecological studies indicate that precipitation patterns have been the major constraining element of tropical rainforest distribution during the Quaternary (Ledru et al., 2015). Through a compilation of proxy data based largely on geochemical analyses of cave deposits, it has been proposed that levels of precipitation in the eastern portion of Amazonia dynamically matched those in the northern portion of the Atlantic Forest, whereas precipitation patterns in western Amazonia have been linked to those in the southern Atlantic Forest (Cheng et al., 2013). These two macroclimatic systems presumably have been acting as a dipole, with multiple cycles of inversion of precipitation regimes between regions that followed the precession component of Earth's orbital cycles. For instance, although decreased rainfall has been inferred to impact eastern Amazonia and the northern Atlantic Forest ~20,000 ya, an opposite pattern (increased precipitation) is recorded for western Amazonia and the southern Atlantic Forest at that time (Cheng et al., 2013). The possibility that environmental drivers other than glacial–interglacial temperature oscillations have impacted tropical biotas worldwide has important implications for phylogeographic investigations, suggesting that at least some regions in the tropics are subjected to processes fundamentally different from those in the better-known temperate regions. We ask whether these proposed out-of-phase precipitation dynamics led to contrasting phylogeographic patterns between rainforest regions, testing recently proposed hypotheses about linked biological responses to synchronous but contrasting climatic regimes in South American forests.

Our approach tests for temporal congruence of demographic trends, using coalescent simulations under a hierarchical demographic model in combination with approximate Bayesian computation (ABC) to test for synchronous shifts in population sizes (Xue and Hickerson, 2015). We examine the timing and magnitude of population changes in different forest regions occupied by a single species as well as codistributed species within the same region. Given the geographic distribution of the three target lizard species, which are largely absent from the southern Atlantic Forest (*A. punctatus* and *P. marmoratus* occupy a limited extent of the southern Atlantic Forest in the states of Rio de Janeiro and São Paulo), we focus on concerted demographical shifts between western Amazonia, eastern Amazonia, and the northern Atlantic Forest.

Further exploring the molecular data and their potential contribution to conservation, we then expand on a recently proposed framework (Brown et al., 2016) to model the plausible distribution of genomic diversity of the two focal *Anolis* species under future climates. To illustrate this approach, we use a 2080 climate model based on the Representative Concentration Pathway (RCP) 6.0 emission scenario, which has been used extensively to represent a medium carbon emission trajectory (Van der Linden and Mitchell, 2009) and results in mid-range estimates of average global changes (IPCC, 2015). For this analysis, we infer species-specific demographic parameters from a combination of species distribution models, forward-in-time demographic simulations, and backward-in-time coalescent simulations that are compared with the empirical SNP data. To avoid potential modeling limitations imposed by Amazonia's sparsely distributed weather stations (Waltari et al., 2014), we restrict our forecasts to the highly endangered Atlantic Forest hotspot (Carnaval et al., 2009).

RESULTS AND DISCUSSION

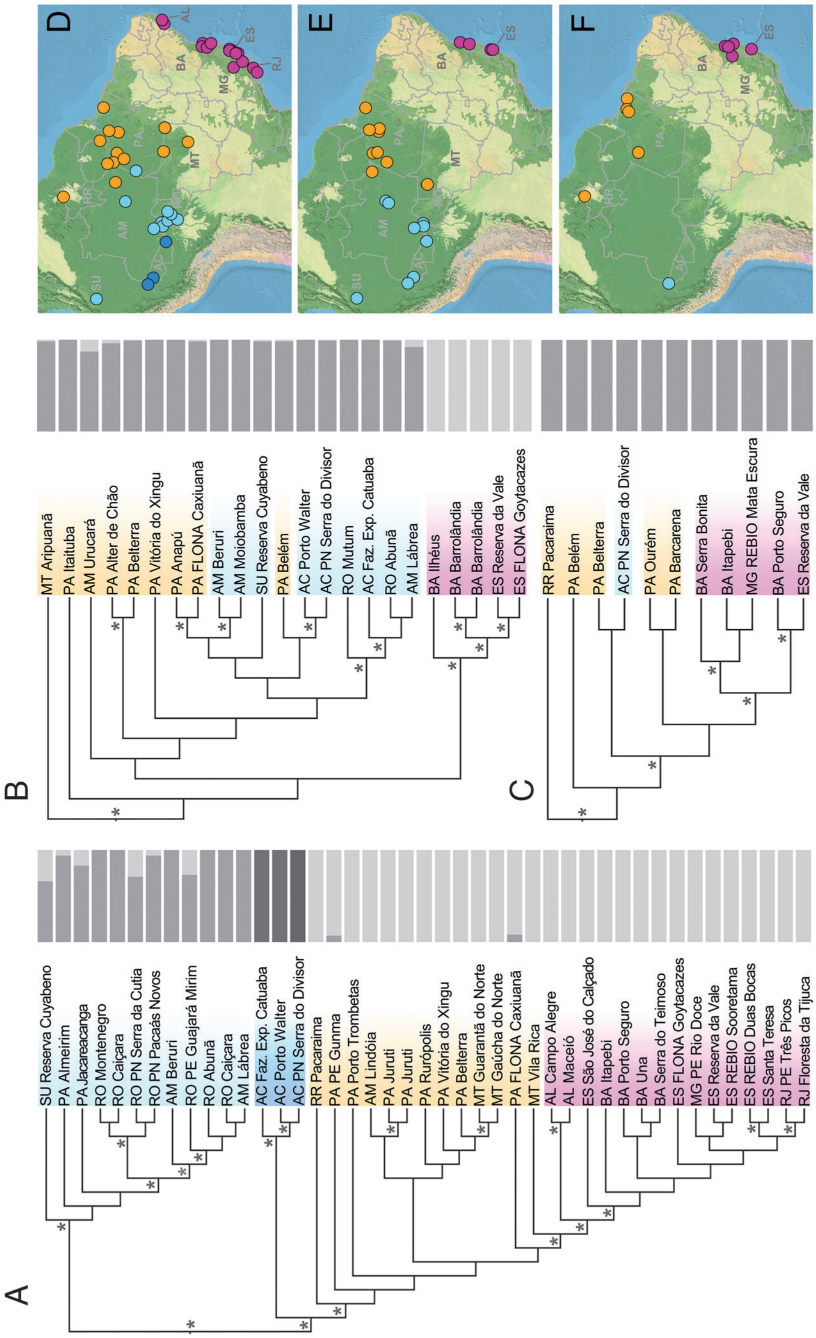
The spatial distribution of phylogenetic structure within all three species supports a scenario of colonization of the Atlantic Forest from the Amazonian domain (Fig. 3.1 and *Datasets S1–S3*[1]). This pattern is consistent with the hypothesis that former increase in precipitation promoted rainforest expansion and facilitated biotic exchange between Amazonia and the Atlantic Forest during the Quaternary (Cheng et al., 2013). It also agrees with patterns of phylogeographic structure reported for other vertebrate taxa (Costa, 2003; Batalha-Filho et al., 2013; Prates et al., 2015). In all species, we found that samples from the Atlantic Forest are monophyletic and are nested among Amazonian lineages, and more closely related to lizards from eastern than from western Amazonia (although the relationship is weakly supported for *A. ortonii* and *P. marmoratus;* Fig. 3.1). Nonetheless, spatial patterns of population genetic structure, inferred by a clustering algorithm (Frichot et al., 2014), differ across taxa (Fig. 3.1): *P. marmoratus* is recovered as largely panmictic, *A. ortonii* is composed of two genetic clusters (Atlantic Forest and Amazonia), and three clusters are recognized within *A. punctatus* (western Amazonia near the Brazil–Peru border, southwestern and central Amazonia, and eastern Amazonia plus the Atlantic Forest).

Species Show Different Historical Demographic Syndromes

Demographic analyses across spatial groups within each species recover signals of population size shifts in western Amazonia, eastern Amazonia, and the northern Atlantic Forest. However, the demographic scenarios that best fit the genomic data, as well as the estimated time of demographic change, differed across species. These observations are in

---

[1] Supporting information for this chapter, which includes *Datasets 1* through *8*, *Figs. S1* through *S6*, and *Table S1*, is available online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1601063113/-/DCSupplemental.


A
SU Reserva Cuyabeno
PA Almeirim
PA Jacareacanga
RO Montenegro
RO Caiçara
RO PN Serra da Cutia
RO PN Pacaás Novos
AM Beruri
RO PE Guajará Mirim
RO Abunã
RO Caiçara
AM Lábrea
AC Faz. Exp. Catuaba
AC Porto Walter
AC PN Serra do Divisor
RR Pacaraima
PA PE Gunma
PA Porto Trombetas
AM Lindóia
PA Juruti
PA Juruti
PA Rurópolis
PA Vitória do Xingu
PA Belterra
MT Guarantã do Norte
MT Gaúcha do Norte
PA FLONA Caxiuanã
MT Vila Rica
AL Campo Alegre
AL Maceió
ES São José do Calçado
BA Itapebi
BA Porto Seguro
BA Una
BA Serra do Teimoso
ES FLONA Goytacazes
MG PE Rio Doce
ES Reserva da Vale
ES REBIO Sooretama
ES REBIO Duas Bocas
ES Santa Teresa
RJ PE Três Picos
RJ Floresta da Tijuca
B
MT Aripuanã
PA Itaituba
AM Urucará
PA Alter de Chão
PA Belterra
PA Vitória do Xingu
PA Anapú
PA FLONA Caxiuanã
AM Beruri
AM Moiobamba
SU Reserva Cuyabeno
PA Belém
AC Porto Walter
AC PN Serra do Divisor
RO Mutum
AC Faz. Exp. Catuaba
RO Abunã
AM Lábrea
BA Ilhéus
BA Barrolândia
BA Barrolândia
ES Reserva da Vale
ES FLONA Goytacazes
C
RR Pacaraima
PA Belém
PA Belterra
AC PN Serra do Divisor
PA Ourém
PA Barcarena
BA Serra Bonita
BA Itapebi
MG REBIO Mata Escura
BA Porto Seguro
ES Reserva da Vale
D
AL
BA
ES
MG
RJ
PA
MT
RR
AM
RO
AC
SU
E
ES
BA
PA
MT
RO
AM
AC
SU
F
ES
BA
MG
PA
RR
AC

disagreement with the expectations of biological responses to out-of-phase climate dynamics as described above (Cheng et al., 2013). In the case of *A. punctatus,* the best-fit demographic syndrome (Excoffier et al., 2013) infers a scenario of population expansion [posterior probability (PP) ≥ 0.98] in each of the three regions, with an ~8- to 17-fold expansion dating to ~62 kya in western Amazonia [95% credibility interval (CI) = 15–124 kya; see *Dataset S4* for posterior estimates of all model parameters; posterior estimates of all model parameters had well-defined peak-shaped distributions; see plots in *Figs. S1–S4*], to ~120 kya in eastern Amazonia (CI = 34–207 kya), and to ~60 kya in northern Atlantic Forest (CI = 14–118 kya). The analyses of *A. ortonii* support an ~10- to 12-fold population contraction in all three regions (PP ≥ 0.88) dating to ~240 kya in western Amazonia (CI = 12–907 kya), to ~204 kya in eastern Amazonia (CI = 9–863 kya), and to ~81 kya in northern Atlantic Forest (CI = 10–300 kya). A scenario of an ~14-fold population contraction at ~72 kya (CI = 11–256 kya) was supported for *P. marmoratus* in the Atlantic Forest (PP = 1.0), the only region with sufficient sampling of SNPs per individual to enable historical demographic inference in this species.

Shifts in Population Size Are Asynchronous

Extending beyond region-specific inferences of population size change within each species, we applied hierarchical demographic models to combine data across regions and species and test for assemblage-wide synchrony in population shifts. To implement this multiregion, multitaxon framework, we used the aggregate site frequency spectrum (aSFS) to

FIGURE 3.1 Phylogenetic relationships between sampled individuals. Relationships for *A. punctatus* (*A* and *D*), *A. ortonii* (*B* and *E*), and *P. marmoratus* (*C* and *F*) were inferred through SVD quartets (Chifman and Kubatko, 2014) based on unlinked SNP data. Asterisks denote bootstrap support >0.70. Bars to the right of trees represent population genetic structure, inferred through sNMF (Frichot et al., 2014). Western Amazonian samples are indicated in blue, eastern Amazonian samples in orange, and Atlantic Forest samples in pink. In the case of western Amazonian *A. punctatus,* two distinct genetic populations (light and dark blue) were inferred with sNMF. Given the availability of samples, and to avoid combining different genetic populations in a single spatial group, we did not include the western Amazonia samples indicated by dark blue in our demographic analyses.

summarize patterns of genetic variation across groups and coupled it with coalescent simulations and a hierarchical approximate Bayesian computation (hABC) approach that allows species-specific parameters to vary independently (Xue and Hickerson, 2015). In an analysis that combined data from both *A. punctatus* and *A. ortonii* in all regions (western Amazonia, eastern Amazonia, and northern Atlantic Forest), we recovered a signal of largely asynchronous demographic change [dispersal index, the variance/mean of the time of population size change (DI) = 76,549; proportion of groups in a synchronous pulse (ζ) = 0.50; *Dataset S5* and *Fig. S5*]. A joint analysis restricted to the northern Atlantic Forest, where sufficient sampling of SNPs per individual is also available for *P. marmoratus*, supports a scenario of partially synchronous population size changes across the three lizard taxa (DI = 17,361; ζ = 0.67). These asynchronous or partially synchronous responses across species are consistent with the results of multiregion analyses conducted within each taxon separately. In a multiregion analysis restricted to *A. ortonii*, the data are consistent with a pattern of partially synchronous contractions, suggesting that two of the three regions responded in synchrony (DI = 41,314; ζ = 0.67). In *A. punctatus,* however, we detected a strong signal of synchronous population size shifts across all geographic regions, with coexpansion happening at ~119 kya [DI = 0.0; ζ = 1.0].

Although the population-level analyses largely support the occurrence of population changes within the time frame of documented atmospheric moisture shifts over the past 250 ky (Cheng et al., 2013), the contrasting syndromes among species (e.g., expansions in *A. punctatus* vs. contractions in *A. ortonii* and *P. marmoratus*) and the uniformity of demographic responses across forest regions and within species (in *A. punctatus* and *A. ortonii*) are inconsistent with the hypothesis that these lizards similarly tracked out-of-synchrony contractions and expansions of South America's lowland forests (Cheng et al., 2013). The general similarity of within-species population responses may indicate that demographic patterns are largely determined more by species-specific biological attributes than by spatial patterns of environmental variation across time. Future analyses of biological responses to the recently proposed South American macroclimatic zones (Cheng et al., 2013) will benefit from a more extensive spatial sampling, particularly in underexplored regions such as the extreme west of Amazonia and the Andean foothills, and also will profit from the inclusion of additional species.

Genetic Diversity Under Future Climates

The distinct responses to environmental shifts observed across these broadly codistributed taxa, as shown by the historical analyses, are also prevalent in predictions of the distribution of genetic diversity under future climates. Although forecast models predict severe loss of genetic diversity for both *A. punctatus* and *A. ortonii* by 2080, the projected spatial distribution of genetic diversity in these two species is markedly distinct (Fig. 3.2*I* and *J*). To develop models of the plausible future spatial distribution of allelic diversity, we adapted a recently proposed framework (Brown et al., 2016) that takes into account not only the distribution of suitable habitats under future climates but also species-specific ecological and demographic constraints inferred from molecular data. This framework first uses carefully parameterized species distribution models (SDMs) to approximate the spatial distribution of suitable habitats and landscape permeability to dispersal. Then empirical genetic (here reduced genomic) data are used to approximate the maximum density of individuals and the dispersal rates across the landscape (Brown et al., 2016).

Projection of present-day SDMs (Fig. 3.2*A* and *C*) onto 2080 climate predicts a dramatic decrease of suitable habitat along the current range of both species (Fig. 3.2*B* and *D*), including areas where present-day levels of genetic (SNP) diversity are high (Fig. 3.2*E* and *G*). On the other hand, the SDM projections into the future predict high suitability in presently cold areas of southern Brazil, where both species currently are absent. Despite such similar predicted habitat suitability, models of the plausible distribution of genetic diversity of the two species by 2080 differ dramatically (Fig. 3.2*F* and *H*). For *A. punctatus,* genetic diversity is predicted to decline strongly across the northern Atlantic Forest but to increase in the south (Fig. 3.2*F*). *A. ortonii* is expected to maintain some genetic diversity in the northern Atlantic Forest but is not expected to colonize the potentially new suitable habitats arising in the south (Fig. 3.2*H*).

Although models of species distribution under future climates have been widely used to guide conservation (Hijmans and Graham, 2006; Sinclair et al., 2010), these results suggest that relevant insight can be gained by incorporating genomic data in predictive studies. A naive interpretation of the SDMs of the two *Anolis* could flag southern Brazil as highly suitable for both species by 2080 (Fig. 3.2*B* and *D*), potentially preventing future loss of genetic diversity. However, our spatial demographic models indicate that these regions may in fact hold genetic diversity for *A. punctatus* (Fig. 3.2*F*) but not for *A. ortonii* (Fig. 3.2*H*). Such different predictions result from contrasting population densities and capacities for dispersal: When tied to the spatial data, SNP variation in *A. punctatus* indicated that, on average, 9.97% of each grid cell's occupants may emigrate at every generation, compared with 1.23% for *A. ortonii.* Our estimates of maximum carrying capacities further suggest that *A. punctatus* can attain much higher densities of individuals than *A. ortonii* (5,563 vs. 1,181 individuals per 30 km^2; for all parameters, PP = 1). (*Fig. S6* and *Table S1*). These results

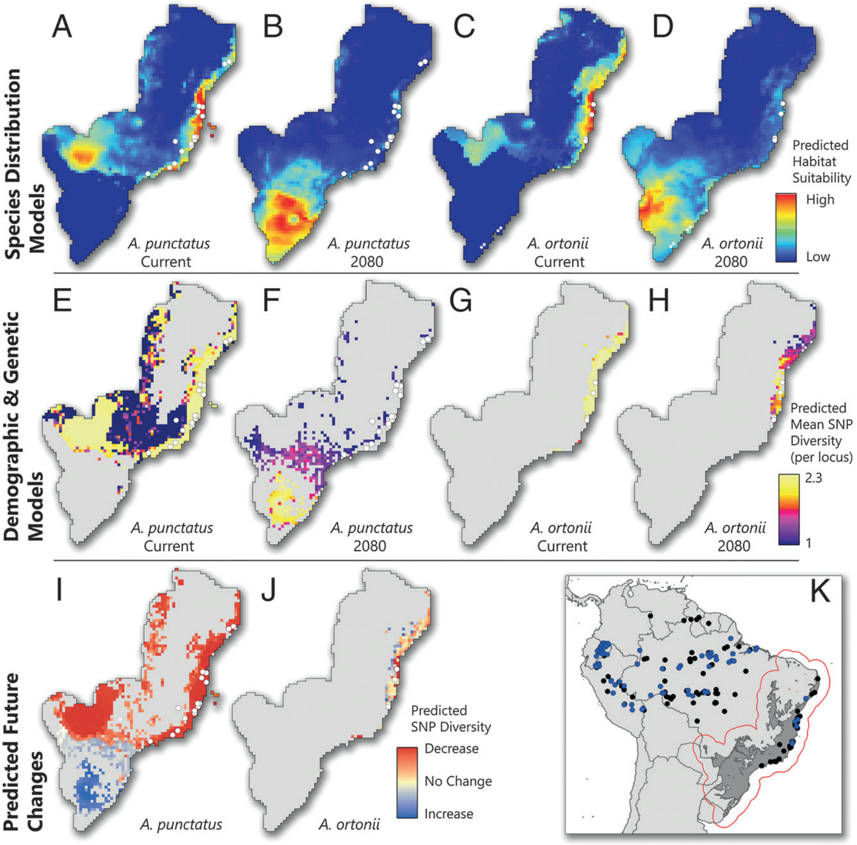


FIGURE 3.2 Predicted habitat suitability and genetic changes resulting from future climate change. (*A*–*D*) Species distribution models for the present time and 2080 for *A. punctatus* (*A* and *B*, respectively) and for *A. ortonii* (*C* and *D*, respectively). (*E*–*H*) Results from demographic and genetic modeling for the present time and 2080 for *A. punctatus* (*E* and *F*, respectively) and *A. ortonii* (*G* and *H*, respectively). (*I* and *J*) Projected plausible changes in genetic diversity by 2080 for *A. punctatus* (*I*) and *A. ortonii* (*J*). White dots in maps indicate the localities with empirical genetic data that were used for ABC parameter estimation. (*K*) Overview of demographic modeling extent (area enclosed by red line), SDM modeling extent along with vetted localities (blue, *A. punctatus*; black, *A. ortonii*), and historic distribution of the Atlantic Forest (depicted in dark gray).

correlate well with differences in body size and perceived abundances in the field: *A. punctatus* is much larger and also is observed much more frequently than *A. ortonii* in most of the areas we sampled. Importantly, these differences are consistent with the notion that dissimilar ecological and demographic constraints lead to distinct species responses to shared environmental change and are relevant for understanding demographic trends (Brown et al., 2016), even in species that occupy broadly similar climatic niches. Although challenging, the incorporation of species-specific ecological traits appears central to models of community responses to ongoing global change, both in the past and in the future.

CONCLUDING REMARKS

Through a comparative analysis of widely ranged and codistributed Neotropical forest lizards, we find concordant spatial patterns of phylogeographic structure across taxa, an observation consistent with previous mitochondrial DNA-based studies throughout the tropics (Schneider et al., 1998; Carnaval et al., 2009). Similar to the relatively few studies targeting multiple loci and tropical taxa (Lorenzen et al., 2012; d'Horta et al., 2013; Demos et al., 2014), we nonetheless find that the demographic processes underscoring those patterns vary across species. This view is supported by tests of synchronous population size shifts that allow for species-specific parameters as well as by plausible models of the distribution of genetic diversity under future carbon emissions. Taxa associated with broadly similar climatic conditions exhibit contrasting responses to common environmental changes, a scenario also supported by simulation-based macroecological studies (Zamudio et al., Chapter 11, this volume).

By combining genomic data with knowledge of the distribution of past, present, and future climates, we learn that species responses to environmental shifts are likely mediated, at least in part, by taxon-specific features that can be approximated through demographic simulations and molecular inference. The use of biological data to document or refine our understanding of the geological and climatic history of the Earth (Baker et al., 2014)—topics so fundamental to the description of the evolution of tropical biodiversity—will benefit from the incorporation of ecological differences among species. As we embark on a new genomic era, comparative phylogeography brings novel insights and exciting advances that bridge ecology and evolution and strengthens their links with the environmental sciences and conservation.

MATERIALS AND METHODS

Sequencing Protocols and Phylogenetic Analyses

We generated SNPs from 43 *A. punctatus* and 23 *A. ortonii* (Dactyloidae) and from 11 *P. marmoratus* (Polychrotidae) individuals (*Dataset S6*). For phylogenetic inference, we included one individual of *Anolis dissimilis, Anolis fuscoauratus, Anolis phyllorhinus, Anolis tandai, Anolis trachyderma, Anolis transversalis,* and *Polychrus liogaster* as outgroups. Genomic DNA was extracted from liver or tail fragments preserved in 100% ethanol through high-salt extraction and RNase treatment. Extractions were verified on agarose gels to ensure that the DNA was not degraded. DNA concentration was measured in a Qubit 2 fluorometer (Invitrogen) to ensure a final concentration of 30–100 ng DNA/µL in a total volume of 30 µL [in Tris EDTA buffer (Thermo Fisher Scientific, Waltham)].

We generated a restriction site-associated DNA library through genotype-by-sequencing (Elshire et al., 2011) at the Institute of Biotechnology at Cornell University. Genomic DNA was first reduced to a representation library using the EcoT22I restriction enzyme. Resulting fragments were tagged with individual barcodes, PCR-amplified, multiplexed, and sequenced in an Illumina HiSeq platform (Excoffier et al., 2013). The number of total single-end reads per individual ranged from ~300,000 to 6.3 million. We used a Pyrad 3.0.6 pipeline (Eaton, 2014) to demultiplex and assign reads to individuals based on sequence barcodes, to perform de novo read assembly (clustering similarity threshold > 0.9), to align reads into loci, and to identify SNPs and genotypes. We enforced a minimum Phred quality score (Waltari et al., 2014) and minimum sequence coverage (10×) to filter out poor-quality reads and reduce base-calling error. To reduce potential paralogs, we enforced a maximum number of heterozygous individuals per DNA site ($n = 5$). Final locus length ranged from ~50 to 150 bp. A single SNP per locus was randomly selected to reduce linkage disequilibrium across sites and ensure sampling of independent SNP histories. The final dataset was composed of 18,672 (*A. ortonii*), 15,421 (*A. punctatus*), or 4,589 (*P. marmoratus*) unlinked SNPs. Genetic datasets used in all analyses were deposited in Dryad (available at dx.doi.org/10.5061/dryad.2v1p6).

Phylogeographic Structure and Demographic Syndromes

We first inferred phylogenetic relationships with singular value decomposition scores for species quartets (SVD quartets) (Chifman and Kubatko, 2014), sampling all possible quartets and allowing for up to 30% of missing data. Node support was estimated with 1,000 bootstrap replicates. We then used sNMF 1.2 (Frichot et al., 2014) to test the best-fit number of genetic clusters within each species (*k*) and to assign individuals to populations. We tested multiple *k* values (1–10), with 20 replicate runs for each *k* value, and allowed a maximum of 30% missing data (each site was present in at least 70% of individuals). Robustness of the results was assessed by testing four values of the alpha regularization parameter (10, 100, 1,000, and 10,000).

Once we ensured that individuals from each spatial group (eastern Amazonia, western Amazonia, and northern Atlantic Forest) for each species did not violate the assumption of panmixia (based on the sNMF test of genetic clustering), we tested which historical demographic syndrome—instantaneous population expansion, instantaneous contraction, or constant size—best describes lizards in each region. For downstream site frequency spectrum (SFS)-based analyses, we used dadi 1.7 (Gutenkunst et al., 2009) to convert individual SNP data to an unpolarized SFS that summarizes patterns of site polymorphism across all sampled individuals within a population. The SNP data of each spatial group were down-projected to an SFS with equivalent sampling sizes across groups, allowing downstream aSFS analyses of groups composed of different numbers of sampled individuals (see below) while decreasing the effect of different levels of missing data between groups. Depending on the analysis, spatial groups of *A. punctatus* were down-projected to 10, 5, or 4 diploid individuals, *A. ortonii* groups were down-projected to five and four diploid individuals, and the *P. marmoratus* northern Atlantic Forest group was down-projected to four diploid individuals (see below and *Dataset S4* for the number of SNPs and down-projection scheme used in each analysis). The number of individuals in each down-projecting scheme was chosen to maximize the number of sampled individuals while minimizing levels of missing data in each group.

To estimate best-fit demographic syndromes, we used *fastsimcoal2* 2.5 (Excoffier et al., 2013) to simulate 100,000 coalescent simulations per demographic syndrome (for a total of 300,000 simulations) for each sampling level per spatial group. We then used an ABC procedure to compare each observed SFS with the appropriate simulations. Prior distributions used were time of instantaneous change ~*uniform* (*U*)(5 k, 1 million) generations ago, size change magnitude ~*U*(0.04, 0.2), with subsequent conversion to inverse in the case of contraction, current effective population size change for expansion ~*U*(100 k, 1 million), current effective population size change for constant size ~*U*(50 k, 500 k), and current effective population size change for contraction ~*U*(5 k, 100 k). An ABC model selection procedure was accomplished using the *R* package *abc* (Csilléry et al., 2012), using the model selection function *postpr()* to estimate the best-fit syndrome and the parameter estimation function *abc()* to infer posterior distributions for time, magnitude of population size change, and effective population size. Under ABC simple rejection sampling, we accepted the 1,500 (0.5%) simulations with the shortest Euclidian distance between the simulated and the empirical SFS. The model with the greatest posterior distribution then was selected as the best-fit demographic syndrome for each spatial group. To determine qualitatively whether tested models were discriminable based on the SFS, we visually inspected principal component analysis (PCA) plots based on a random selection of 100 simulated SFS under each model (for a total of 300 simulations). Moreover, to confirm that the best-fit model could generate the observed data, we conducted a PCA on the posterior-accepted simulated SFS, projected the empirical SFS into the principal components (PCs), and plotted the empirical data point, along with the accepted simulated data, on axes PC2 vs. PC1.

Estimating Magnitude of Synchronicity in Population Size Change

We tested the degree of synchronicity in population size changes across spatial groups through three sets of analyses. In analysis 1, we tested the degree of synchronicity in population size changes at a species level, comparing the three spatial groups within the better-sampled species *A. punctatus* and *A. ortonii* separately. Given the substantially greater sampling of *A. punctatus,* we down-projected the data of this species to 10 diploid individuals. In the case of *A. ortonii,* the data were down-projected to five diploid individuals. In analysis 2, we combined data from *A. punctatus* and *A. ortonii* spatial groups, all projected to five diploid individuals, to estimate the degree of synchronicity in population size shifts across all three geographic regions in both species. In analysis 3, we inferred the degree of synchronicity in demographic responses across the three species within the northern Atlantic Forest, a region for which enough sampled SNPs per individual of *P. marmoratus* were available. For this analysis, all three species in the northern Atlantic Forest were down-projected to four diploid individuals. In each analysis, we combined the empirical (observed) SFS of the spatial groups under consideration into an aSFS (Xue and Hickerson, 2015).

We then used an hABC procedure, which compares the observed aSFS with aSFS simulated under a hierarchical demographic model with varying degrees of synchronicity, to estimate (*i*) the proportion of groups in synchrony, ζ; (*ii*) the variance/mean of demographic change timing across spatial groups (with lower values signaling greater synchronicity), DI; and (*iii*) the timing of synchronous population size change (Xue and Hickerson, 2015). We performed 1 million coalescent simulations of the aSFS using *fastsimcoal2,* implementing the same prior distributions as the demographic syndrome analysis; the exception was the prior for ε in the expander populations, set to $\sim U(0.02, 0.2)$, and for current effective population size, set to ~*U*(500 k, 2 million) for expanders and ~*U*(50 k, 250 k) for contractors, because single-population analyses often had L-shaped posterior distributions for these parameters. The hyperprior for ζ was distributed according to a discrete uniform prior across all possible values from a single spatial group (i.e., completely idiosyncratic) to all spatial groups (i.e., complete synchrony) within a synchronous pulse. We also implemented a time buffer to ensure that asynchronous or idiosyncratic spatial groups responded at >20,000 generations apart from the synchronous time. This strategy aimed to increase the identifiability of ζ by reducing the temporal proximity of asynchronous responses to spatial groups in synchrony. In analyses in which the spatial group-level analyses supported expansion in some spatial groups but contraction in others (*Dataset S4*), we built the aSFS in three steps. First, an aSFS was built based only on the SFS of the spatial groups found to have expanded. Then another aSFS was built based only on the SFS of groups found to have contracted. Lastly, these two aSFS were concatenated to form a single summary statistic vector. Simulated aSFS were constructed accordingly to each respective empirical aSFS. With this method we were able to reduce statistical noise resulting from improper model specification and conflicting population signals.

Following hABC hyper-parameter inference, we accepted the 1,500 simulations whose aSFS had the shortest Euclidian distance from the empirical aSFS, using the *R* package *abc.* The function *postpr()* was used to infer ζ, with each discrete value treated as a separate model, and the function *abc()* was used to estimate the remaining parameter summaries (i.e., DI and synchronous time). Posterior distributions then were estimated based on the values of the retained simulations. As in the procedure for the spatial group-specific demographic syndrome analysis, we performed a PCA on the accepted simulations and plotted the simulated data with the empirical aSFS along the first two PCs to assess if the proposed scenarios could have generated the empirical aSFS. Furthermore, to assess hABC performance using the aSFS, we performed 50 leave-one-out cross-validations for the set of simulations per each empirical aSFS, with each cross-validation involving the extraction of a single simulation to act as a pseudo-observed dataset (POD) and using the remaining simulations for hABC estimation. This procedure was conducted using the functions *cv4postpr()* for ζ and *cv4abc()* for the parameter summaries, both contained in the *R* package *abc.* For each set of cross-validations, we calculated Pearson's correlation and the rms error between the known and estimated values across the 50 leave-one-out PODs. PCA plots based on the simulated SFS confirmed that the scenarios tested by the demographic analyses were generally discriminable (*Figs. S1–S4*). The PCA plots of the aSFS vectors calculated from the hABC rejection-sampled posterior, as well as from the aSFS calculated from the observed data, confirmed that best-fit models could generate the observed data. The leave-one-out cross-validation experiment indicated few biases and adequate accuracy in parameter estimates and model testing based on hABC analyses and the aSFS (*Dataset S5*).

Species Distribution Models to Inform Analyses of the Future Distribution of Genetic Diversity

To parameterize predictive models of future shifts in the spatial distribution of genomic diversity, we estimated migration rates and maximum carrying capacity at occupied sites from spatial demographic simulations that used friction and carrying-capacity layers derived from SDMs projected into former climates. We used the empirical SNP data to validate the simulated demographic scenarios (and hence to enable the estimation of demographic parameters) by following multiple forward-in-time spatial demographic simulations with a corresponding backward-in-time genetic simulation. The simulations closest to the empirical genetic data were retained for estimation of posterior distributions of model parameters. Given the complexity of the analysis, we restricted this inference to the Atlantic Forest and to the two best-sampled species, *A. punctatus* and *A. ortonii.*

Species distribution models were generated in MaxEnt 3.3.3k (Phillips et al., 2006), as implemented in SDMtoolbox 1.1c (Brown, 2014). We used museum records, data from collaborators, and data from our own fieldwork to compile occurrence records of *A. punctatus* and *A. ortonii,* which then were vetted for spatial and taxonomic accuracy. We then randomly selected one of multiple occurrence records per species within a 10-km radius using SDMtoolbox (Brown, 2014). The final dataset for SDM building was composed of 129 records of *A. punctatus* and 72 records of *A. ortonii* (*Dataset S7*). All models used eight standard bioclimatic variables summarizing patterns of precipitation and temperature, at a 2.5-arc-minute resolution (Bio1, 4, 10–12, 15–17 available at www.worldclim.org) (Hijmans et al., 2005). Models were built with occurrence points from the entire range of the species and later were clipped to the Atlantic Forest (–2.126 N, –59.567 W to –36.076 N, –34.717 W). Pseudoabsence (background) points were sampled from a minimum convex polygon defined as a 100-km buffer drawn around the occurrence localities (Barbet-Massin et al., 2012).

To parameterize the SDMs properly, we evaluated the performance of various combinations of five feature classes (linear; linear and quadratic; hinge; linear, quadratic, and hinge; and linear, quadratic, hinge, product, and threshold), and 10 regularization multipliers (from 0.5 to 5, in increments of 0.5) (Shcheglovitova and Anderson, 2013). To this end, we evaluated the performance of SDMs built under each combination of model parameters through a geographically structured *k*-fold cross-validation (i.e., the occurrence records were partitioned into *k* equal geographically clustered subsamples, here $k = 5$, and the models were trained with four of the groups and then evaluated with the excluded group until all group combinations were run). Model fit was assessed through inspections of the omission rate, area under the curve (AUC), and model feature class complexity (Brown, 2014). After optimum model parameters were determined (those leading to the lowest omission rate, highest AUC, and lowest complexity, in the order listed), a final SDM was built with all occurrence sites. For both *A. punctatus* and *A. ortonii*, the best-fit models were parameterized with linear, quadratic, and hinge feature classes and a regularization multiplier of 0.5 and 1, respectively. The best resulting SDMs for *A. punctatus* and *A. ortonii* (of 250 SDMs with alternative parameter combinations for each species) show high predictive accuracy (average omission rates 0.320 and 0.243, respectively) and high average AUCs (0.758 and 0.918). These final models were used to infer the range of each anole into the past and future. When projecting to variable values outside the observed range, all projections of the response curves were reset ("clamped") to match the upper or lower values found in the study area (Phillips et al., 2006).

Models of the paleodistribution of the species were derived with snapshot simulations of climate spanning the last 70 ky (at 1-ky time intervals back up to 22 kya, and before that at 2-ky time intervals, with 46 time periods in total) derived from the Hadley Centre Climate model (HadCM3) (Fuchs et al., 2013; Carnaval et al., 2014). To project species ranges and genetic diversity under future climates, we used bioclimatic variables estimated from the Hadley Center model for the years 2030, 2050, 2070, and 2080 (HadGEN2-ES) (IPCC, 2015), which is based on the RCP 6.0 emission scenario (Van der Linden and Mitchell, 2009).

Estimating Carrying Capacity and Friction from SDMs

The SDMs generated under each former climatic period were rendered into two landscapes, friction and carrying capacity, for use in spatial demographic simulations (Knowles and Alvarado-Serrano, 2010; Brown and Knowles, 2012). The carrying capacity represents the maximum number of individuals that can occupy a grid cell at that time and is proportional to the suitability values given by the SDM. The friction values depict the relative difficulty of movement of individuals across different points in the landscape and are inversely proportional to the suitability inferred for the adjacent cells. To convert the SDMs into these friction and carrying-capacity landscapes, we tested five transformation functions:

linear, sigmoid (with two different inflection points), and binary (with two different cutoffs) (Brown and Knowles, 2012). The sigmoid function best matched current distribution data. Conversion was performed with a normal cumulative distribution function, where the inflection point of the sigmoidal curve was 0.5 and the slope of the curve was 0.091 for *A. punctatus* and 0.073 for *A. ortonii.* These values were defined based on suitability values (from the SDM) measured at sites with known occurrences, where half of the SD equaled the slope of the curve (Brown and Knowles, 2012).

Spatial Demographic Simulations for the Past

A total of 200,000 forward-in-time simulations of dispersal were performed under different migration rates and maximum carrying capacities. These 2D stepping-stone demographic colonizations (Kimura and Weiss, 1964) were simulated based on the per-deme maximum carrying capacity and migration limitations imposed by the friction and carrying capacity landscapes, using the program SPLATCHE 2 (Ray et al., 2010). In the demographic modeling, the colonization of the landscape proceeded in a generation-by-generation manner, from a predetermined ancestral source area(s) (see below). In each generation, population growth in each deme (i.e., an occupied cell) followed a logistic model characterized by a per-generation intrinsic growth rate (set to 0.2) and the time-specific SDM-informed maximum carrying capacity. This step was followed immediately by migration of individuals into neighboring demes, conditioned by the time-specific SDM-based friction of the landscape. The number of emigrants (dispersers) per generation from a given deme was modeled as a function of the migration rate (*m*, the proportion of individuals of a deme emigrating at each generation) and of the total number of occupants in that deme (which in turn depends on the local carrying capacity and on the population growth rate, under a logistic growth trend). A spatial resolution of ca. 30 km^2 (~0.35°) was used for the demographic modeling because the spatial autocorrelation of the observed genetic data approached zero at this resolution [measured in SAM 4.0 using a spatial correlogram (Rangel et al., 2010)].

To define the ancestral area from which spatial simulations would start, as well as the initial population density in those cells, we considered cells whose suitability score corresponded to the top 0.1 quartile of suitability values from the SDM projected to 70 kya [the oldest time period considered in our spatial demographic analyses (Knowles and Alvarado-Serrano, 2010)]. At the start of the simulations, each cell in which the species was considered to be present was assigned the maximum number of individuals (i.e., the maximum per deme carrying capacity). If the cell's actual carrying capacity, as informed by the corresponding projected SDM, was lower than the maximum capacity (i.e., if its suitability was lower than the maximum), the remaining individuals were distributed in neighbor cells. Because of computational limitations on the number of generations and on spatial resolution, a generation time of 10 years was used in the spatial demographic simulations. To adjust the results of the simulations to the more biologically plausible generation time of 1 year for anole lizards (Tollis et al., 2012; Muñoz et al., 2013), we multiplied the mutation, recombination, and growth rates by 10.

Spatial Genetic Simulations for the Past

To use the empirical data to validate the simulated demographic scenarios, and hence enable estimation of demographic parameters, each forward-in-time spatial demographic simulation was followed by a corresponding backward-in-time genetic simulation. The latter was parameterized by the spatially explicit demographic conditions simulated first. Specifically, the record of the per-generation movement of individuals and deme population sizes provided by the demographic simulations informed the probability of within-deme coalescent events and the probability of backward-in-time movement to neighboring demes of each sampled SNP. SNPs from focal populations were sampled from the resulting genealogies (Knowles and Alvarado-Serrano, 2010; Brown and Knowles, 2012). The simulated genetic data matched the empirical genetic dataset used in this analysis: We simulated and sampled the same number of markers (4,000 SNPs) and the same number of individuals per sampled cell (15 sampling localities for *A. punctatus* and 4 for *A. ortonii*) (*Dataset S8*). Given the uncertainty in mutation rates for SNP data, we simulated SNPs into four equal groups (1,000 SNPs each) with mutation rates at each order between 1×10^{-5} to 1×10^{-8} mutations per site per generation. Rates of recombination were set to 0.822 cM/Mb based on genomic averages (Dumont and Payseur, 2008). Prior distributions were set as 3 to 5 for maximum landscape carrying capacity (*maxK*) and –3 to –0.5 for *m* (values in $\log_{10}$). Simulations were summarized into a total of 243 summary statistics for *A. punctatus* and 45 for *A. ortonii* with Arlequin v3.5 (Excoffier and Lischer, 2010), including the statistics for between-site comparisons [e.g., mean number of alleles (*r*), *r* SD, mean heterozygosity (*H*), *H* SD, mean total *H*, mean number of pairwise differences, number of private polymorphic sites, and mean number of sites with segregating substitutions] and global descriptions (over all sites, e.g., total number of alleles, total number of polymorphic sites, total *H*). Because the use of a large number of summary statistics may reduce the accuracy of the ABC estimation as a result of high dimensionality (Blum, 2012), summary statistics were transformed into partial least squares (PLS) components

(Boulesteix and Strimmer, 2007) using the first 10,000 simulations as random calibration. After exploring the reduction of the rms errors associated with each additional PLS component for each parameter (Wegmann et al., 2009), we retained only the first nine PLS components. The 1,000 simulations (0.5%) closest to the empirical genetic data were retained for estimation of posterior distributions of model parameters (Beaumont et al., 2002), after adjustment with a general linear model (Leuenberger and Wegmann, 2010). The fit of the model to the empirical data was evaluated by assessing the fraction of retained simulations that had likelihood smaller than or equal to that of the observed data (a small fraction suggests that the model does not fit well with the empirical data) (Wegmann et al., 2010). Accuracy of parameter estimation was assessed by calculating the coefficient of determination (R^2) between each parameter and the nine retained PLS components (Neuenschwander et al., 2008).

Estimating the Genetic Consequences of Future Climate Change

Once we estimated the demographic parameters for each species, we proceeded to map the distribution of genomic diversity currently and in the face of projected future climate change. To do so, we ran a final demographic simulation starting at 70 kya, which was run forward until two points in time: the present (ca. 2010) and the year 2080. For each past time period, the posterior estimates (modes) of ancestral population size, migration rate, and carrying capacity derived from the historical analyses were used to parameterize the simulations. All other parameters, including growth rate, were kept at the values used for the initial spatial demographic simulations. Genetic constitution of every occupied locality was sampled at the end of each model run (on average, 1,560 and 183 grid cells for the present time and 1,508 and 172 cells for 2080, for *A. punctatus* and *A. ortonii,* respectively). Demographic and genetic simulations were repeated 100 times for each temporal scenario, through 20 replicate demographic models, each with five genetic simulation replicates (Gehara et al., 2013). To measure predicted changes compared with present-day spatial patterns of genetic diversity, the genetic landscape predicted for 2080 was subtracted from that for the present day. Because output summary statistics of the simulated genetic data were highly correlated ($r = 0.81$, $P < 0.001$), we chose to present only mean allelic diversity here, which depicts the average allelic diversity (r) among the 4,000 simulated SNPs.

ACKNOWLEDGMENTS

We thank all the Brazilian collectors who made this study possible and particularly students in the M.T.R. laboratory for field support. Discussions with the A.C.C. and M.J.H. laboratories greatly improved this manuscript. Ana Prudente provided samples deposited at the Museu Paraense Emilio Goeldi. Frederick H. Sheldon provided samples deposited at the Louisiana State University Museum of Natural Science. The Instituto Chico Mendes de Conservação da Biodiversidade issued Sistema de Autorização e Informação em Biodiversidade (SISBIO) collection permits 36753-1, 36753-4, and 27290-3. This work was largely cofunded by São Paulo Research Foundation (FAPESP) Grant BIOTA 2013/50297-0, National Science Foundation (NSF) Grant DEB 1343578, and the National Aeronautics and Space Administration through the Dimensions of Biodiversity Program. Additional funding included NSF Doctoral Dissertation Improvement Grant DEB 1601271 and a City University of New York (CUNY) Graduate Center Doctoral Student Research Grant (to I.P.), NSF DEB Grant 1253710 (to A.T.X., D.F.A.-S., and M.J.H.), FAPESP Grants 03/10335-8 and 11/50146-6 (to M.T.R.), and NSF-DEB Grant 1120487 and a CUNY Junior Faculty Research Award in Science and Engineering (to A.C.C.). CUNY's High Performance Computing, used in data analyses, is supported by NSF Grants CNS-0855217, CNS-0958379, and ACI-1126113.

4

Comparative Phylogeography of Oceanic Archipelagos: Hotspots for Inferences of Evolutionary Process

KERRY L. SHAW*‡ AND ROSEMARY G. GILLESPIE†

Remote island archipelagos offer superb opportunities to study the evolution of community assembly because of their relatively young and simple communities where speciation contributes to the origin and evolution of community structure. There is great potential for common phylogeographic patterns among remote archipelagos that originate through hotspot volcanism, particularly when the islands formed are spatially isolated and linearly arranged. The progression rule is characterized by a phylogeographic concordance between island age and lineage age in a species radiation. Progression is most likely to arise when a species radiation begins on an older island before the emergence of younger islands of a hotspot archipelago. In the simplest form of progression, colonization of younger islands as they emerge and offer appropriate habitat, is coincident with cladogenesis. In this chapter, we review recent discoveries of the progression rule on seven hotspot archipelagos. We then discuss advantages that progression offers to the study of community assembly, and insights that community dynamics may offer toward understanding the evolution of progression. We describe results from two compelling cases of progression where the mosaic genome may offer insights into contrasting demographic histories that shed light on mechanisms of speciation and progression on remote archipelagos.

---

*Department of Neurobiology and Behavior, Cornell University, Ithaca, NY 14853; and †Environmental Science Policy and Management, University of California, Berkeley, CA 94720. ‡To whom correspondence should be addressed. Email: KLS4@cornell.edu.

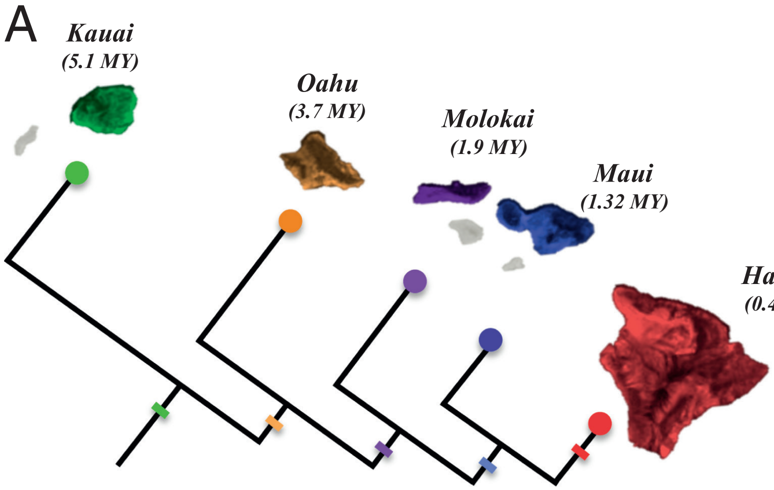
Evolutionists are drawn to the study of island biotas for their unique species (Otte, 1994; Rubinoff and Haines, 2005; Grant and Grant, 2008; Bowen et al., Chapter 1, this volume; Yoder et al., Chapter 12, this volume), as refuges of extant, "relict" organisms (Buckley et al., 2009; Wood et al., 2015), and most famously, for their displays of adaptive radiation (Losos and Ricklefs, 2009; Gillespie, 2015). Likewise, ecologists have long recognized the value of islands as microcosms of the processes of community assembly (MacArthur and Wilson, 1967; Warren et al., 2015). In recent years, there has been growing interest in combining these elements to study the evolution of community assembly, with particular focus on islands within (and beyond) the "radiation zone" (MacArthur and Wilson, 1967), where in situ speciation can be a major contributor to the origin of ecological communities (Heaney, 2007; Rominger et al., 2016). Under such circumstances, rules for community assembly can be illuminated by comparative phylogeographic approaches, revealing common evolutionary histories of codistributed, endemic taxa both within and between island archipelagos.

Remote island archipelagos offer relatively simple arenas for the evolutionary dynamics of community assembly because they are generally small in size and are often characterized by spatial isolation beyond the probable dispersal range of most organisms (Gillespie and Baldwin, 2009; Rosindell and Phillimore, 2011). As G. G. Simpson (1940) hypothesized, the probability of colonization should decline with increasing remoteness of an island from a mainland source pool, which he aptly named "sweepstakes" dispersal. Moreover, remote island archipelagos are veritable specks of land in a wide ocean world, again reducing the probability of colonization. Evidence of these hypothesized filters to colonization shows that remote islands frequently harbor disharmonious biotas (where there is an imbalance of taxonomic representation compared with mainland source pools). Moreover, the degree of disharmony increases with increasing distance from a probable source pool offering additional evidence of these spatial sieves and the potential for a relatively simpler community assembly (Gillespie and Roderick, 2002). Together with small size and extreme isolation, the assemblage of biotas may be further reduced in remote oceanic archipelagos because they are generally formed without life due to the nature of their geological origins (Gillespie et al., 2012). In contrast, species and community evolution in island systems in close proximity to continental species pools (e.g., continental "fragment" islands) (Gillespie and Roderick, 2002) can be extraordinarily complex (Steppan et al., 2003) in particular because they are often formed with a full complement of species and have more frequent connections with source pools.

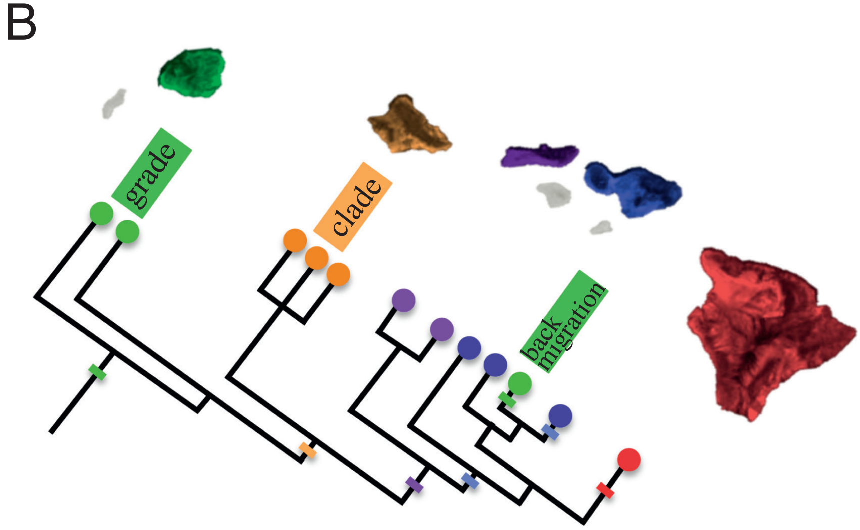
Along with the effects of small size, spatial isolation, and dispersal origins on community assembly of remote islands, common phylogeographic patterns may further enhance the "laboratory" for study of the evolution of community assembly. The evidence of disharmony, and potential for geographically diverse source pool origins (Gillespie and Roderick, 2002; Steppan et al., 2003; de Queiroz, 2005; Keeley and Funk, 2011; Gillespie et al., 2012), might lead us to expect somewhat haphazard phylogeographic patterns in taxa distributed across multiple remote archipelagos (Gillespie et al., 2012), for example, as appears to be the case for the spider genus *Tetragnatha* across Polynesia (Gillespie, 2002). However, once a lineage has initially established within a remote archipelago, it is largely cut off from its source population (qualified, of course, on the dispersal ability of the taxon). Thereafter, colonization among the constituent islands is considerably more predictable, and geological history at this more restricted geographic scale inspires hypotheses of common phylogeographic patterns among codistributed taxa (Funk and Wagner, 1995b; Avise, 2000).

Many remote archipelagos are volcanic in origin and sometimes exist in a linear age progression. These "hotspot" archipelagos are formed by molten lava rising from relatively fixed spots on the sea floor. Plate tectonic drift creates a conveyor belt motion that carries newly formed islands away from the hotspot location in a consistent direction at a steady rate, the mechanism that generates the linear age progression (Wilson, 1963; Clouard and Bonneville, 2005). These geological origins create a geographical and chronological context for hypotheses of phylogeographic congruence among codistributed taxa that track the ages of the islands, termed the "progression rule" (Figs. 4.1 and 4.2).

Common evolutionary histories of taxa, as well as deviations, can have a profound impact on our understanding of community assembly in the radiation zone. In this paper, we examine the progression rule to gain insight into the process of community assembly. In the following sections, we first present a brief overview of the progression rule from exemplar remote oceanic archipelagos. We then consider a simple question: can the progression rule be explained on the basis of island hot spot theory alone? We evaluate recent studies of progression from remote oceanic islands, highlighting two cases where contrasting phylogeographic histories are implicated within the same species radiations. The potential for distinct genetic histories held within the same demographic taxon, a phenomenon now widely recognized (Seehausen et al., 2014), illuminates the link between ecological community assembly and evolutionary assembly in isolated archipelagos.



Simple progression



Clades and grades

FIGURE 4.1 Hypothetical progression patterns. (*A*) A simple progression pattern with cladogenesis coincident with interisland migration. Hatch marks show interisland migrations. (*B*) A complex progression showing clades, grades, and backmigration. Hatch marks show interisland migrations. Colors correspond to islands: green, Kaua'i; orange, O'ahu; purple, Moloka'i; blue, Maui; and red, Hawai'i.

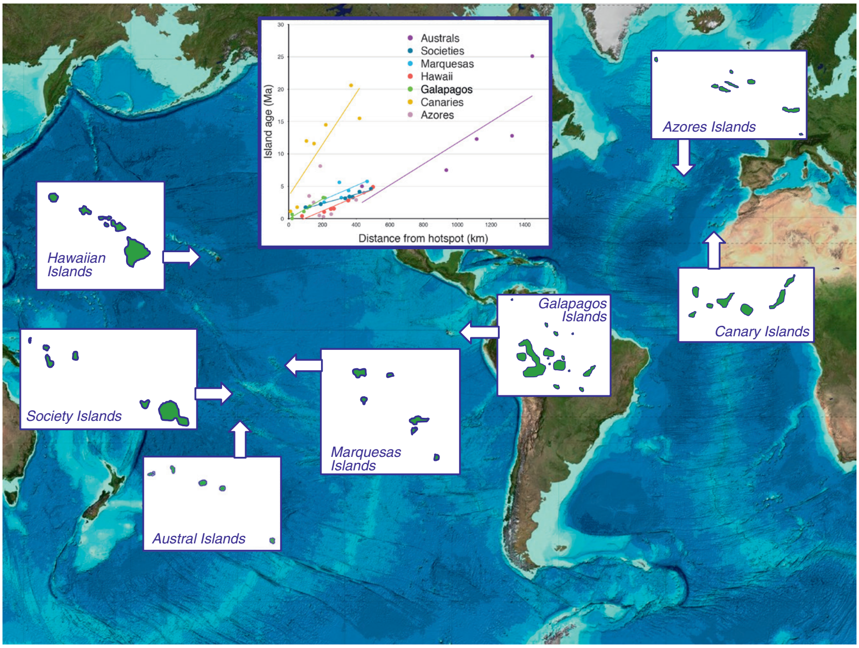


FIGURE 4.2 Map of seven archipelagoes that show a variable geological age chronology. (*Inset*) Relationship between island age and distance from the hotspot for each archipelago (estimated in the case of the Azores) (Cannat et al., 1999). This reflects the match of the geographical arrangement of islands to their age, with tighter relationships in archipelagoes such as that of Hawai'i, in which islands are produced in a conveyor-belt fashion. The figure includes only islands of area >10 km^2, as smaller islands tend to be atolls with communities limited to coastal strand. Trend lines are shown (except for the Azores in which there is no clear trend line). Background image reproduced from the GEBCO world map 2014, www.gebco.net.

THE PROGRESSION RULE IN COMPARATIVE ISLAND PHYLOGEOGRAPHY

In the context of islands, a progression rule (or pattern) refers to a phenomenon of phylogeographic concordance with island age, whereby older lineages map to older islands within an archipelago, and younger lineages map to progressively younger islands in that system (Fig. 4.1*A*) (Funk and Wagner, 1995b). Progression is hypothesized to result from the early colonization by a lineage when older islands existed but younger islands had not yet formed. Subsequent colonization occurs as new islands form. In his seminal work on phylogenetic theory, Hennig (1966) proposed the progression rule as a consequence of a general speciation mechanism, wherein species ranges consist of an ancestral, central portion within an older geographic area, and a derived, peripheral portion in a younger geographic area. Progression becomes evident in the phylogeny once speciation occurs; the ancestral, central portion splits from the derived, peripheral portion of the ancestral species. Thus, the ancestral lineage is concordant with older area, whereas the derived lineage is concordant with newer geological areas (Fig. 4.1*A*). Hennig's (1966) model would not predict a progression to the extent of present-day discussions, unless the next new habitat that forms is spatially closer to the newest of the previous habitats.

Oceanic hotspot archipelagos offer, at a minimum, the starting conditions where progression could begin. Indeed, at least superficially, progression is not difficult to explain, as colonists of younger islands are most likely to come from spatially proximate, older islands. However, there is ample variation among taxonomic lineages in both the degree and form of progression (see below). In general, progression is more likely to evolve with (*i*) increasing spatial isolation of the archipelago, (*ii*) increasing spatial linearity of the islands, and (*iii*) enough dispersal to ensure colonization, such that when new habitat arises after the emergence of a new island, propagules are spatially poised to colonize from the next oldest habitat.

Variation in the starting conditions for progression may arise for a variety of reasons (we treat the persistence of progression below). For example, some archipelagos are less isolated from potential source areas (relative to the dispersal ability of the organisms) than others, resulting in repeated colonizations by the same lineages, producing conditions for either biotic turnover or anagenesis, rather than the within-archipelago cladogenesis associated with the progression pattern (Rosindell and Phillimore, 2011). Effectively, too much dispersal prevents the development of a progression pattern. In addition, although less likely, some organisms may be dispersal limited, and while habitat is available, they do not disperse to it predictably. In addition, some archipelagos do not show a linear spatial arrangement concordant with age, resulting in potential within archipelago colonizations from islands of mixed age. This nonlinear age arrangement would interfere with the older to younger colonization and ensuing phylogeographic pattern characteristic of progression. Likewise, we would not predict a progression for taxa arriving after the current major islands came into existence.

Progression can be complex; Funk and Wagner (1995a) discuss a variety of reasons for this complexity, providing a general framework for interpreting more complex patterns. A strict progression pattern is the simplest, where all cladogenic events arise coincident with migration to newly arisen volcanoes along a linear chronological island sequence (Fig. 4.1*A*). Funk and Wagner (1995a) discuss this pattern with reference to Hawaiian endemics, where it is rare for taxa to conform exactly to this simple manifestation of the progression rule, although some taxa do illustrate the progression perfectly (Magnacca and Price, 2015).

Another more common pattern shows older to younger island colonization followed by some degree of within-island speciation (a pattern of progressive clades; Fig. 4.1*B*) (Funk and Wagner, 1995a). This progression pattern would arise when the interisland colonists from an older island migrate early in the history of the extant, older island clade. Progressive clades can give way to progressive grades (Fig. 4.1*B*), where a formerly monophyletic clade within an older island becomes paraphyletic when a subsequent colonization to a younger island occurs by a terminal taxon within that clade. In addition to these basic progression histories, instances of backmigration (migrations from younger to older islands) may also occur (Fig. 4.1*B*).

Armed with some understanding of the complexities of the progression pattern, we now examine progression in some well-recognized hotspot archipelagoes, with particular focus on the Pacific Ocean (Fig. 4.2).

Hawaiian Islands

Progression has been established most famously as a general pattern in the Hawaiian Archipelago (Fig. 4.2), where the formation of new, colonizable habitats does follow such a chronological sequence (Funk and Wagner, 1995b). Hawai'i's indigenous biota is highly disharmonious, and chance colonization has undoubtedly played a large part in the initial establishment of the native diversity that we see today. However, within the Hawaiian Islands, a taxonomically broad expression of the progression rule is evident, including examples from plants [e.g., Hawaiian silverswords (Baldwin and Robichaux, 1995); *Schiedea* (Nepokroeff et al., 2003); *Psychotria* (Ree and Smith, 2008); and lobeliads (Givnish et al., 2009)], insects [e.g., *Megalagrion* damselflies (Jordan et al., 2003); *Laupala* crickets (Mendelson and Shaw, 2005); *Banza* katydids (Shapiro et al., 2006); *Hyposmocoma* moths (Haines et al., 2014); and picture-winged *Drosophila* (Bonacum et al., 2005; Magnacca and Price, 2015)], land snails [e.g., succineid species (Rundell et al., 2004)], spiders [e.g., *Orsonwelles* (Hormiga et al., 2003)], and birds (VanderWerf et al., 2010), among many others. Funk and Wagner (1995b) documented progression in 18 of 25 endemic lineages. More recent reviews (Roderick and Gillespie, 1998; Cowie and Holland, 2008) have further summarized this impressive phylogeographic result.

Austral Islands

The Austral Archipelago (Fig. 4.2) is considered geologically continuous with the Cook Islands (located to the northwest), which together were formed from repeated episodes of volcanism at several sites (Dickinson, 1998; Bonneville et al., 2002). Like the other Pacific hotspots, they are sequentially ordered from southeast to northwest by increasing age, although there has been secondary volcanic activity in the older islands, making them older than other Pacific hotspots (Clouard and Bonneville, 2005) (Fig. 4.2, *Inset*; and *Table S1*[1]). Among independent lineages of spiders, in particular the crab spider *Misumenops rapaensis* (Garb and Gillespie, 2006), the orb web spider *Tangaroa tahitiensis* (Uloboridae) (Gillespie et al., 2008), and *Rhyncogonus* weevils (Claridge et al., 2016), a similar pattern of sequential colonization of islands is emerging, with large genetic distances between island populations. Thus, despite the modest extent and topography of the Austral islands and the widespread and generalist nature of their taxa, all studies to date show strong support for a progression rule in this archipelago.

Society Islands

Age progression within the Society Islands is in good agreement with the fixed hotspot hypothesis (Clouard and Bonneville, 2005), although the islands are considerably smaller in size than those of the Hawaiian chain (Fig. 4.2 and *Table S1*). However, there is considerable topographic diversity. Among insects, blackflies in the genus *Simulium* (Joy and Conn, 2001; Craig, 2003) show no evidence of a progression rule, which may be because of extinctions of habitat-specialized species on the older islands, due to loss of habitat. Likewise, among weevils in the genus *Rhyncogonus*, there appear to have been multiple independent colonizations of the island of Tahiti from neighboring island chains and no evidence of progression from older to younger islands (Claridge, 2006; Claridge et al., 2016). A similar conclusion results from phylogenetic analyses of Polynesian reed-warblers (genus *Acrocephalus;* Cibois et al., 2011). One reason that has been suggested is that there are ancient islands in close proximity to the Societies that may have served as a source of propagules (Gillespie et al., 2008).

---

[1] Supporting information for this chapter, which includes *Table S1*, is available online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1601078113/-/DCSupplemental.

Marquesas Islands

The chronological arrangement of the Marquesas Islands is not strictly regular (Clouard and Bonneville, 2005). Nevertheless, among birds, an approximate progression from older to younger islands is found in the genus *Pomarea* (Cibois et al., 2004). Among spiders, the endemic taxa fall into a northern and southern lineage (Gillespie, 2003), consistent with the progression, although without strong phylogenetic support. A similar pattern of northern and southern lineages is found in *Rhyncogonus* weevils (Claridge, 2006) and partulid land snails of the genus *Samoana* (Johnson et al., 2000).

Galapagos Islands

The Galapagos Islands, although more geographically clustered than other Pacific hotspots (White et al., 1993), are still arranged chronologically (Fig. 4.2). This spatial arrangement implies a potential diversification sequence from southeast to northwest, paralleling the geological formation of the archipelago's islands (Poulakakis et al., 2008). Some lineages in the Galapagos show a very clear progression from older to younger islands (Parent et al., 2008), well illustrated by the Galapagos giant tortoise (*Geochelone nigra*), in which both the species-level phylogeographic pattern based on mtDNA data and the pattern of lineage sorting suggest diversification in parallel with the island geological formation (Caccone et al., 2002; Beheregaray et al., 2004). One of the two lineages of Galapagos lava lizards (*Microlophus*) (Wright, 1983; Lopez et al., 1992) has also diversified in concert with the geological formation of island clusters of similar age (Kizirian et al., 2004). Among marine iguanas (*Amblyrhynchus cristatus*), mtDNA shows population differentiation concordant with geographical isolation of populations across the archipelago, a result largely in agreement with nuclear microsatellite data (Steinfartz et al., 2007). Among birds, Galapagos mocking birds also appear to follow the progression rule (Arbogast et al., 2006), although Darwin's finches show limited evidence of diversification closely associated with the geological formation of the islands (Grant and Grant, 2008). A progression pattern has been inferred for Galapagos bulimulid land snails (Parent and Crespi, 2009). In contrast, insects including Galapagos flightless weevils (*Galapaganus*) (Sequeira et al., 2008) and the microlepidopteran genus *Galagete* (Schmitz et al., 2007) do not follow the progression rule. Evidently, many Galapagos terrestrial faunal groups follow the progression rule, the major exceptions coming from more vagile lineages (Galapagos finches, *Galagete* lepidopterans, and *Galapaganus* weevils) (Parent et al., 2008). Interestingly, in no situation is adaptive radiation associated with progression.

Canary Islands

Phylogeographically, the Canary Islands are the most thoroughly studied of all of the Atlantic island groups (Juan et al., 2000). In this archipelago, the islands farther from the mainland are younger, and those closer to the mainland (Fuerteventura and Lanzarote) are older (Fig. 4.2, *Inset*). Compared with the Pacific volcanoes, the archipelago is much older and moves much more slowly. Older islands closer to the continent are drier, as well as lower in elevation; thus, Fuerteventura and Lanzarote do not contain any wet forest habitats. The progression rule appears to be a common pattern of colonization shown by several groups of organisms (Juan et al., 2000; Sanmartín et al., 2008; Faria et al., 2016). However, species may, at least sometimes, also disperse from younger to older islands, even when these islands are occupied by close relatives (Juan et al., 2000). Among spiders, a progression rule has been documented for *Loxosceles* (Planas and Ribera, 2014) and *Dysdera* (Arnedo et al., 2001; Macías-Hernández et al., 2008), whereas *Pholcus* and *Spermophorides* probably colonized the older Fuerteventura and Lanzarote from the younger Gran Canaria (López-Mercader, 2005; Dimitrov et al., 2008). The mixed support for a progression across the Canary Islands presents the possibility of an alternative explantion for the origin of the endemics of the oldest islands, which could be the result of secondary replacement of its original fauna by new colonists better adapted to increasingly arid conditions (Cardoso et al., 2010). Older islands in the Canaries may show loss of old resident species due to aridification and orographic simplification; in some lineages, this may have been compensated by colonization and subsequent diversification of new, better-adapted organisms that could take advantage of empty niches and new opportunities.

Azores Islands

The Azores have also been used for examination of the role of island progression in dictating biogeographic patterns (Carvalho et al., 2015). However, the islands are located over a complex microplate rather than a single hotspot. Nevertheless, there is some evidence for a directional mode of dispersal from older to younger islands [plants (Díaz-Pérez et al., 2008; Rumeu et al., 2011), arthropods (Amorim et al., 2012; Parmakelis et al., 2015)], although a simple progression rule is not common [yet Van Riel (2005) presents intriguing evidence from gastropods]. Much of the current evidence consists of age estimates of haplotypes/alleles rather than robustly structured cladogenetic branching between islands. This pattern may be in part due to young radiations, since much of the landmass in the Azores is younger than 1 Ma. Moreover, the islands have suffered extensive habitat modification due to more recent volcanic activity and deforestation of native forests over many centuries of occupation (Carvalho et al., 2015).

PERSISTENCE OF THE PROGRESSION PATTERN AND THE EVOLUTION OF COMMUNITY ASSEMBLY

As with most evolutionary processes, community assembly in the radiation zone happens more slowly than can be observed. Progression patterns on remote oceanic archipelagos are valuable in that they suggest a temporal frame of reference among extant communities as well as comparative insights into past events. For example, a progression pattern is typically interpreted as evidence of an historical path of colonization. A phylogenetic estimate that allies early branches with older islands and more distal branches with younger islands suggests that a lineage has been established within a given archipelago because the earliest islands were habitable by that taxon. Subsequent colonizations occur coincident with the emergence of younger islands, initiating new communities at progressively younger times. In archipelagos that meet conditions for progression, codistributed taxa with common phylogeographic histories therefore can be studied as part of their communities on specific islands in the sequence, representing slices of time in the history of the archipelago. Thus, a progression pattern provides snapshots of communities through time, for as long as the oldest islands have been supporting such communities.

The history of community assembly over evolutionary time also may be informed by the geographical polarity evident in the progression pattern of a lineage. In an archipelago with a linear spatial and chronological sequence, like those often present in hotspot regions, a simple progression pattern is coincident with a unidirectional path of colonization from older to younger islands. A possible explanation for this is that a regular supply of propagules on older islands maintains the constant potential for colonization as new habitat emerges with the origin of a new island. This dynamic suggests that once new habitat is available (on progressively younger islands), colonists from spatially proximate, older habitat have the potential to arrive, establish, and for many such radiations, diverge.

Biotic Resistance on an Evolutionary Timescale

Although a progression rule may be a valuable tool to understand the evolution of community assembly, there is also the possibility that community assembly may help us understand the development of the progression rule. One of the more prominent features of the progression pattern, and oceanic colonization pathways in general, is that they are largely unidirectional (Waters et al., 2013). In hotspot regions, the chronological sequence of the archipelago landscape promotes a progression pattern because a probable source of colonists is available to establish the next stepping stone in the series once the next island habitat emerges. Thus, the initial progression arising from older to younger island colonizations seems relatively easy to explain.

On the other hand, dispersal vectors (wind, ocean currents. or biological vehicles) (Gillespie et al., 2012) are not expected to promote a unidirectional migration route (e.g., in Hawai'i, complex wind currents have the potential to bring propagules in a variety of directions, and tropical storm tracks, in particular, at times flow in a reverse direction to island age). Indeed, radiations that evidently postdate the Hawaiian Archipelago formation [e.g., the spider genus *Havaika* (Arnedo and Gillespie, 2006); the plant genus *Tetramolopium* (Lowrey, 1995); some Hawaiian birds (Fleischer and McIntosh, 2001)] show colonization routes at odds with the chronosequence, which suggests that there is no physical feature dictating an older to younger colonization route after the islands exist.

Although an early progression pattern arising from the initial colonization opportunities in hot spot archipelagos might often be expected, the unpredictable nature of dispersal might be expected to erase the pattern over evolutionary time. Drawing on ideas dating back to MacArthur and Wilson (1967), it has been argued that an apparent unidirectional pathway of colonization toward younger islands, and by extension the progression rule, may be bolstered by niche preemption, a type of priority effect (Gillespie and Roderick, 2002; Fukami et al., 2007; Gillespie and Baldwin, 2009; Waters et al., 2013; Fukami, 2015).

Priority effects, which arise from the impact that species have on one another within a community, depend on the order of arrival in the community (Fukami, 2015). By this process, once the first colonist (a founder) establishes in a newly available habitat, it soon monopolizes resources in the critical dimensions of the niche and blocks subsequent propagules from establishing. Key to the concept of priority effect is order of arrival. Although priority effects can be positive or negative (Fukami, 2015), this manifestation of the priority effect is negative, yielding advantage to the first colonist to arrive: once a niche has been filled, it is more difficult for ecologically similar individuals to enter (MacArthur and Wilson, 1967).

Priority effects would have an impact on phylogeographic patterns potentially in two ways. First, and most dramatically, patterns that suggest older to younger island colonization would unfold over evolutionary time as a hotspot archipelago forms in a spatiotemporal sequence. On each successively younger island, initial (early) colonists would arrive from the most proximate older island, establish in an empty habitat, and increase in population size and density, that is, preempting the relevant niche. At any point in the future, ecologically similar propagules from younger islands backcolonizing to older islands would be second to arrive (late invaders), suffer a disadvantage, and fail to establish due to arriving in the preempted niche (Gillespie and Roderick, 2002). Thus, the ecological consequences of niche preemption by the early colonists is that backmigrations are discouraged from establishing. Second, niche preemption may further accentuate the initial older to younger island progression pattern by bolstering the genetic legacy of the early colonists via a "founder takes all" dynamic, whereby repeated older to younger island colonizations are likewise thwarted, again due to niche preemption (Waters et al., 2013). Both of these scenarios suggest that ecological dynamics and species interactions are important in giving rise to progression. In other words, to explain features of a progression rule pattern, particularly for a species radiation where all available habitat has been colonized (suggesting propagules are not rare), but speciation has occurred in concert with interisland colonization (suggesting that gene flow has terminated), priority effects may be critical.

Moreover, the ability of new colonists to enter a community is frequently thought to decline as the diversity of species already in the community increases; the established species are considered to dictate biotic resistance against further colonization or invasion (Chapin et al., 2000; Kennedy et al., 2002; Hooper et al., 2005). Thus, biotic resistance results from a complementarity effect; the larger the number of species in a community, the more resource used, in turn creating more community resistance to new invaders. Diversity can also generate biotic resistance through a selection effect under which the most effective competitors, such as those most similar to focal invaders, are more likely to be present in more diverse mixes (Chapin et al., 2000; Loreau and Hector, 2001). Priority effects can play a synergistic role in each of these dynamics. Curiously, although the consequences of priority effects are well understood over evolutionary timescales, the opposite appears to be true at ecological timescales.

Biotic Turnover on an Ecological Timescale

MacArthur and Wilson (1967) proposed that species richness on islands is the result of a dynamic balance between stochastic immigrations and extinctions. Their thinking was a radical shift from most community assembly theory with rejection of a deterministic equilibrium driven by species interactions. Today, the proximity to the mainland, the size of the mainland species pool, and the area of the island are hypothesized to be the three main factors explaining species richness on islands. The equilibrium theory of island biogeography assumes a dynamic equilibrium with continual species turnover, because of colonization and extinction events that constantly modify community composition, a proposition subsequently developed in metapopulation ecology (Hanski, 1999). Thus, biotic turnover prevents genetic differentiation from accumulating, and thereby continually erases the development of a progression pattern.

How Do We Reconcile the Concepts of Biotic Turnover with Biotic Resistance?

Based on many (perhaps most) models of ecological community assembly, biotic resistance seems unlikely. So why, then, is biotic resistance so apparent on an evolutionary timescale (Strauss et al., 2006)? Put another way, how does ecological turnover (and associated ecological processes) give way to the biotic resistance and community "lockup" dynamics that might sustain a progression pattern over evolutionary timescales?

Possible hypotheses to explain community lockup dynamics that maintain progression patterns might be as follows. (*i*) Species and/or communities might be impervious to subsequent invasion due to "high-density blocking" (Waters et al., 2013). This hypothesis posits that, although additional colonizations occur (e.g., through backmigration), numerical dominance of the early, resident species deters establishment or dilutes the genetic contributions of late colonists with similar niche requirements. A prediction of this hypothesis is that the more similar late colonists are to resident species, the less likely they are to successfully invade. A genetic manifestation of this idea is that the distinct alleles of late invaders that manage to hybridize (admix) with residents have a high likelihood of extinction, thereby obscuring repeated- and backcolonization events. (*ii*) Late invaders suffer a mating disadvantage, or mate with residents but have no genetic impact, because they either fail to reproduce or their offspring have low fitness. (*iii.a*) Early invaders have adapted to conditions specific to each island, so that late invaders are always inferior competitors, or (*iii.b*) early colonizers undergo ecological release, resulting in extensive niche overlap among members of the early community; over time, and as the environment becomes more stable, they may specialize, placing subsequent colonizers at a disadvantage. Alternatively, (*iv*) after some initial turnover, the combined island communities lock in at a state where "by chance" no island can be invaded. Although it is difficult to separate these possible explanations (and they may not be mutually exclusive), genomic data promise an opportunity at least to test whether priority effects are real, and if so, what consequences would arise for genetic admixture and introgression.

THE PROGRESSION RULE AND THE MOSAIC GENOME

To address the lockup hypothesis will require a multidisciplinary approach. However, the first explanation suggested above, that late invaders hybridize with species in residence and their genes admix into the resident population, falls in the domain of comparative phylogeography. This hypothesis is predicated on the idea of a mosaic genome where specific regions can have separate historical identities (Baum and Shaw, 1995; Maddison, 1995; Shaw, 2001; Wu, 2001; Seehausen et al., 2014), some reflecting a history of progression, and others reflecting additional demographic complexities (such as backmigrations). Teasing apart complex histories within the same species lineage requires separate analyses of distinct hereditary units. Whether sufficient data exist to mark these separate histories is an empirical issue, but to be sure, if we combine data across data partitions that represent distinctive histories, we will lose insights into these complexities.

In a recent phylogeographic context, it is possible for different gene partitions to reveal conflicting but true histories that manifest more than one of these progression patterns within the same lineage. Such situations are extremely valuable because they suggest a multidimensional insight into past demography of the lineage.

A potential example comes from the Hawaiian cricket genus *Laupala*, a morphologically cryptic group of 38 flightless, single-island endemic, species (Otte, 1994). Multiple datasets reveal a compelling case for progressive clades and grades in *Laupala*, the most resolved of which emerges from a nuclear (presumably) phylogeny based on amplified fragment length polymorphisms (AFLPs; Mendelson and Shaw, 2005) (Fig. 4.3*A*). The phylogeny is well supported, is concordant with the species taxonomy, and is consistent with a less-resolved nuclear sequence phylogeny (Shaw, 2002). The area cladogram shows that the group began on Kaua'i (or an older island) and split into two species groups once on the younger O'ahu. Subsequent diversification occurred via fine-scale progression to increasingly younger islands, colonizing first either Moloka'i (absent in one group) or West Maui, then East Maui, and, last, Hawai'i Island. Thus, the nuclear phylogeny suggests a unidirectional pattern with rare interisland migrations.

However, the unidirectional pattern of colonization is contradicted by the mtDNA phylogeny (Fig. 4.3*B*). Although there is a general pattern of progression in the mtDNA area cladogram as well, there is extensive conflict between an mtDNA tree and both the nuclear DNA trees (Shaw, 2002). Importantly, the phylogeographic patterns in the mtDNA include two potential backmigrations between Hawai'i and Maui and one additional backmigration between Maui and Moloka'i. These backmigrations are between neighboring islands, a trend shared by other taxa where back-

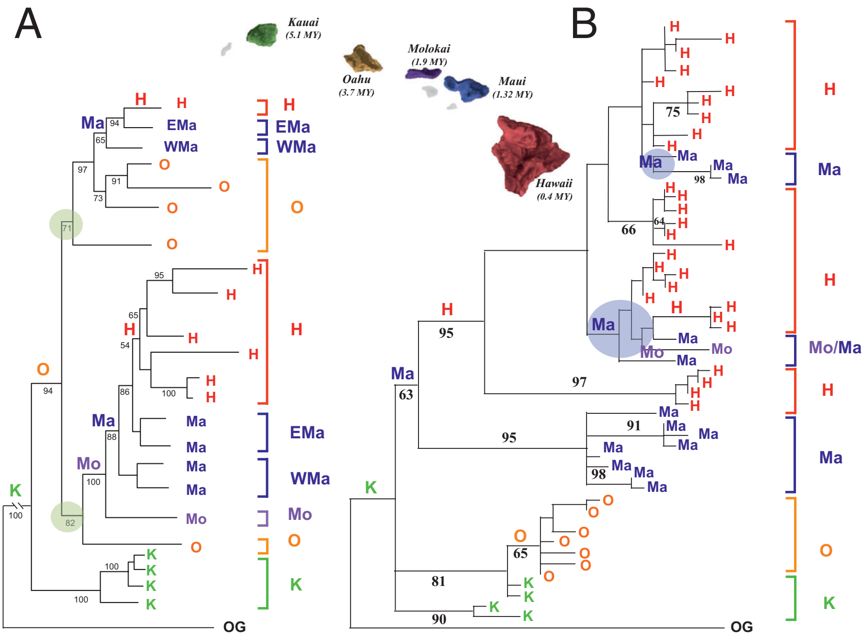


FIGURE 4.3 Progression and the mosaic genome. (*A*) Area cladogram derived from the nuclear AFLP phylogeny of *Laupala* (Mendelson and Shaw, 2005), a genus of flightless crickets endemic to Hawai'i. Two species groups are shown (green discs overlaying resolved nodes). An overall progression is evident, with two progression subpatterns in each species group. Minimum interisland migrations are marked by K (Kaua'i), O (O'ahu), Mo (Moloka'i), Ma (Maui), and H (Hawai'i). Additional abbreviations include OG (outgroup), EMa (East Maui), and WMa (West Maui). A neighbor joining tree based on Nei-Li distances is shown. Terminal taxa are species represented by 2–10 individuals color coded to the island to which they are endemic. Bootstrap values are shown for 1,000 neighbor joining replicates below branches. (*B*) An area cladogram derived from the mtDNA phylogeny of *Laupala* (Shaw, 2002). Shown is the maximum parsimony tree with the highest likelihood score; terminal taxa are unique sequences from concatenated 12s, 16s, and tRNA^{val} regions. The mtDNA topology is not congruent with the AFLP tree, and species groups are not evident in the mtDNA tree. An overall progression is evident, with two backmigrations marked by blue discs. One possible minimum interisland migration scenario is shown, marked by K (Kaua'i), O (O'ahu), Mo (Moloka'i), Ma (Maui), and H (Hawai'i). Bootstrap values are shown for 1,000 parsimony replicates below branches.

migrations occur over a pattern of progression (Funk and Wagner, 1995a). In short, the mtDNA evidence suggests that there has been considerably more interisland movement than is revealed by the nuclear partitions.

A recent study of the flightless, species-rich weevil *Laparocerus* (Faria et al., 2016) from the Canary and Madeira Islands comes to a similar conclusion. The mtDNA and nuclear gene trees sampled from *Laparocerus* show conflicting topologies, similar to *Laupala*, along with a progression from older to younger islands evident in both data partitions. Like *Laupala*, the nuclear phylogeographic history is simpler, whereas the mtDNA phylogeographic history captures considerably more interisland movement. Although it is unclear in the case of *Laupala*, the nuclear data show evidence of subsequent admixture in cases of multiple colonization in *Laparocerus*. Importantly, in both systems, the patterns of mtDNA variation do not conform to expectations of incomplete lineage sorting, and instead appear to be informative about interisland colonization patterns.

Recent phylogeographic studies of Hawaiian planthoppers and spiders using next-generation sequencing approaches are starting to reveal similar patterns of higher-than-expected movement, at least between the younger islands. Together, the data suggest that there is some movement between islands, at least among younger islands (Roderick et al., 2012). The promise of these new genetic data technologies is that the heterogeneity of progression patterns can be fully investigated, and potentially lead to an understanding of the causes of such heterogeneous genetic signatures.

DISCUSSION

The progression rule is one of the most pervasive phylogeographic patterns yet documented, at least for the well-studied remote archipelago of Hawai'i. There is considerable evidence accumulating from two other well-studied archipelagos, the Canary Islands (Juan et al., 2000) and the Galapagos (Parent et al., 2008; Faria et al., 2016), despite the complicating factors of spatial nonlinearity and greater proximity to continents. Additional evidence is accumulating for progression in other remote archipelagos as well (summarized here for the Australs and the Marquesas) (Cibois et al., 2011). Evidence for progression, thus far, is mixed in the Azores, presumably due to the geological complexity of the hotspot (Carvalho et al., 2015), and the Society Islands, perhaps due to the close proximity to much older islands, meaning that there were other islands that served as a source of colonists, or a "bridge" between other archipelagos to the younger islands of the Society chain.

Progression is a feature that emerges over evolutionary time and can only develop once a community transitions from biotic turnover (manifest on an ecological timescale) to biotic resistance (on an evolutionary timescale). To a large degree, the transition from biotic turnover to biotic resistance mirrors the transition from population genetic cohesion to the genetic differentiation that characterizes speciation. In species formation, eventually, a nascent species closes genetic connections with its past (represented by its sister species) and persists as an independent evolutionary entity. Likewise, in the transition from ecological turnover to biotic resistance, a new community forms. Whereas formerly, the community was in a dynamic state of revision due to immigration and extinction (under some theories), eventually, the community resists repeated immigration of colonists. The speciation process occurs within the context of this community transition, and embodies the zone of radiation.

The geographic and chronological settings of remote, hotspot archipelagos offer conditions under which progression patterns can arise. However, the ecological features of the organisms involved play an important role in the subsequent patterns of diversification, and the rate at which they evolve. Key are the rates of dispersal and likelihood of colonization and establishment (Buckley et al., 2013). Obviously, if little to no dispersal occurred, a progression pattern would be unlikely to arise as the species would likely remain in the ancestral habitat and not radiate (e.g., see potential examples reviewed in Keeley and Funk, 2011). Likewise, if dispersal and subsequent gene flow facilitated panmixia, neither speciation nor progression would be expected to arise because biotic turnover would prevail. However, if dispersal were infrequent but predictable (the likely condition for the majority of taxa), colonization would occur as new island habitats become available. Gene flow would be reduced, and depending on its magnitude, would not continue to homogenize gene frequencies between founder and source.

Once the founder population, now the resident species, established and has filled the available niche, priority effects could come into play by a number of possible mechanisms: (*i*) The community might be resistant to subsequent invasion due to the monopolization of resources by the numerically dominant resident species; the more similar late colonists are to resident species, the less likely they are to successfully invade. (*ii*) Late colonists may suffer incompatible mating encounters with the numerically dominant resident species, leading to a failure to reproduce. (*iii*) The resident species may have adapted to local conditions, giving it a competitive edge over later colonists. The last mechanism involves natural selection and competitive exclusion, whereas numerical dominance drives the first, and may contribute to the second interaction. These possible mechanisms could, in theory, be tested with appropriate experiments.

Depending on the degree of differentiation, the late colonist may nonetheless hybridize with the resident species and its genes may diffuse into the resident species gene pool, leaving some trace of its attempt at establishment. As discussed in the two examples described earlier, mosaic histories contained within the genome of a given lineage may harbor evidence regarding how a progression pattern evolves in this intermediate dispersal domain. It seems unlikely that species radiations manifesting a progression have occurred in the clean manner that a phylogeny might suggest. Even in the most straightforward cases, such as the patterns of progressive clades and grades in the AFLP phylogeographic pattern in *Laupala* crickets, additional data cautions us from concluding that interisland migrations do not occur. More efforts are needed to examine contrasting evidence of progression patterns as they may provide insights into how the pattern develops. If we can understand the development of progression, we may begin to understand how ecological turnover (where progression would not arise) gives way to biotic resistance (where progression could arise).

As elucidated by Fukami (2015), priority effects by niche preemption are most likely to occur when two species show highly overlapping resource use, the first colonist has a high impact on the overlapping niche dimensions, and when the growth rate of the late colonist is heavily dependent on the environment. Thus, we might expect that ecological shifts create opportunities for late colonists to escape the impact of priority effects as a result of exploring new niche dimensions. Highly dynamic adaptive radiations that appear to violate the progression rule via backmigration, such as some branches of the Hawaiian picture-winged *Drosophila* radiation (Magnacca and Price, 2015), provide an opportunity to test this idea. Perhaps herein lies the explanation for the finding that adaptive radiations in the Galapagos were not found to adhere to the progression rule (Parent et al., 2008). Likewise, species that have more plastic attributes to their resource use might have a heightened immunity to priority effects, which would act to discourage the development of a progression pattern in the lineage. Better characterization of ecological traits, reproductive behaviors, divergence times, and genetic admixture among lineages of island radiations should allow for more rigorous evaluation of priority effects on the development of progression within a phylogenetic context.

The study of community assembly in and beyond the radiation zone is exciting in part because it brings together two fairly disparate disciplines: the ecological study of community structure and the evolutionary study of the origin of species. The presence of progression, and its more nuanced manifestation revealed by the mosaic genome, is fortunate for the study of community assembly because the pattern provides a temporal framework for both ecological and evolutionary studies of communities and their interrelationship. The key importance of the progression pattern is that multiple lineages are establishing and assembling, interacting and adapting, over a similar time frame that plays out over extended evolutionary time. Thus, we can measure ecological metrics at different time slices of the community assembly process to find how properties (species diversity, abundance, body size distributions, trophic interactions) change over time (Loreau et al., 2002; Harmon et al., 2009), and how the origin of new species affects these properties. Moreover, as genomic data become available across multiple lineages that appear to follow a progression in a given system (Gillespie, 2015), we are gaining insight into how taxa differ in mode, rates, and patterns of establishment and diversification (Rominger et al., 2016). Integrating multidimensional datasets across stages of the progression will allow us to understand how interactions develop and evolve and the importance of such interactions in dictating properties of stability, turnover, and the evolution of biotic resistance in a community. Effectively, with multiple lineages being formed over the same time frame, we can examine the feedback between ecology and evolution and hence generate insights into the processes involved in the formation and loss of biodiversity.

ACKNOWLEDGMENTS

We thank Francisco Ayala, John Avise, and Brian Bowen for the invitation to participate in the symposium and special issue. We thank two reviewers for comments that helped improve the manuscript. We also thank our students and colleagues of the Dimensions of Hawaiian Biodiversity team for many stimulating discussions. This work was supported by National Science Foundation (NSF) Division of Environmental Biology Grants 1241060 (to K.L.S.) and 1241253 (to R.G.G.).

Part II

COMPARATIVE PHYLOGEOGRAPHY IN A GENOMIC SENSE

Throughout the early phylogeographic era, cytoplasmic genomes (mitochondrial DNA in animals and chloroplast DNA in plants) provided the bulk of empirical genetic information for phylogeographic reconstructions. However, a long-appreciated fact is that these cytoplasmic genomes represent only a minuscule fraction of a species' total hereditary pedigree, the vast majority of which is ensconced in nuclear DNA. Early attempts to extract useful phylogeographic information from the nuclear genome met mostly with failure, due to technical difficulties coupled with sex-based genetic recombination. In recent years, this situation has changed (to an arguable degree) due to the ongoing technological revolutions in next-generation sequencing and "big-data" genomics. How might phylogeographic inferences be impacted as new technologies extract more and more genetic information from loci in the recombining nuclear genome?

Each species has its own true phylogeographic past (extended intraspecific pedigree) through which its various loci have been transmitted and which therefore constitute a plethora of gene genealogies. For any species, genealogy and historical population demography are like opposite sides of the same coin: intimately connected. Coalescent theory offers a robust conceptual framework for translating the empirical phylogeographic structure of a "gene tree" (such as that provided by mtDNA) into inferences about a population's demography history. Although the population pedigree of any species in effect contains (or consists of) multitudinous quasi-independent nuclear gene trees, the actual degree to which a given pedigree constrains the topologies of multiple unlinked loci has received rather scant attention. In Chapter 5, John Wakeley and colleagues address this issue using population genetic models and computer simulations. By focusing attention on extreme demographic events (the occurrence of very large families in a pedigree, and on strong selective sweeps in the population's recent past), the authors conclude that "only rather extreme versions of such events can be expected to structure population pedigrees in such a way that unlinked loci will show deviations from the standard predictions of population genetics, which average over population pedigrees."

Rohan Mehta and colleagues (Chapter 6) continue this general theme of the fundamental distinction between gene trees and population (or species) trees, by addressing the probability of monophyly of a gene lineage on a species tree. Although such models (based on neutral coalescent theory) have long been employed to calculate the probability that a set of gene lineages is reciprocally monophyletic under the simplest case of a pair of sister taxa, the current authors extend such analyses to probabilities of gene-tree monophyly for genetic studies that span arbitrary numbers of multiple isolated populations or species.

One of several new types of nuclear data stemming from next-generation sequencing involves the recovery of SNPs (single nucleotide polymorphisms) from many thousands of unlinked nuclear genomic regions, even in nonmodel species. SNPs are increasingly being used to supplement more traditional phylogeographic datasets based on cytoplasmic genomic sequences or allelic profiles at relatively small numbers of microsatellite loci. In Chapter 7, Maria Thomé and Bryan Carstens employ a case-history approach (involving Brazilian frogs) to illustrate how such molecular information can be used to estimate historical population demographic parameters (such as population size and gene flow) under a wide variety of evolutionary models. Their take-home message is that an objective approach to phylogeographic inference should entail calculating the probability of multiple demographic models given the data and then subsequently ranking these models using information theory. The chapter's framework also allows the authors to express their own views on the ever-changing epistemology of phylogeographic inference.

Anna Papadopoulou and Lacey Knowles (Chapter 8) expand on the topic of phylogeographic models by tracing and critiquing the historical emphasis on genealogical concordance in comparative phylogeography. Such concordance generally refers to shared phylogeographic patterns, either across multiple loci within extant species or across particular genes of multiple codistributed species. The authors question whether concordance in general is a uniformly useful criterion for evaluating alternative phylogeographic hypotheses, by emphasizing taxon-specific traits that may predict concordance or discordance among datasets and species. The authors bolster their reservations with case studies illustrating the many possible ways that genealogical discordance (the antithesis of concordance) can arise in particular situations.

Scott Edwards and colleagues (Chapter 9) conclude this section of the book by providing an overview of the impact of cutting-edge molecular technologies (such as various expressions of next-generation sequencing) on the trajectory of the fields of phylogeography and phylogenetics in the genomics era. As empirically illustrated by the authors' comparative genetic research spanning diverse vertebrate taxa across northern Australia, the emerging discipline of phylogenomics will call for a greater appreciation of reticulation during the evolutionary process, both within genomes in the form of genetic recombination, and across populations and species in the forms of gene flow and introgression. In this important sense, the arenas of comparative phylogeography and interspecific phylogenetics can again be seen as lying along a conceptual continuum of historical evolutionary genetic phenomena, from population-level separations to deeper organismal divergences.

5

Effects of the Population Pedigree on Genetic Signatures of Historical Demographic Events

JOHN WAKELEY,*† LÉANDRA KING,* AND
PETER R. WILTON*

Genetic variation among loci in the genomes of diploid biparental organisms is the result of mutation and genetic transmission through the genealogy, or population pedigree, of the species. We explore the consequences of this for patterns of variation at unlinked loci for two kinds of demographic events: the occurrence of a very large family or a strong selective sweep that occurred in the recent past. The results indicate that only rather extreme versions of such events can be expected to structure population pedigrees in such a way that unlinked loci will show deviations from the standard predictions of population genetics, which average over population pedigrees. The results also suggest that large samples of individuals and loci increase the chance of picking up signatures of these events, and that very large families may have a unique signature in terms of sample distributions of mutant alleles.

The degree to which a sample may be considered representative of a population is a fundamental question in any application of statistics. In the complicated world of evolutionary and population genetics, where it is sometimes not even clear which aspects of ancestry or data should be modeled as random processes, questions of this sort

---

*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138. †To whom correspondence should be addressed. Email: wakeley@fas.harvard.edu.

assume greater significance still, and simple mistakes can have drastic effects on inference. These issues are brought to the fore in the field of phylogeography, which was first developed by Avise and colleagues in the 1980s after the introduction of genotyping technologies into evolutionary biology and which takes as its starting point the fact that hierarchical patterns of genetic variation contain information about the locations of populations and species in the past, as well as their relative population sizes and other factors of biological interest (Avise, 2000).

The core debate about randomness in the subsequent development of phylogeography was about whether individual gene genealogies should be treated as outcomes of highly variable random processes, which need to be modeled, or as simple observations from which conclusions about the past may be drawn more or less directly (Knowles and Maddison, 2002; Hey and Machado, 2003; Templeton, 2008, 2009; Nielsen and Beaumont, 2009). There will be cases in which the size and shape of a single gene genealogy contain substantial information about population-level or intraspecific ancestry but, as noted in a recent review (Bloomquist et al., 2010), this debate has come down on the side of modeling. The reasons for this are that gene genealogies are in fact the results of random processes, likely at the population level but certainly at the level of Mendelian genetic transmission, and that it is not known a priori whether a given set of data comes from one of those cases in which gene genealogies are individually informative (Ewens, 1990; Hudson, 1990; Rosenberg and Nordborg, 2002). Although this particular issue may be considered settled, debates about the proper application of random models in phylogeography continue to arise (Lemey et al., 2009; De Maio et al., 2015).

We consider an additional question about the application of random models that has received comparatively little attention either in phylogeography or population genetics. Namely, what is the extent to which genealogies in the family sense—also known as organismal pedigrees (Ball et al., 1990) or population pedigrees (Wakeley et al., 2012)—constrain gene genealogies and thus genetic variation? Two points distinguish this question from the initial core debate about randomness in phylogeography.

First, whereas in phylogeography the focus has been on the undesirable effects of making inferences conditional on a single gene genealogy estimated from data, here it is on the validity of inferences based on standard population-genetic models that average over population pedigrees when in fact there is only one. It turns out that in relatively large well-mixed populations with constant demography over time, the predictions of standard models are generally quite accurate even though they involve this conceptual error (Ball et al., 1990; Wakeley et al., 2012). The second point is that the variation we are interested in here is variation among loci for a set of sampled individuals. Even though the population pedigree may itself be the outcome of a random process, all loci in the genome share the same pedigree. The population pedigree should thus be considered a given, fixed quantity because peculiarities of genetic variation among loci in the genome may be due to peculiarities in the pedigree.

Work on the effects of population pedigrees began in 1990 with Ball et al. (1990), who made the fundamental observation that standard-model predictions for a single well-mixed population fit the distributions of pairwise measures of diversity among independent loci on a given pedigree surprisingly well. Follow-up work on subdivided populations came to similar conclusions but also illustrated that sampling small numbers of transmission pathways through a pedigree can give results quite different from corresponding standard-model predictions (Wollenberg and Avise, 1998) and that pedigrees can substantially affect the probabilities of gene-tree topologies in isolation-by-distance migration models (Kuo and Avise, 2008). These works used simulations to generate pedigrees and to model genetic transmission within each pedigree.

Chang (1999) explored two key aspects of ancestry within population pedigrees analytically, proving for a population of N individuals that (*i*) the most recent common ancestor of all present-day individuals in the pedigree sense (i.e., an individual through which all present-day individuals are cousins) will typically be observed at $\log_2 (N)$ generations in the past, and (*ii*) by about $1.77 \log_2 (N)$ generations in the past, the ancestries of all present-day individuals overlap completely. Underlying these results is the key fact that the number of pedigree ancestors of an individual grows by a factor of two each generation. Rohde et al. (2004) used simulations and analysis of human population structure and history to suggest that our ancestries overlap in these same ways only a few thousand years ago. The $\log_2 (N)$-generation timescale for pedigree ancestry is dramatically shorter than the N-generation timescale for common ancestry in the genetic sense (Hudson, 1990), which for humans corresponds to hundreds of thousands of years (e.g., Schiffels and Durbin, 2014).

Subsequent work using both analysis and simulations has emphasized the rapid approach to equilibrium of shared ancestry in pedigrees. Reproductive values of individuals across the population (Fisher, 1930), which are proportional to the probabilities that a genetic lineage sampled randomly today traces back to each individual in a given past generation, reach a stationary distribution on this same $\log_2 (N)$ timescale (Derrida et al., 2000; Barton and Etheridge, 2011). Correspondingly, deviations from the predictions of standard population-genetic models, in particular those of coalescent theory, are evident for distributions of times to common ancestry within the recent $\log_2 (N)$ generations but disappear as lineages are traced into the more distant past (Wakeley et al., 2012).

Pedigrees are, of course, a mainstay of medical genetics, where they allow powerful inferences about the genetics of human disease (Thompson, 1975). These are not population pedigrees, which cover entire populations or species for all times, but partial recent pedigrees of sampled individuals. Pedigree analyses of this sort are being applied to a growing number of natural populations, ones for which patterns of reproductive relationship are known, to disentangle the genetics of complex traits and understand patterns and consequences of inbreeding (Pemberton, 2008). Observed partial pedigrees have also been used to make inferences about recent historical demography—for example, the French settlement of Quebec (Moreau et al., 2011)—directly from pedigree shape without genetics.

Population pedigrees have less frequently made their way into the models of population genetics. Beyond the examples above (Ball et al., 1990; Wollenberg and Avise, 1998; Derrida et al., 2000; Kuo and Avise, 2008; Barton and Etheridge, 2011; Wakeley et al., 2012), they have been invoked to study the length distribution of admixture tracts in a descendant population (Liang and Nielsen, 2014) as well as to describe the ways in which ancestors in the pedigree sense are numerous, whereas the genetic ancestors among them are comparatively few (Matsen and Evans, 2008; Gravel and Steel, 2015).

Here, we use simulations to assess the potential for two kinds of demographic events to alter the shape of population pedigrees so dramatically that they have marked signatures on genetic variation across the genome, specifically among independently segregating loci without intralocus recombination. We begin by emphasizing the assumptions of standard population-genetic models, which determine how they should be applied, and the resulting conceptual error involved in using standard models to explain variation across the genome in diploid biparental organisms. The first kind of demographic event we consider is the case of a very large family at some generation in the past. The second is the introduction and sweep through the population of a strongly advantageous mutant allele. In both cases, we ask whether data from unlinked loci will deviate from standard predictions for the same demographies without these special events. We restrict our attention to well-mixed populations. This provides a baseline set of results against which subsequent work (e.g., on geographically structured populations) may be compared.

TWO CONCEPTUALLY DIFFERENT RANDOM EXPERIMENTS

One of the most familiar results of population genetics is the probability there will be j copies of an allele in the next generation given there are currently i copies of it in a population of N individuals,

$$P(j \mid i) = \binom{2N}{j}\left(\frac{i}{2N}\right)^{j}\left(\frac{2N-i}{2N}\right)^{2N-j}, \qquad \textbf{[5.1]}$$

which is derived from the diploid monecious Wright–Fisher model (Fisher, 1930; Wright, 1931) with the possibility of selfing as a result of random mating or random union of gametes (Watterson, 1970). There is no reference to the specific outcome of reproduction among the N individuals (i.e., to what could be called the single-generation pedigree) because Eq. 5.1 is an average over all possible outcomes of reproduction. Using the theory of Markov processes or diffusion approximations for large N, predictions over longer periods of time can be derived from Eq. 5.1 (Ewens, 2004). Such predictions about the probabilities of outcomes of evolution from a given starting point can be compared directly to the results of laboratory experiments, in which allele frequencies are measured but pedigrees typically are not.

The classic experiments of Buri (1956), in which the entire evolutionary process was repeated independently a large number of times,

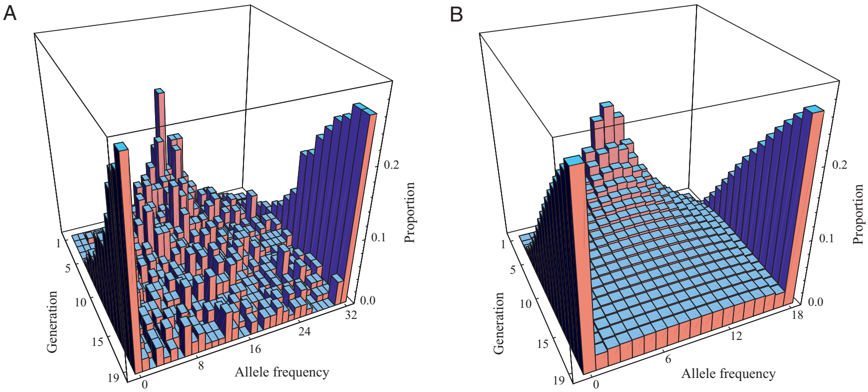


FIGURE 5.1 (*A*) Data from series I (table 13 of Buri, 1956) for generations 1–19. In each generation, the proportion for each allele frequency is the fraction of the total 107 populations that showed that particular frequency. Generation 0 is not depicted but would have allele frequency equal to 16 and proportion equal to 1. (*B*) Corresponding theoretical prediction using Eq. 5.1 iteratively, but with the effective population size $N_e = 9$ estimated by Buri (1956) rather than $N = 16$ as in the experiment. With $N = 16$, only about 23% of the populations would have been monomorphic by generation 19 instead of the ~54% observed in the experiment.

provide the appropriate sort of data. In one experiment, Buri recorded allele frequencies of a selectively neutral mutation (bw^{75}) at the *brown* (eye-color) locus in *Drosophila melanogaster* over 19 generations in 107 replicate laboratory populations. Populations were founded each generation by a random sample of eight male and eight female offspring of the adult flies of the previous generation. Every population began with a relative frequency of 0.5, or 16 copies of the mutant allele out of a total of $2N = 32$. The results are displayed in Fig. 5.1*A*, with corresponding predictions providing a fit to the data shown in Fig. 5.1*B*. Over the course of the 19 generations, each population's allele frequency drifted randomly. Some populations became fixed for and others lost the bw^{75} allele. By the end of the experiment, roughly 54% of the populations were monomorphic and the remainder were distributed more or less evenly among the polymorphic allele frequencies.

Now consider another standard population-genetic prediction, in this case for the distribution of the number (K_2) of SNP differences between a pair of sequences at a locus,

$$\begin{aligned} P(K_2 = k) &= \int_0^\infty f_{T_2}(t) P(K_2 - k \mid T_2 = t) dt \\ &= \frac{1}{\theta + 1} \left(\frac{\theta}{\theta + 1} \right)^k \quad k = 0, 1, 2, \ldots. \end{aligned} \qquad \textbf{[5.2]}$$

Eq. 5.2 holds under the infinitely many sites mutation model with parameter $\theta = 4N_e u$, in which N_e is the coalescent effective population size (Sjödin et al., 2005), and without intralocus recombination (Watterson, 1975). The first line of Eq. 5.2 shows how a typical derivation of this result proceeds by conditioning on the underlying, unknown coalescence time (T_2) between the pair of sequences, that is, with $T_2 \sim$ Exponential(1) and $K_2 \mid T_2 = t \sim$ Poisson(θt). Because the distribution of T_2 is obtained starting from the single-generation probability of coalescence which, like Eq. 5.1, is an average over the process of reproduction, the exponential distribution of T_2 is an average over the long-term process of reproduction, or over the population pedigree.

Thus, Eq. 5.2 is an equilibrium result that captures the balance between genetic drift and mutation. It predicts what would be observed if two sequences at a locus were sampled at random from such a population. For most organisms, it is not feasible to perform long-term experiments analogous to those of Buri (1956) to create multiple replicate populations for comparison with Eq. 5.2 or other similar predictions. Instead, these predictions are applied to datasets of multiple loci genotyped in the same set of individuals sampled from a single population (or species). Although this type of application is conceptually wrong because the loci share the pedigree, standard-model predictions match simulated pedigree-coalescent data surprisingly well for large, well-mixed populations (Ball et al., 1990; Wakeley et al., 2012).

An example of this standard type of application is given in table 3 of Sachidanandam et al. (2001), which gives the numbers of loci showing zero, one, two, three, or four SNP differences between pairs of sequences at 12,027 loci ranging in length between 400–700 bp in one of the first major SNP-typing studies in humans. Fig. 5.2 plots these data alongside the corresponding predictions from Eq. 5.2. The coalescent model in Fig. 5.2 and the more sophisticated one in table 3 of Sachidanandam et al. (2001), which takes variation in the lengths of loci and the mutational opportunity among loci into account, can both be rejected using a χ^2 test. However, it is not clear that this is due to the pedigree, because humans deviate from the assumptions of standard models in other ways (e.g., growth and population structure).

This standard type of application is assumed to be appropriate for loci that are far enough apart in the genome (on different chromosomes in the extreme case) that they assort essentially independently into gametes. Whether or not they assort independently, Eq. 5.2 is not the correct predic-

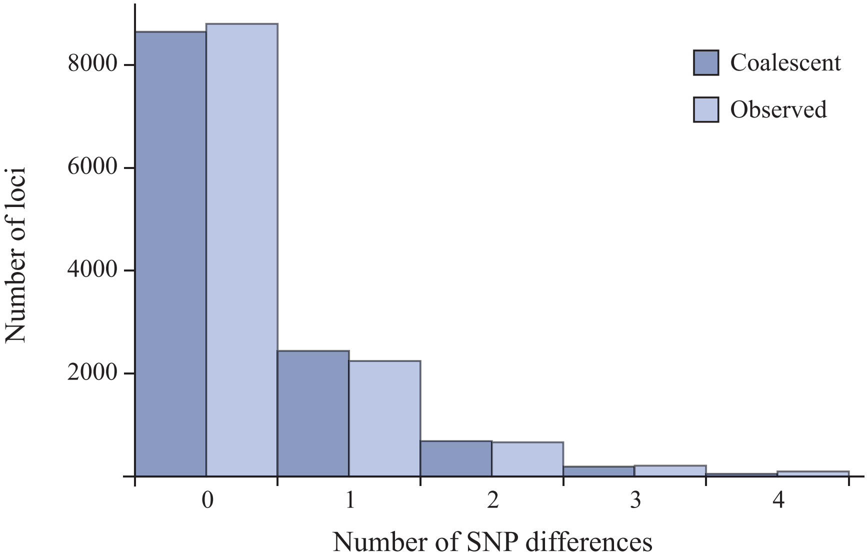


FIGURE 5.2 Observed data from table 3 in Sachidanandam et al. (2001) and coalescent expectations fitted to have the same average number of SNP differences (0.394), with the distribution truncated at four SNP differences for technical reasons described in Sachidanandam et al. (2001).

tion because Eq. 5.2 involves the implicit assumption that the loci do not share the same pedigree. Loci on different chromosomes are independent, but only conditional on the population pedigree. They might collectively show patterns of times to common ancestry or genetic variation that depend on the specific features of the pedigree.

In fact, the population pedigree completely determines the probabilities of coalescence in any given generation. Fig. 5.3 shows a four-generation piece of the Spanish Hapsburg royal family from a study of inbreeding in the demise of this ruling family line (Alvarez G et al., 2009). Two alleles, one sampled from Mary of Portugal and one sampled from Philip II, would have zero chance of coalescing in the previous two generations, then a substantial probability of coalescing in past generation 3. Thus, the probability of coalescence is not constant over time, as assumed in standard models, and it may not be clear whether it should ever be equal to familiar result $P(\text{coal}) \approx 1/(2N_e)$ even under the idealized diploid, monecious Wright–Fisher model.

Simulations for a variety of models of reproduction show that standard predictions, such as the exponential distribution of T_2, are robust to the presences of the shared population pedigree (Ball et al., 1990; Wakeley

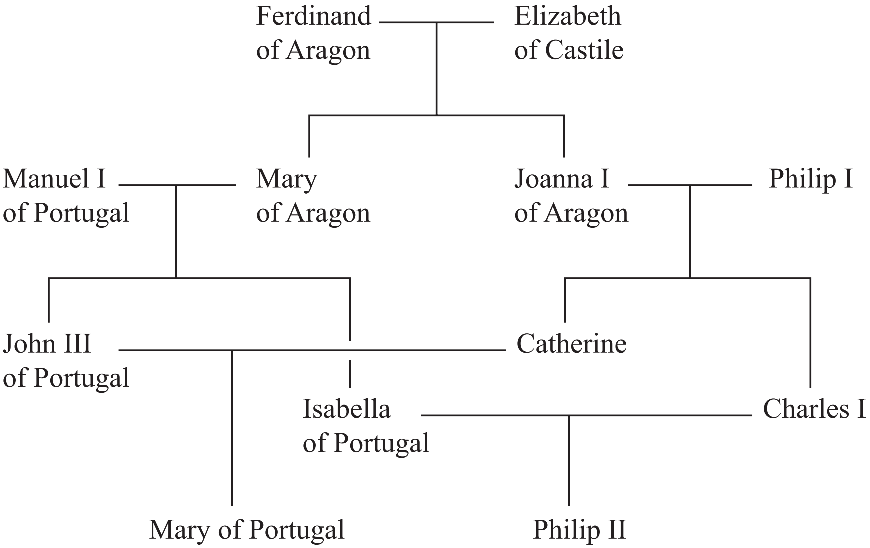


FIGURE 5.3 A small portion of the human population pedigree, from G. Alvarez et al. (2009). Spanish Hapsburg King Charles II, who is not shown but would be three generations below, is inferred to have had an inbreeding coefficient of $F = 0.254$ and could neither rule effectively nor continue the family line.

et al., 2012). One exception is when the sample being analyzed has recent common pedigree ancestors, in which case predictions such as Eq. 5.2 are drastically wrong. However, it is unlikely to sample related individuals from a large population, so the main effect of the shared population pedigree is to make coalescence impossible (as in Fig. 5.3) until the ancestries of the sampled individuals overlap (Wakeley et al., 2012).

In what follows, we consider the effects of extreme pedigrees on distributions of time to coalescence, pairwise SNP differences, and frequencies of mutations in a sample. The results are from simulations of population pedigrees and coalescence of alleles from sampled individuals within pedigrees. In large part, our findings provide further support of the robustness of standard models that average over pedigrees but also suggest that some demographic events might leave signatures detectable in large samples of loci and individuals.

PEDIGREE EFFECTS OF A LARGE FAMILY

An extensive recent study of human Y-chromosome variation (Balaresque et al., 2015) identified a number of descent clusters present at unusually high frequencies in Asia and inferred that these represent the genetic heritages of a corresponding number of highly reproductively successful men. It was surmised that one of these men was Genghis Khan, who had previously been suggested as the source of a particular Y-chromosome haplotype found at ~8% frequency across a large region of Asia (Zerjal et al., 2003). The larger sample and finer-scale geographical sampling seemed to uphold this finding and further revealed substantially higher frequencies of this haplotype in some local populations in central Asia, with one from Middle Kyrgyzstan, for example, showing a sample frequency of ~68% (Balaresque et al., 2015).

We consider a hypothetical, extreme scenario based on these inferences about Genghis Khan, in which a single man has a very large number of children at generation 28 in the past. Details of our simulations are given in *Materials and Methods.* We present results for distributions of pairwise coalescence times among autosomal loci in a pair of individuals, assuming independent assortment but conditional on a single shared population pedigree. We also present results for pairwise SNP differences and site frequencies, for which we use $\theta = 0.5$ per locus and assume the infinitely many sites mutation model without intralocus recombination (Watterson, 1975). Considering the observed population heterozygosity of about 7×10^{-4} in humans (e.g., see Palamara et al., 2015), $\theta = 0.5$ corresponds to loci of length ~700 bp, and an average number of SNP differences between pairs of sequences equal to 0.5. Note that the per-site recombination rate is of a similar order of magnitude as the per-site mutation rate in humans (Reich et al., 2002) and in many other organisms (see table 4.1 in Lynch, 2007). Modeling these relatively short loci, which should have on average only about 0.5 recombination events between a pair of sequences, is one way to minimize the consequences of assuming no intralocus recombination.

Fig. 5.4*A* shows the probabilities of pairwise coalescence, or the proportion of loci expected to coalesce, in each of the past 40 generations

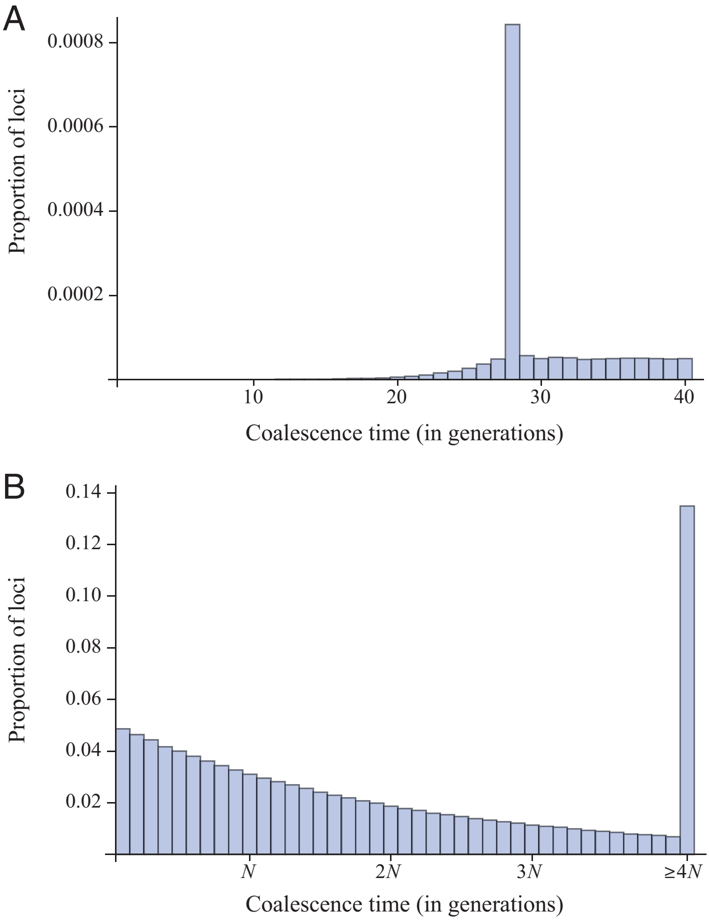


FIGURE 5.4 Simulated distributions of coalescence times conditional on a population pedigree for the case of a large family described in the text, in which the children comprise 8% of the population in generation 27. Each panel is based on a single population pedigree and single pair of sampled individuals. (*A*) Only the most recent generations. (*B*) The whole range of coalescence times on the coalescent timescale of the ancestral population (N = 10,000). Proportions in *A* are estimated based on 10^8 replicates and in *B* from 10^6 replicates.

assuming "Genghis Khan's" children comprise 8% of the population. There is very little coalescence in the most recent generations, 1–20, due the strong population growth assumed, but there would still be little coalescence during this time in a population of constant size (here $N = 10{,}000$). In generation 28, there is a spike in the chance of coalescence. Its height is small, though, because coalescence occurs only when both lineages are among that 8% of the population, both trace back to the father, and they descend from the same allele. Thus, the increase in probability is only $0.08^2 \times (1/2)^2 \times 1/2 = 0.0008$.

Looking at the same scenario over the much longer time frame relevant to coalescence, in Fig. 5.4*B*, this extra mass of coalescence probability has no discernible effect on recent coalescence (leftmost bin in Fig. 5.4*B*) now corresponding to coalescence within the recent $0.1N$, or 1,000, generations. In sum, we cannot expect to observe the effects of even this fairly dramatic demographic event in a large sample of loci from a pair of individuals, which would amount to taking many random draws from the distribution in Fig. 5.4. Fig. 5.4*B* is indistinguishable from the simple coalescent predictions from an exponential distribution with mean 1 corresponding to $2N$ generations.

The situation changes when the children make up 68% of the population. Fig. 5.5*A* shows a dramatic effect even on the overall distribution of coalescence times. In this case the increase in the chance of coalescence is $0.68^2/2^3 = 0.0578$, which roughly doubles the proportion of loci expected to coalesce within the first $0.1N$, or 1,000, generations. We might expect this increase to be observable in data, for example, in pairwise SNP differences. However, for the relatively short (~700 bp or $\theta = 0.5$) loci we model here, a fairly large proportion of loci should be monomorphic even if their coalescence times are greater than $0.1N$ generations. Fig. 5.5*B* compares a simulated distribution of pairwise SNP differences among loci on a single population pedigree for this case to a simulated distribution for a pedigree with the same demography but without any special demographic event. The distributions differ, but it would take more than 8,300 loci to distinguish between them at the 1% level using a χ^2 homogeneity test.

We also investigated the possibility there would be greater power to detect the pedigree effects of a large family using site-frequency data. We again simulated ancestries of very many loci starting from the same set of individuals sampled without replacement from the current generation, only now we sampled 1,000 individuals and followed 1,000 genetic lineages, creating pseudodata for each locus then counting the number of copies of each mutant in the sample. Fig. 5.6 shows these "unfolded" site-frequency distributions (Akashi, 1999) for the case in which the children comprise 8% of the population (Fig. 5.6*A*) and in which they comprise 68% of the population (Fig. 5.6*B*).

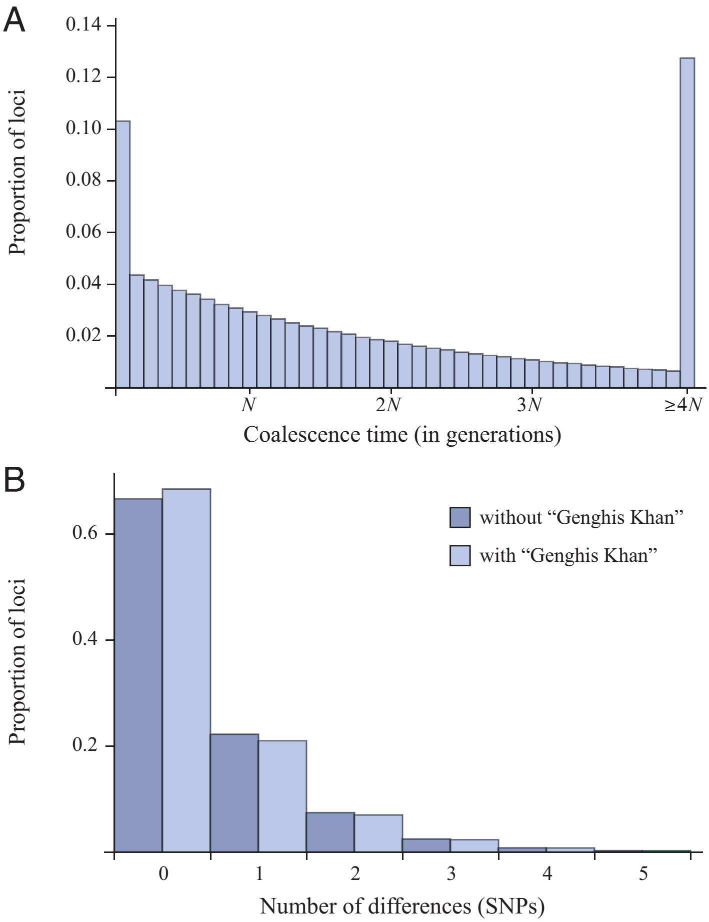


FIGURE 5.5 Simulated distributions of pairwise coalescence conditional on a population pedigree in which the children in generation 27 comprise 68% of the population. Each panel is based on a single population pedigree and single pair of individuals sampled. (*A*) A plot of coalescence times, analogous to Fig. 5.4*B*. (*B*) The distribution of genetic variation among loci with or without the demographic event of such a very large family. Proportions are estimated based on 10^6 replicates.

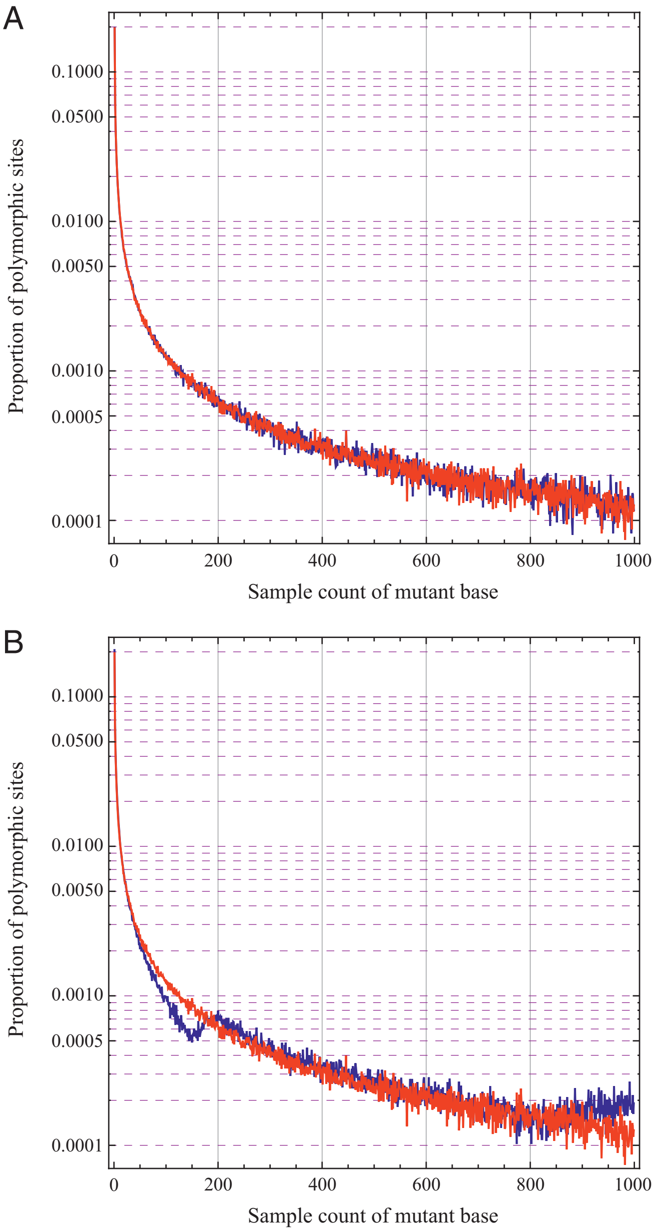
When the children make up 8% of the population, there seems to be no discernible effect on site frequencies, but a striking pattern is observed when the children make up 68% of the population. Differences appear in two parts of the distribution. First, there is a deficit of polymorphic sites at which the mutant is found in about 50–200 copies in the sample. The explanation for this is that many potential branches in the gene genealogy that would have had between roughly 50 and 200 descendants in the sample will be collapsed to zero when bunches of lineages coalesce in "Genghis Khan." Without a large family, these branches would have positive lengths, and mutations on them would produce polymorphisms in these site-frequency classes. In the simulations for Fig. 5.6*B*, an average of 934 lineages remained by generation 27 in the past, so each of the two clusters of coalescent events in "Genghis Khan" involve an average of 943 × 0.68/4 ≈ 159 ancestral lineages. Thus, there is a deficit of branches with roughly 1,000 − 943 = 66 descendants up to the size of these two clusters (159 lineages each). These calculations are based on average numbers of lineages, whereas the simulations in Fig. 5.6 include a great deal of variation in each of these numbers and in patterns of coalescence.

The second effect on the site-frequency distribution is an increase in the number of high-frequency derived mutations. Similar patterns have been ascribed to positive selection (Fay and Wu, 2000), but U-shaped distributions of allele frequencies are observed within local populations subject to migration (Wright, 1931) and are not unexpected when multiple-merger coalescent events can occur (Sargsyan and Wakeley, 2008). We do not have a quantitative explanation of this pattern in Fig. 5.6*B*, but, roughly speaking, it is due to the fact that both of the large clusters may be on one side of the root of the gene genealogy. As described in *Materials and Methods,* we verified the overall pattern of site frequencies for this case using a modified set of standard coalescent simulations.

PEDIGREE EFFECTS OF A SELECTIVE SWEEP

We also investigated the potential of a strong selective sweep to structure the population pedigree in such a way that a genomewide deviation from the predictions of the standard neutral model would be observed. Whereas the genetic effects of selective sweeps are known to be dramatic for loci linked to a locus under selection (Smith and Haigh, 1974; Kaplan et al., 1989; Kim and Stephan, 2002), it is generally understood that unlinked loci are not affected by sweeps. In fact, there is some small effect of a selective sweep even on unlinked loci, which may be attributed to a transient increase of the variance of offspring numbers during a selective sweep (Robertson, 1961; Barton, 1995). To investigate this effect of a sweep as mediated by the population pedigree, we simulated very strong selective sweeps beginning at generation 50 in the past in a population of constant size $N = 10{,}000$, with additive fitness effects of two alleles (*Materials and Methods*).

The pedigree effects of a sweep may be likened to those of a large family, with the family now defined in genetic terms and where the event


A
0.1000
0.0500
0.0100
0.0050
0.0010
0.0005
0.0001
Proportion of polymorphic sites
0
200
400
600
800
1000
Sample count of mutant base
B
0.1000
0.0500
0.0100
0.0050
0.0010
0.0005
0.0001
Proportion of polymorphic sites
0
200
400
600
800
1000
Sample count of mutant base

unfolds over a larger number near here of generations. Another conceptually similar phenomenon is cultural inheritance of fertility, or correlation in offspring numbers, across generations, evidence for which has been inferred from the shapes of human mitochondrial gene genealogies (Blum et al., 2006).

Fig. 5.7 shows probabilities of pairwise coalescence, or proportion of loci expected to coalesce, in each of the past 56 generations assuming a selection coefficient of $s = 10$ (Fig. 5.7*A*) or $s = 1$ (Fig. 5.7*B*). When selection is extremely strong, such that individuals homozygous for the advantageous mutant allele have an average of 11 offspring for every one offspring of a wild-type homozygote (Fig. 5.7*A*), there is a sharp peak in the distribution of coalescence times around the time of the sweep. However, the overall effect on the proportion of loci expected to coalesce during the event is only about four times greater than for our "Genghis Khan" whose children comprise 8% of the population (Fig. 5.4*A*), and analogously we may infer that even this exceedingly strong selective sweep should have little impact on patterns of genetic variation.

Not surprisingly, the effects of lesser sweeps are very subtle. Fig. 5.7*B* shows the effect of a sweep with $s = 1$ on probabilities of pairwise coalescence, plotted over the same number of generations as Fig. 5.7*A* but with a notably different scale on the vertical axis. In this case, where homozygotes for the advantageous mutant allele have an average of two offspring for every one offspring of a wild-type homozygote, there is just a small bump in the proportion of loci expected to coalesce during the sweep, here centered around generation 26.

In contrast to the large-family simulations that included population growth, and therefore showed little coalescence in the first ~20 generations, both panels in Fig. 5.7 illustrate the effect of recent pedigree structure on probabilities of coalescence. In the most recent ~$\log_2 (N)$ generations,

FIGURE 5.6 Unfolded site-frequency distributions when the children of the large family comprise 8% (*A*) versus 68% (*B*) of the population. In both panels, the lines in red display results for the assumed background demography with growth but no large family and are identical in both panels, and the lines in blue show results when there is a large family. The lines in blue in *B* are based on 100,000 replicate loci; the others are based on 10,000 loci.

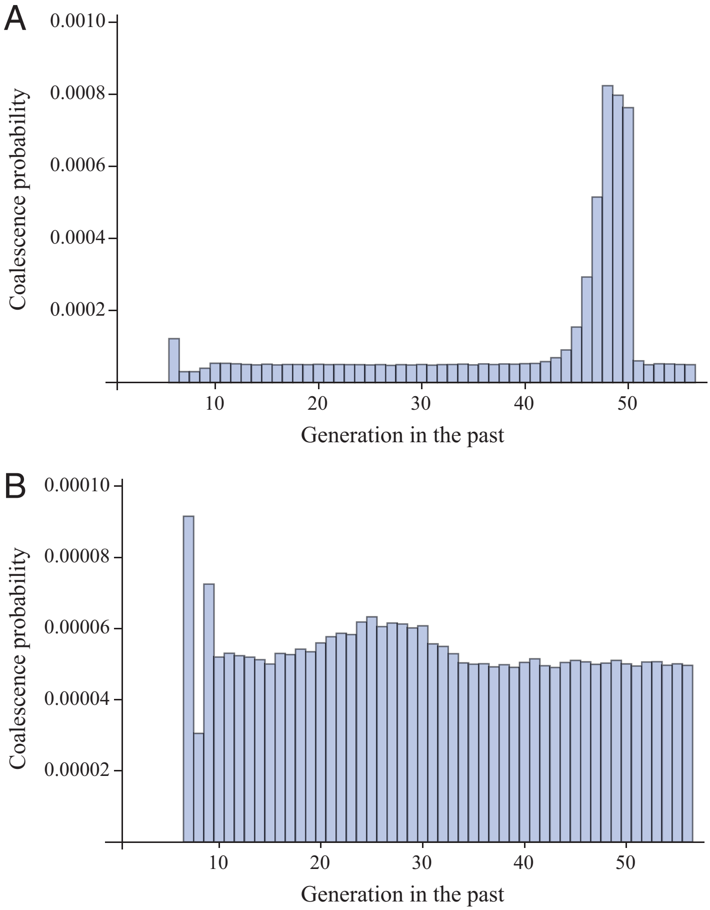


FIGURE 5.7 Distributions of pairwise coalescence times conditional on the population pedigree for the case of a selective sweep, with either $s = 10$ (*A*) or $s = 1$ (*B*). Each panel is based on a single population pedigree and single pair of sampled individuals. In both panels, coalescence probabilities have been computed numerically given each pedigree and pair of sampled individuals.

here about 13 generations, probabilities of coalescence depend strongly on the ancestries of the two sampled individuals. In the case of Fig. 5.7*A*, these ancestries did not overlap until generation 6 in the past and in the case of Fig. 5.7*B* they did not overlap until generation 7 in the past. Tracing farther back, in both cases, the probability then equilibrates and stays near $1/(2N)$, which here is equal to 0.00005 because $N = 10{,}000$.

Fig. 5.8 provides a more detailed view of the pedigree effects of strong selective sweeps. Ten replicate populations, each with a sweep beginning in generation 50 in the past, were simulated. The probabilities of both coalescence for a pair of lineages and the frequency of the advantageous allele were computed for every generation in the pedigree. These two quantities are shown in Fig. 5.8 with thicker and thinner lines, respectively, and using different colors for each of the 10 replicates. Fig. 5.8*A* shows that sweeps with $s = 10$ occur very quickly, in about 15 generations, whereas the sweeps in Fig. 5.8*B* for $s = 1$ take longer, about 50 generations. Coalescence probabilities for the sweeps in both panels display the relatively great variation over time and among pedigrees in the recent ~$\log_2 (N) \approx 13$ generations as well as the characteristic settling near $1/(2N) = 0.00005$ in the more distant past.

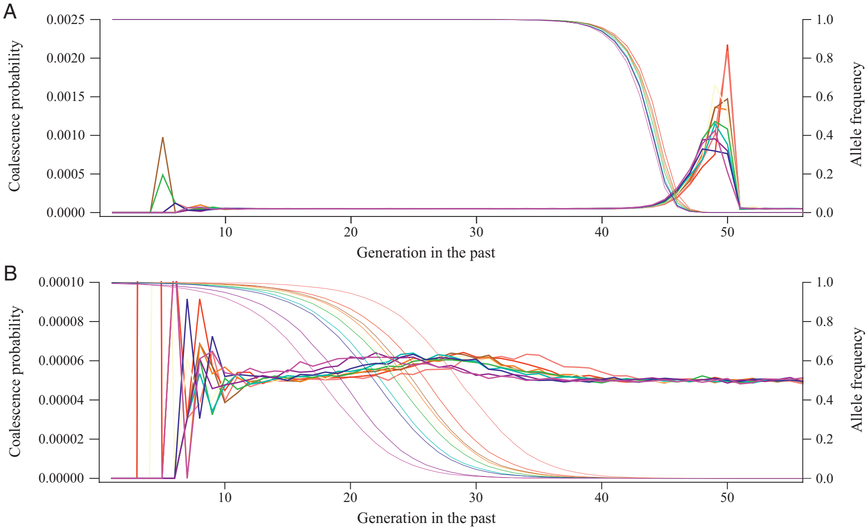


FIGURE 5.8 Distributions of pairwise coalescence times and trajectories of selective sweeps for 10 different replicate populations. As in Fig. 5.7, $s = 10$ in *(A)* and $s = 1$ in (*B*). The left vertical axes and thicker colored lines plot probabilities of coalescence and the right vertical axes and thinner colored lines plot frequencies of the favored allele as the sweeps progress, in each case beginning with a single copy in generation 50 in the past. In each panel, lines with the same color apply to the same replicate population.

A greater level of variation in the timing of the 10 sweeps is visible in Fig. 5.8*B*, with $s = 1$, than in Fig. 5.8*A*, with $s = 10$. Fig. 5.8*B* also shows that differences in the timing of the increase in coalescence probability track differences in the timing of sweeps (distinguished by color). Variation in the timing of a sweep is attributable to the time it takes the favored allele to escape the effects of genetic drift when it is in low copy number in the population. Especially in Fig. 5.8*A*, it can be seen that coalescence tends to happen earlier in the sweep, when the favored allele is in low frequency (Pennings and Hermisson, 2006). Finally, there is greater variation in the additional density of coalescence events among sweeps in Fig. 5.8*A* ($s = 10$) than in Fig. 5.8*B* ($s = 1$). We interpret this as a consequence of sweeps happening so quickly when $s = 10$ that coalescences depend strongly on the details of the initial increase of the favored allele.

CONCLUSIONS

We have explored two ways in which demographic events within populations may alter the structure of organismal genealogies, or population pedigrees, so as to produce unexpected patterns of variation across genomes. Our simulations of the effects of recent very large families and strong selective sweeps on variation among unlinked loci have primarily yielded negative results. Standard population-genetic predictions that average over pedigrees, such as $P(K_2 = k)$ in Eq. 5.2, seem quite robust even to fairly extreme versions of these events. However, we have also shown that frequencies of mutant alleles across the genome in very large samples of individuals provide more sensitive indicators of extreme demographic events, compared with simpler measures such as pairwise sequence differences. Following Keinan and Clark (2012), large samples have been of particular interest in human population genetics. For example, the recent update of the 1000 Genomes Project presented site frequencies in a sample of 2,504 people at more than 80 million SNPs (The 1000 Genome Project Consortium, 2015), so the potential is there to generate similar data for more geographically localized populations and for species other than humans to investigate the detailed effects of population pedigrees.

Finally, the genetic signatures of recent demographic events that we have uncovered apply marginally to single sites—they do not take linkage and recombination into account—and we note that the pedigree effects of such events might be relatively strong for multilocus measures such as the length distribution of blocks of identity by descent (Chapman and Thompson, 2003; Palamara et al., 2012).

MATERIALS AND METHODS

Simulations of Population Pedigrees and Coalescence

We simulated pedigrees according to the diploid, two-sex version of the Wright–Fisher model of random mating. That is, each individual in the next generation (forward in time) has a mother and a father chosen uniformly at random from the female and male adults of the current generation. Given a pedigree, neutral genetic loci are transmitted according to Mendel's laws. Importantly, multiple loci are independent conditional on the pedigree. For each simulated population pedigree, a sample of individuals is taken at random without replacement from the current population, which is generation 0 in the model. A single genetic lineage is followed backward in time from each sampled individual according to Mendel's law of independent segregation (i.e., going with 50% chance to the mother or the father in each generation). When two lineages trace back to the same individual, they coalesce with probability 1/2 and remain distinct in that individual with probability 1/2. For each pedigree and sample, we simulated large numbers of loci that were assumed also to follow Mendel's law of independent assortment. The programs used in this research may be downloaded from wakeleylab.oeb.harvard.edu/resources.

Pedigree Simulations Coalescent with a Large Family

We set the generation in which there was a large family to be generation 28 in the past using the fact that Genghis Khan lived about 800 years ago and a current estimate of 29 years as the average length of one human generation (Fenner, 2005). We assume that the children of our "Genghis Khan" comprised either 8% or 68% of the population, and that for the next 27 generations the population grew at rate 0.3 per generation (Keinan and Clark, 2012), which is similar to estimates of growth for descent clusters in Balaresque et al. (2015). The results we present do not depend strongly on this growth because, either way, generation 28 in the past is very recent compared with average coalescence time. We assume an ancestral population size of $N = 10{,}000$, and in every generation there are equal numbers of males and females in the population. Pedigree simulations were as above, except in generation 27 in the past. Depending on the case, in this generation, 27, either 8% or 68% of individuals have our hypothetical "Genghis Khan" as their father. The mothers of these individuals are chosen uniformly at random as in every other generation. When considering sample frequencies of mutations, or site frequencies, we sampled 1,000 individuals. With multiple lineages, multiple coalescent events can occur in single generations, either in different individuals or within single individuals. These multiple mergers are especially important in generation 28 in the past, when large numbers of lineages may trace back to "Genghis Khan." If k lineages trace back to a single individual, each of them has chance 1/2 of descending from each of the two alleles in that individual. Therefore, a binomially distributed number of lineages, with parameters k and 1/2, will trace back to one allele and the remainder will trace back to the other allele in that parent, creating two clusters of coalescence.

In simulating genetic data, we assumed that all mutations are selectively neutral and that each mutation produces a unique polymorphic site (Watterson, 1975). For each gene genealogy we placed a Poisson number of mutations randomly on the branches in the standard way to create pseudodata (Hudson, 2002), with the modification that our gene genealogies are not necessarily simple bifurcating trees. For a gene genealogy with total length t generations, the number of mutations would be Poisson($\theta t/(4N)$), where N is the ancestral population size, which we set to 10,000. We assumed the mutant state could be distinguished from the ancestral state at each polymorphic site when compiling the site-frequency distribution and simply counted the number of copies of the mutant in the sample of size 1,000. To verify that the site-frequency distribution shown in Fig. 5.6*B,* with a deficit of mutant counts around 150 and an increase above about 850, we performed simulations in which a sample of size 1,000 was subject to two rounds of binomial sampling. First, the number of lineages that trace back to "Genghis Khan" was given by a random draw from a binomial distribution with parameters 1,000 lineages and 0.68 × 0.5 for the probability of being among the children and tracing back to the father. All of these (k) lineages then coalesce into two groups, of sizes k_1 and $k_{2,}$ one for each of the two chromosomes in the father. We modeled this with a second random draw as described above, that is $k_1 \sim$ binomial(k, 1/2) and $k_2 = k - k_{1.}$ We then generated a standard coalescent tree (Hudson, 2002) with $1{,}000 - k + 2$ tips, where two tips had k_1 and k_2 descendants in the sample instead of the usual 1 descendant. These simulations assumed that the large-family event occurred instantaneously at time 0 and did not account for population growth, but the results were extremely similar to those in Fig. 5.6*B*.

Pedigree-Coalescent Simulations of Strong Selective Sweeps

We assumed that a selectively favored allele *A* was introduced as a mutant in a single copy in the population in generation 50 in the past, into a background of wild-type alleles *a*. The relative fitnesses of the three diploid genotypes were *AA*: $1 + s$, *Aa*: $1 + s/2$, and *aa*: 1. We conditioned on allele *A* not going extinct, by restarting the simulation if *A* was lost. Pedigrees were simulated from generation 50 onward by sampling parents nonuniformly, in proportion to their relative fitnesses. The population size was held constant at $N = 10{,}000$. Here, starting with a single pair of sampled individuals, we computed probabilities of coalescence in each generation in the past numerically (e.g., see equation 3 in Wakeley et al., 2012) rather than by simulating large numbers of independent loci. In addition, for each simulated pedigree with a sweep, we recorded the numbers of copies of allele A present in each generation.

ACKNOWLEDGMENTS

We thank Noah Rosenberg for helpful comments and discussion.

6

The Probability of Monophyly of a Sample of Gene Lineages on a Species Tree

ROHAN S. MEHTA,*‡ DAVID BRYANT,† AND
NOAH A. ROSENBERG*

Monophyletic groups—groups that consist of all of the descendants of a most recent common ancestor—arise naturally as a consequence of descent processes that result in meaningful distinctions between organisms. Aspects of monophyly are therefore central to fields that examine and use genealogical descent. In particular, studies in conservation genetics, phylogeography, population genetics, species delimitation, and systematics can all make use of mathematical predictions under evolutionary models about features of monophyly. One important calculation, the probability that a set of gene lineages is monophyletic under a two-species neutral coalescent model, has been used in many studies. Here, we extend this calculation for a species tree model that contains arbitrarily many species. We study the effects of species tree topology and branch lengths on the monophyly probability. These analyses reveal new behavior, including the maintenance of nontrivial monophyly probabilities for gene lineage samples that span multiple species and even for lineages that do not derive from a monophyletic species group. We illustrate the mathematical results using an example application to data from maize and teosinte.

---

*Department of Biology, Stanford University, Stanford, CA 94305; and †Department of Mathematics and Statistics, University of Otago, Dunedin 9054, New Zealand. ‡To whom correspondence should be addressed. Email: rsmehta@stanford.edu.

Mathematical computations under coalescent models have been central in developing a modern view of the descent of gene lineages along the branches of species phylogenies. Since early in the development of coalescent theory and phylogeography, coalescent formulas and related simulations have contributed to a probabilistic understanding of the shapes of multispecies gene trees (Tajima, 1983; Takahata and Nei, 1985; Neigel and Avise, 1986), enabling novel predictions about gene tree shapes under evolutionary hypotheses (Rosenberg, 2003; Degnan and Salter, 2005), new ways of testing hypotheses about gene tree discordances (Wu, 1991; Yu et al., 2012), and new algorithms for problems of species tree inference (Liu et al., 2009a; Wu, 2012) and species delimitation (Knowles and Carstens, 2007; Yang and Rannala, 2010). A “multispecies coalescent” model, in which coalescent processes on separate species tree branches merge back in time as species reach a common ancestor (Degnan and Rosenberg, 2009), has become a key tool for theoretical predictions, simulation design, and evaluation of inference methods, and as a null model for data analysis.

A fundamental concept in genealogical studies is that of monophyly. In a genealogy, a group that is monophyletic consists of all of the descendants of its most recent common ancestor (MRCA): every lineage in the group—and no lineage outside it—descends from this ancestor. Backward in time, a monophyletic group has all of its lineages coalesce with each other before any coalesces with a lineage from outside the group.

The phylogenetic and phylogeographic importance of monophyly traces to the fact that monophyly enables a natural definition of a genealogical unit. Such a unit can describe a distinctive set of organisms that differs from other groups of organisms in ways that are evolutionarily meaningful. Species can be delimited by characters present in every member of a species and absent outside the species, and that therefore can reflect monophyly (Sites and Marshall, 2004; De Queiroz, 2007). In conservation biology, monophyly can be used as a prioritization criterion because groups with many monophyletic loci are likely to possess unique evolutionary features (Moritz, 1994). Reciprocal monophyly, in which a set of lineages is divided into two groups that are simultaneously monophyletic, is often used in a genealogical approach to species divergence (Baum and Shaw, 1995; Hudson and Coyne, 2002). The proportion of loci that are reciprocally monophyletic is informative about the time since species divergence and can assist in representing the level of differentiation between groups (Edwards and Beerli, 2000; Rosenberg, 2003).

Many empirical investigations of genealogical phenomena have made use of conceptual and statistical properties of monophyly (Funk and Omland, 2003). Comparisons of observed monophyly levels to model predictions have been used to provide information about species divergence times (Hare and Weinberg, 2005; Syring et al., 2007). Model-based monophyly computations have been used alongside DNA sequence differences between and within proposed clades to argue for the existence of the clades (Birky et al., 2005), and tests involving reciprocal monophyly have been used to explain differing phylogeographic patterns across species (Carstens and Richards, 2007). Comparisons of observed levels of monophyly with the level expected by chance alone (Rosenberg, 2007) have assisted in establishing the distinctiveness of taxonomic groups (Neilson and Stepien, 2009; Kubatko et al., 2011). Loci that conflict with expected monophyly levels have provided signatures of genic roles in species divergences (Wang et al., 1999; Ting et al., 2000; Dopman et al., 2005).

For lineages from two species under a model of population divergence, Rosenberg (2003) computed probabilities of four different genealogical shapes: reciprocal monophyly of both species, monophyly of only one of the species, monophyly of only the other species, and monophyly of neither species. The computation permitted arbitrary species divergence times and sample sizes—generalizing earlier small-sample computations (Tajima, 1983; Takahata and Nei, 1985; Neigel and Avise, 1986; Takahata and Slatkin, 1990; Wakeley, 2000)—and illustrated the transition from the species divergence, when monophyly is unlikely for both species, to long after divergence, when reciprocal monophyly becomes extremely likely. Between these extremes, the species can pass through a period during which monophyly of one species but not the other is the most probable state.

Although this two-species computation has contributed to various insights about empirical monophyly patterns (Birky et al., 2005; Hickerson et al., 2006a; Carstens and Knowles, 2007; Carstens and Richards, 2007; Syring et al., 2007; Bergsten et al., 2012), many scenarios deal with more than two species. Because multispecies monophyly probability computations have been unavailable—except in limited cases with up to four species (Rosenberg, 2002, 2003; Degnan, 2010; Zhu et al., 2011; Eldon and Degnan, 2012)—multispecies studies have been forced to rely on two-species models, restricting attention to species pairs (Baker et al., 2009; Neilson and Stepien, 2009; Bergsten et al., 2012) or pooling disparate lineages and disregarding their taxonomic distinctiveness (Ting et al., 2000; Carstens and Richards, 2007).

Here, we derive an extension to the two-species monophyly probability computation, examining arbitrarily many species related by an evolutionary tree. Furthermore, we eliminate the past restriction (Rosenberg, 2003) that the lineages whose monophyly is examined all derive from the same population. This generalization is analogous to the assumption that in computing the probability of a binary evolutionary character (RoyChoudhury et al., 2008; Bryant et al., 2012; RoyChoudhury and Thompson, 2012), one or both character states can appear in multiple species. Our approach uses a pruning algorithm, generalizing the two-species formula in a conceptually similar manner to other recursive coalescent computations on arbitrary trees (Efromovich and Kubatko, 2008; RoyChoudhury et al., 2008; Bryant et al., 2012; RoyChoudhury and Thompson, 2012; Stadler and Degnan, 2012; Wu, 2012).

Like the work of Degnan and Salter (2005), which considered probability distributions for gene tree topologies under the multispecies coalescent model, our work generalizes a coalescent computation known only for small trees (Rosenberg, 2002, 2003) to arbitrary species trees. We study the dependence of the monophyly probability on the model parameters, providing an understanding of factors that contribute to monophyly in species trees of arbitrary size. Finally, we explore the utility of monophyly probabilities in an application to genomewide data from maize and teosinte.

RESULTS

Model and Notation

Overview

Consider a rooted binary species tree $\mathcal{T}$ with ℓ leaves and specified topology and branch lengths. For each of the ℓ species represented by leaves of $\mathcal{T}$, a number of sampled lineages is specified. Given a specified partition of the lineages into two subsets, we consider a condition describing whether one, the other, both, or neither of the two subsets of lineages is monophyletic. Our goal is to provide a recursive computation of the probability that the condition is obtained under the multispecies coalescent model. Notation appears in *Table S1*.[1]

Lineage Classes

The initial sampled lineages are partitioned into class S (subset) for lineages within a chosen subset, and class C (complement) for all lineages not included in S. Coalescence between an S lineage and a C lineage produces an M (mixed) lineage. Any coalescence involving an M lineage also produces an M lineage. Coalescences between two S or two C lineages produce S and C lineages, respectively (Table 6.1).

---

[1] Supporting information for this chapter, which includes *Table S1*, *Figure S1*, and *Datasets S1* and *S2*, is available online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1601074113/-/DCSupplemental.

TABLE 6.1 Lineage Classes Produced by Coalescence Events

		Class 2		
		S	*C*	*M*
Class 1	*S*	S ⅄ S S	M ⅄ S C	M ⅄ S M
	C	M ⅄ C S	C ⅄ C C	M ⅄ C M
	M	M ⅄ M S	M ⅄ M C	M ⅄ M M

NOTE: Intraclass coalescences between pairs of lineages preserve the class; interclass coalescences result in *M* lineages.

Letting the number of *S* and *C* lineages present initially in the *i*th leaf be S_i and C_i, respectively, the model parameters are S_i and C_i for $1 \le i \le \ell$, and the species tree $\mathcal{T}$. For convenience, we aggregate the S_i and C_i with $\mathcal{T}$ into a parameter collection $\mathcal{T}_{SC}$ that we call the initialized species tree.

Monophyly Events

A monophyly event E_i is an assignment of labels to lineage classes *S* and *C*. We can choose to label a class "monophyletic" or "not monophyletic," or assign no label at all, so that nine monophyly events are possible, six of which are relevant for our purposes (Table 6.2). All lineages in a monophyletic class must coalesce within the class to a single lineage before any coalesces outside the class. If multiple classes are labeled monophyletic, then each class must be separately monophyletic.

Species-Merging Events

We orient the species tree vertically, "up" toward the root and "down" toward the leaves. From a coalescent backward-in-time perspective, at every internal node of the species tree—representing a species-merging event—lineages enter from two branches directly below the node. We label one of these branches "left" and the other "right," based on an arbitrarily labeled diagram of species tree $\mathcal{T}$. These labels are used only for bookkeeping; the labeling does not affect subsequent calculations. Lineages entering from the left and right branches are called "left inputs" and "right inputs," respectively. Each node *x* of $\mathcal{T}$ is associated with exactly

TABLE 6.2 Possible Monophyly Events for Two Disjoint Lineage Classes, S and C

Monophyletic Groups	Description	Notation
S	Monophyly of S	E_S
C	Monophyly of C	E_c
Only S	Paraphyly of C	$E_{SC'}$
Only C	Paraphyly of S	$E_{S'C}$
Both S and C	Reciprocal monophyly	E_{SC}
Neither S nor C	Polyphyly	$E_{S'C'}$

one branch, leading from node x to its immediate predecessor on $\mathcal{T}$. We refer to this branch with the shared label x.

For an internal branch x in $\mathcal{T}$, the number of class-S left inputs is s_x^L (c_x^L for class C, m_x^L for class M); the number of class-S right inputs is s_x^R (c_x^R for class C, m_x^R for class M). The total number of class-S inputs of x is $s_x^I = s_x^L + s_x^R$ ($c_x^I = c_x^L + c_x^R$ for class C, $m_x^I = m_x^L + m_x^R$ for class M). The number of lineages that exit branch x, entering a branch farther up the species tree, is the set of outputs of branch $x : s_x^O, c_x^O$, or m_x^O.

We combine the input and output values into two three-entry vectors: the "input states" $\mathbf{n}_x^I = (s_x^I, c_x^I, m_x^I)$ and the "output states" $\mathbf{n}_x^O = (s_x^O, c_x^O, m_x^O)$. Note that $\mathbf{n}_x^I = \mathbf{n}_x^L + \mathbf{n}_x^R$. We refer to the nodes directly below node x corresponding to its left and right incoming branches by x_L and x_R, respectively, and to nodes farther down the tree by sequences of Ls and Rs, which, read from left to right, give the steps needed to reach them from x. For example, x_{RL} follows down from x to the right (x_R), then from x_R to the left (x_{RL}).

The time interval associated with node x is T_x, the length of branch x. Branch lengths are measured in coalescent time units of N generations, where N represents the haploid population size along the branch and is assumed to be constant. Thus, larger population sizes correspond to shorter lengths of time in coalescent units. Coalescences between inputs during time T_x yield the outputs of x. The root branch of $\mathcal{T}$ has infinite length.

The outputs of any nonroot branch are exactly the left or the right inputs of another branch farther up the tree; the outputs of the root are the outputs of the species tree. The root has only one output lineage: $\mathbf{n}_{\text{root}}^O = (0,0,1)$. Inputs of a node x are the outputs of x_L and x_R, so that $\mathbf{n}_x^L = \mathbf{n}_{x_L}^O$ and $\mathbf{n}_x^R = \mathbf{n}_{x_R}^O$. For convenience, when node x corresponds to leaf i, we let $\mathrm{s}_x^I = \mathrm{s}_x^L = S_i$ and $\mathrm{c}_x^I = \mathrm{c}_x^L = \mathrm{C}_i$ (Fig. 6.1).

We define $\mathcal{T}_{SC}^x$ to be the initialized species subtree with root x and E_i^x to be the monophyly event E_i for the subtree with root x, ignoring the rest of the species tree.

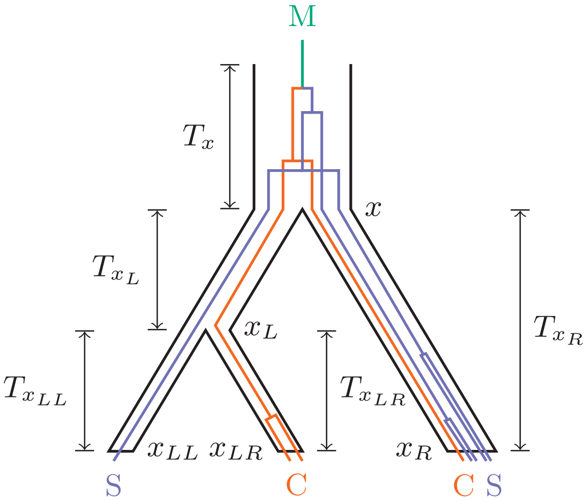


FIGURE 6.1 Notation for computing monophyly probabilities above a species tree node x. Nodes x_{LL}, x_{LR}, and x_R are leaves. S lineages appear in blue, C lineages in orange, and M lineages in green. The figure illustrates reciprocal monophyly. Sequentially listing the numbers of S, C, and M lineages as a vector, the outputs of branch x are $\mathbf{n}_x^O = (0,0,1)$. Inputs are $\mathbf{n}_x^L = (1,1,0)$ and $\mathbf{n}_x^R = (2,1,0)$. Farther down the tree, branch x_L has inputs $\mathbf{n}_{x_L}^L = (1,0,0)$ and $\mathbf{n}_{x_L}^R = (0,1,0)$. Adopting the convention that leaf inputs enter from the left, branch x_R has inputs $\mathbf{n}_{x_R}^L = (4,1,0)$ and $\mathbf{n}_{x_R}^R = (0,0,0)$. Descending one more level—which is only possible for x_L—the inputs for branch x_{LL} are $\mathbf{n}_{x_{LL}}^L = (1,0,0)$ and $\mathbf{n}_{x_{LL}}^R = (0,0,0)$, and for branch x_{LR}, they are $\mathbf{n}_{x_{LR}}^L = (0,2,0)$ and $\mathbf{n}_{x_{LR}}^R = (0,0,0)$. Branch widths represent constant population sizes but do not indicate relative magnitudes of these sizes.

Coalescence Sequences

A coalescence sequence is a sequence of coalescences that reduces a set of lineages to another set of lineages. As an example, consider four lineages—labeled A, B, C, and D—that coalesce to a single lineage. One sequence has A and C coalesce first, followed by B and D, then the lineages resulting from the AC and BD coalescences. This sequence could be described as (A, C), (B, D), (AC, BD). If the first two coalescences happened in opposite order, the sequence would be (B, D), (A, C), (AC, BD).

Combinatorial Functions

The probability $g_{n,j}(T)$ that n lineages coalesce to j lineages in time T is given by equation 6.1 of Tavaré (1984). It is nonzero only when $n \geq j \geq 1$ and $T \geq 0$, except that we set $g_{0,0}(T) = 1$.

Following equation 4 of Rosenberg (2003), the number of coalescence sequences that reduce n lineages to k lineages is $I_{n,k} = [n!(n-1)!]/[2^{n-k}k!(k-1)!]$. This function is nonzero only when $n \geq k \geq 1$, with the convention $I_{0,0} = 1$.

Finally, the binomial coefficient,

$$W_2(r_1, r_2) = \binom{r_1 + r_2}{r_1},$$

by equation 5 from Rosenberg (2003), gives the number of ways that separate coalescence sequences consisting of r_1 and r_2 coalescences can be ordered in a larger sequence containing them both as subsequences. $W_2(r_1,r_2)$ is defined when $r_1,r_2 \geq 0$.

The Central Recursion

Overview

We develop a recursion for the probability of a particular output state $\mathbf{n}_x^O$ and monophyly event E_i^x for a branch x given the initialized species subtree $\mathcal{T}_{SC}^x$. We use the law of total probability to write the desired probability as a sum over all possible input states $\mathbf{n}_x^I$ of the probability of the input state multiplied by the conditional probability of the output given the input. Keeping in mind that {inputs of x} = {outputs of x_L}∪{outputs of x_R}, we then use the independence of the outputs for branches x_L and x_R to decompose the probability of the input state of x into a product of the probabilities of the output states of x_L and x_R. Schematically,

$$\begin{aligned}\mathbb{P}\left(\text{outputs of } x, E_i^X \middle| \mathcal{T}_{SC}^x\right) = \sum_{\substack{\text{possible} \\ \text{inputs of } x}} & \mathbb{P}\left(\text{outputs of } x_L, E_i^{X_L} \middle| \mathcal{T}_{SC}^{x_L}\right) \\ & \times \mathbb{P}\left(\text{outputs of } x_R, E_i^{X_R} \middle| \mathcal{T}_{SC}^{x_R}\right) \\ & \times \mathbb{P}\left(\text{outputs of } x, E_i^X \middle| \text{inputs of } x, \mathcal{T}_{SC}^x\right). \qquad \textbf{[6.1]}\end{aligned}$$

The third term on the right-hand side of Eq. 6.1, which we represent by F, is the probability that the inputs coalesce to the specified outputs during time T_x in accord with the monophyly event. We write the random variable for the output state of branch x as $\mathbf{Z}_x$, labeling the particular values attained by the random variable by $\mathbf{n}_x^O$. By formalizing Eq. 6.1, we can write the central recursion of our analysis:

$$\mathbb{P}\left(\mathbf{Z}_x = \mathbf{n}_x^O, E_i^x \middle| \mathcal{T}^x_{SC}\right) = \sum_{\mathbf{n}_x^L=0}^{\left(S^{\text{subt}}_{x_L}, C^{\text{subt}}_{x_L}, 1\right)} \sum_{\mathbf{n}_x^R=0}^{\left(S^{\text{subt}}_{x_R}, C^{\text{subt}}_{x_R}, 1\right)} \mathbb{P}\left(\mathbf{Z}_{x_L} = \mathbf{n}^O_{x_L}, E_i^{x_L} \middle| \mathcal{T}^{x_L}_{SC}\right)$$
$$\times\ \mathbb{P}\left(\mathbf{Z}_{x_R} = \mathbf{n}^O_{x_R}, E_i^{x_R} \middle| \mathcal{T}^{x_R}_{SC}\right) F\left(\mathbf{n}_x^O, E_i^x \middle| \mathbf{n}_x^I, \mathcal{T}^x_{SC}\right). \qquad \textbf{[6.2]}$$

In this equation, we denote the total number of inputs of class *S* across all of the leaves subtended by x_L or x_R by $S^{\text{subt}}_{x_L}$ or $S^{\text{subt}}_{x_R}$ ($C^{\text{subt}}_{x_L}$ or $C^{\text{subt}}_{x_R}$ for class *C*). Each of the two summations is a nested triple sum, proceeding componentwise over the three entries in the vectors $\mathbf{n}_x^L$ and $\mathbf{n}_x^R$—for example, for $\mathbf{n}_x^L$, we sum from 0 to $S^{\text{subt}}_{x_L}$, from 0 to $C^{\text{subt}}_{x_L}$, and from 0 to 1. We now explain the basis for this recursion.

Bounds of Summation

The sums in Eq. 6.2 traverse all possible inputs of branch *x*. We use summation bounds that only require information contained in the initialized species subtree $\mathcal{T}^x_{SC}$. Numbers of inputs are nonnegative, and for each lineage class, some branches have the possibility of having no inputs in the class. Thus, all lower bounds are 0.

For the upper bounds, because coalescence does not create new *S* and *C* lineages (Table 6.1), the numbers of *S* and *C* lineages never exceed the numbers of *S* and *C* leaves in the gene tree, respectively. Thus, for branch *x*, an upper bound for the possible number of inputs of class *S* or *C* from one side (*L* or *R*) is $S^{\text{subt}}_{x_L}$ or $S^{\text{subt}}_{x_R}$ for class *S* and $C^{\text{subt}}_{x_L}$ or $C^{\text{subt}}_{x_R}$ for class *C*.

We use Eq. 6.2 to calculate probabilities only for E_S, E_C, and E_{SC} (Table 6.2), using them to obtain probabilities for the remaining events. These three events require complete intraclass coalescence separately in the appropriate classes before interclass coalescences are possible. As a result, they permit exactly one coalescence between an *S* lineage and a *C* lineage. Because the leaves possess no *M* lineages and because only the unique coalescence between an *S* and a *C* lineage creates an *M* lineage (Table 6.1), the number of *M* lineages never exceeds 1.

Probability of the Outputs of a Node Given the Inputs

Separating the function *F* from Eq. 6.2 into a term for the probability that the correct number of outputs is produced from the inputs and a combinatorial term K_i for the probability that the coalescence sequence generating those outputs occurs in accord with the monophyly event E_i, *F* takes the form

$$F\left(\mathbf{n}_x^O, E_i^x \middle| \mathbf{n}_x^I, \mathcal{T}^x_{SC}\right) = g_{|\mathbf{n}_x^I|, |\mathbf{n}_x^O|}\left(T_x\right) K_i\left(\mathbf{n}_x^O, E_i^x \middle| \mathbf{n}_x^I, \mathcal{T}^x_{SC}\right), \qquad \textbf{[6.3]}$$

where $|\mathbf{n}_x^I| = s_x^I + c_x^I + m_x^I$ and $|\mathbf{n}_x^O| = s_x^O + c_x^O + m_x^O$. For the case of $i = S$, in which monophyly of S is of interest, we have:

$$K_S\left(\mathbf{n}_x^O, E_S^x \middle| \mathbf{n}_x^I, \mathcal{T}_{SC}^x\right) = \begin{cases} 1 & \text{Case 1a: } \mathbf{n}_x^I = (0,0,1), \mathbf{n}_x^O = (0,0,1) \\ & \text{Case 1b: } \mathbf{n}_x^I = (0,c_1,0), \mathbf{n}_x^O = (0,c_2,0) \\ & \text{Case 1c: } \mathbf{n}_x^I = (0,c_1,1), \mathbf{n}_x^O = (0,c_2,1) \\ & \text{Case 1d: } \mathbf{n}_x^I = (0,c_1,1), \mathbf{n}_x^O = (0,0,1) \\ & \text{Case 1e: } \mathbf{n}_x^I = (s_1,0,0), \mathbf{n}_x^O = (s_2,0,0) \\ \dfrac{\sum_{k=c_2+1}^{c_1} I_{S_x^{\mathcal{T}},1} I_{c_1,k} W_2\left(S_x^{\mathcal{T}} - 1, c_1 - k\right) k I_{k,c_2+1}}{I_{S_x^{\mathcal{T}}+c_1,c_2+1}} & \text{Case 2: } \mathbf{n}_x^I = (S_x^{\mathcal{T}}, c_1, 0), \mathbf{n}_x^O = (0,c_2,1) \\ \dfrac{I_{s_1,s_2} I_{c_1,c_2} W_2\left(s_1 - s_2, c_1 - c_2\right)}{I_{s_1+c_1,s_2+c_2}} & \text{Case 3: } \mathbf{n}_x^I = (s_1,c_1,0), \mathbf{n}_x^O = (s_2,c_2,0) \\ 0 & \text{otherwise} \end{cases} \qquad \textbf{[6.4]}$$

Here, $s_x^{\mathcal{T}}$ records the total number of class-S lineages in the species tree $\mathcal{T}$ at the species merging event corresponding to node x. For cases 1 and 3, $0 < c_2 \leq c_1$ and $0 < s_2 \leq s_1$. For case 2, $0 \leq c_2 < c_1$, $0 < c_1$, and $0 < s_x^{\mathcal{T}}$. Note that it is not strictly necessary for $s_x^I = s_x^{\mathcal{T}}$ in case 2 (violation of E_S would be accommodated elsewhere in the calculation, on another species tree branch), but we retain this condition for clarity.

Function F (Eq. 6.3) describes the probability of an output state and monophyly event given an input state and the initialized species tree. Its g term records the probability that the correct number of coalescences occur during the time T_x, defining a space of coalescence sequences from the input state to any output state with the same number of lineages as the desired output. K_i (Eq. 6.4) records the fraction of those sequences that produce the correct output and preserve the monophyly event E_i (in this case, E_S).

The cases in Eq. 6.4 represent distinct scenarios for the types of input and output lineages present (Fig. 6.2*A–G*). In case 1 (Fig. 6.2*A–E*), no coalescence violates E_S, as all coalescences have types (S, S) (case 1e), (C, C) (cases 1b, 1c, 1d), or (C, M) (cases 1c, 1d). No coalescences occur in case 1a. The correct output state is guaranteed ($K_S = 1$), as each coalescence decrements the number of S (case 1e) or C lineages (cases 1b, 1c, 1d), and the only change from input to output is a reduction in S or C lineages.

In cases 2 and 3, both S and C lineages are present, and we enumerate the ways to obtain the desired output state from the input state in accord

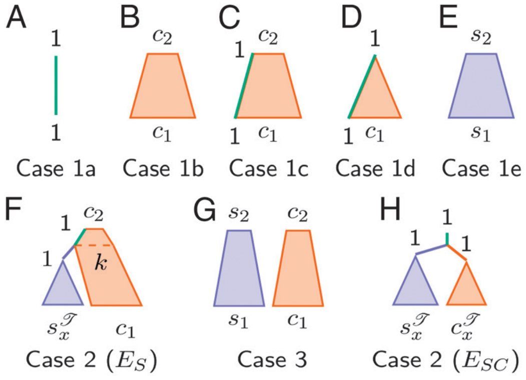


FIGURE 6.2 All cases required for computing combinatorial terms K_S and K_{SC} in monophyly probabilities. (*A*–*G*) Cases for monophyly of *S* (Eq. 6.4). (*H*) A case for reciprocal monophyly (Eq. 6.5). In each panel, lineages coalesce from bottom to top, with the width of a shape corresponding to the number of lineages present. A single lineage is represented by a line, and multiple freely coalescing lineages are represented by shaded polygons with horizontal cross-section proportional to the number of extant lineages. Lineages represented in the same shape or in touching shapes can coalesce with each other. Lineage colors follow Fig. 6.1.

with the monophyly event. To obtain K_S, we divide by the total number of coalescence sequences of correct length.

Case 2 describes the only possible way an *S* lineage and a *C* lineage can coalesce with each other under E_S (Fig. 6.2*F*). All extant *S* lineages at the time of node *x* ($s_x^I = s_x^{\mathcal{T}}$) must coalesce to a single lineage, and that lineage must coalesce with a *C* lineage when *k* class-*C* lineages remain from the $c_x^I = c_1$ extant *C* lineages present in both species at node x. This coalescence results in a single *M* lineage and $k - 1$ lineages of class *C*, which can coalesce in any order to a single class-*M* lineage and $c_x^O = c_2$ class-*C* lineages.

The number of ways that $s_x^{\mathcal{T}}$ lineages can coalesce to one lineage is $\mathrm{I}_{s_x^{\mathcal{T}},1}$. The number of ways that c_1 lineages can coalesce to *k* lineages is $I_{c_1,k}$. These separate sequences of $s_x^{\mathcal{T}} - 1$ and $c_1 - k$ coalescences can be ordered in $W_2(s_x^{\mathcal{T}} - 1, c_1 - k)$ ways. The number of ways that a single *S* lineage can coalesce with one of *k* lineages of class *C* is *k*. Finally, *k* lineages—one *M* lineage and $k - 1$ class-*C* lineages—can coalesce to $c_2 + 1$ lineages in I_{k,c_2+1} ways. The desired number of coalescence sequences of correct length that result in the correct output state without violating E_S is obtained by summing the product of these terms over possible values of k, which ranges from just enough C lineages ($c_2 + 1$) to allow the correct number of output lineages (c_2)—the resultant single S lineage coalesces with one C lineage and then no other coalescence occurs—to the total number c_1 of incoming C lineages, when all of the S lineages coalesce before any of the C lineages coalesce. The denominator of ratio K_S is the total number of ways of coalescing $s_x^{\mathcal{T}} + c_1$ input lineages to $c_2 + 1$ output lineages: $I_{s_x^{\mathcal{T}} + c_1, c_2 + 1}$. Note that setting $c_2 = 0$ in the ratio, reflecting a scenario with only one output lineage, of class M, reduces the formula to the two-species equation 11 from Rosenberg (2003) (*Supporting Information*).

Case 3 describes any situation with S and C lineages present and no interclass coalescence (Fig. 6.2*G*). At node x, the $s_x^I = s_1$ class-S lineages coalesce to $s_x^O = s_2$ class-S lineages, and the $c_x^I = c_1$ class-C lineages to $c_x^O = c_2$ class-C lineages. Group S has not yet coalesced with the other sampled lineages and does not do so within this species tree branch; its monophyly is not necessarily determined on the branch. The number of ways s_1 lineages can coalesce to s_2 lineages is I_{s_1,s_2}; c_1 lineages can coalesce to c_2 lineages in I_{c_1,c_2} ways. These sequences can be ordered in $W_2(s_1 - s_2, c_1 - c_2)$ ways. The numerator in the fraction of coalescence sequences of the correct length that result in the correct output state without violating E_S is the product of these three terms. The denominator is the total number of ways of coalescing $s_1 + c_1$ input lineages to $s_2 + c_2$ outputs: $I_{s_1 + c_1, s_2 + c_2}$.

Any pairing of an input state and an output state that does not belong in cases 1–3 of Eq. 6.4 must violate E_S. This violation yields an output probability of $K_S = 0$.

Reciprocal Monophyly

Monophyly events E_{SC} and E_S differ in that for E_{SC}, unlike for E_S, C and M lineages cannot coexist. Thus, cases 1c and 1d of Eq. 6.4 move to "otherwise" for K_{SC}, producing $K_{SC} = 0$ for the input states of those cases. Additionally, for E_{SC}, an interclass coalescence can occur only after all S lineages have coalesced to a single S lineage and all C lineages have coalesced to a single C lineage, whereas E_S required only that all S lineages coalesce. For E_S, interclass coalescences occur only in case 2 of Eq. 6.4; for E_{SC}, we modify this case by requiring first that before the interclass coalescence, the C lineages must be all C lineages in the tree at the time of node x (as we did for S lineages for case 2 of Eq. 6.4; $c_x^I = c_x^{\mathcal{T}}$). Second we require $k = 1$ and $c_2 = 0$, so all C lineages coalesce to a single lineage before the interclass coalescence. Setting $k = 1$, $c_2 = 0$, substituting $c_x^{\mathcal{T}}$ for

c_1 in case 2 of Eq. 6.4, and noting that $I_{1,1} = 1$, we obtain case 2 for K_{SC} (Fig. 6.2*H*), applicable when $\mathbf{n}_x^I = (s_x^{\mathcal{T}}, c_x^{\mathcal{T}}, 0)$ and $\mathbf{n}_x^O = (0,0,1)$:

$$K_{SC}\left(\mathbf{n}_x^O, E_{SC}^x \middle| \mathbf{n}_x^I, \mathcal{T}_{SC}^x\right) = \frac{I_{S_x^{\mathcal{T}},1} I_{C_x^{\mathcal{T}},1} W_2\left(S_x^{\mathcal{T}} - 1, C_x^{\mathcal{T}} - 1\right)}{I_{S_x^{\mathcal{T}} + C_x^{\mathcal{T}},1}}. \qquad \textbf{[6.5]}$$

For E_{SC}, the input condition for case 2 can be satisfied only at the root of $\mathcal{T}$. For all input states other than those of Eq. 6.5 or cases 1c and 1d of Eq. 6.4, $K_{SC} = K_S$.

Completing the Calculation

Having obtained a recursion that propagates monophyly probabilities through a species tree, we apply Eq. 6.2 at the root to complete the calculation of the probability of a monophyly event on $\mathcal{T}_{SC}$:

$$\mathbb{P}\left(E_i \middle| \mathcal{T}_{SC}\right) = \mathbb{P}\left(\mathbf{Z}_{\text{root}} = (0,0,1), E_i^{\text{root}} \middle| \mathcal{T}_{SC}^{\text{root}}\right). \qquad \textbf{[6.6]}$$

Specifying each possible monophyly event E_i^{root} in Eq. 6.6,

$$\mathbb{P}\left(E_S \middle| \mathcal{T}_{SC}\right) = \mathbb{P}\left(\mathbf{Z}_{\text{root}} = (0,0,1), E_S^{\text{root}} \middle| \mathcal{T}_{SC}^{\text{root}}\right) \qquad \textbf{[6.7]}$$

$$\mathbb{P}\left(E_C \middle| \mathcal{T}_{SC}\right) = \mathbb{P}\left(E_S \middle| \mathcal{T}_{CS}\right) \qquad \textbf{[6.8]}$$

$$\mathbb{P}\left(E_{SC} \middle| \mathcal{T}_{SC}\right) = \mathbb{P}\left(\mathbf{Z}_{\text{root}} = (0,0,1), E_{SC}^{\text{root}} \middle| \mathcal{T}_{SC}^{\text{root}}\right) \qquad \textbf{[6.9]}$$

$$\mathbb{P}\left(E_{SC'} \middle| \mathcal{T}_{SC}\right) = \mathbb{P}\left(E_S \middle| \mathcal{T}_{SC}\right) - \mathbb{P}\left(E_{SC} \middle| \mathcal{T}_{SC}\right) \qquad \textbf{[6.10]}$$

$$\mathbb{P}\left(E_{S'C} \middle| \mathcal{T}_{SC}\right) = \mathbb{P}\left(E_C \middle| \mathcal{T}_{SC}\right) - \mathbb{P}\left(E_{SC} \middle| \mathcal{T}_{SC}\right) \qquad \textbf{[6.11]}$$

$$\mathbb{P}\left(E_{S'C'} \middle| \mathcal{T}_{SC}\right) = 1 - \mathbb{P}\left(E_{SC'} \middle| \mathcal{T}_{SC}\right) - \mathbb{P}\left(E_{S'C} \middle| \mathcal{T}\right) - \mathbb{P}\left(E_{SC} \middle| \mathcal{T}_{SC}\right), \qquad \textbf{[6.12]}$$

where $\mathcal{T}_{CS}$ is $\mathcal{T}_{SC}$ with the labels S and C switched. These recursive computations reduce to the known values for the two-species case (*Supporting Information*).

Effect of Species Tree Height *T*

To illustrate the features of monophyly probabilities, we now examine the effects on the probabilities of model parameters. First, we vary the tree height T and preserve relative branch length proportions, studying the limiting cases of $T = 0$ and $T \to \infty$.

T = *0*

At $T = 0$, nonroot species tree branches have length 0, so the species tree is a single infinitely long branch—the root—with initial sample sizes equal to the sums of the values at the leaves. Formally, because $g_{i,j}(0) = 1$ if $i = j$, every nonroot branch outputs exactly its inputs. All $s = \sum_{i=1}^{\ell} S_i$ class-S lineages and all $c = \sum_{i=1}^{1} C_\mathrm{i}$ class-C lineages enter the root. Using Eq. 6.7, and noting that $g_{i,1}(\infty) = 1$, we find that $\mathbb{P}(E_S \mid \mathcal{T}_{SC})$ is a simple function of the total numbers of S and C lineages:

$$f(s,c) = \frac{\sum_{k=1}^{c} I_{s,1} I_{c,k} W_2(s-1,c-k) k I_{k,1}}{I_{s+c,1}} = \frac{2(s+c)}{s(s+1)\binom{s+c}{s}} \qquad \textbf{[6.13]}$$

with the last equality from equation 11 in Rosenberg (2003). Function f decreases with increasing s or c, as adding any lineage increases the chance of a monophyly-violating interclass coalescence.

T → ∞

As $T \to \infty$, because $\lim_{T \to \infty} g_{i,j}(T) = 1$ when $j = 1$, every branch exhibits complete coalescence. We define the minimal subtree with respect to S, $\mathcal{T}^*_{SC}$, as the smallest subtree of the species tree whose leaves contain all of the initial S lineages in the tree.

For large T, the monophyly probability depends on properties of $\mathcal{T}^*_{SC}$. To be monophyletic, the S lineages must encounter C lineages only above their root. If $\mathcal{T}^*_{SC}$ contains no C lineages, then complete coalescence in each branch implies monophyly of S lineages, and the monophyly probability is 1. If $\mathcal{T}^*_{SC}$ contains C lineages and is at a leaf, k, then the limiting probability is $f(S_k,C_k)$. Complete coalescence in every branch makes this leaf analogous to the root in the $T = 0$ case. Note that if $S_k > 1$ then the limit $f(S_k,C_k)$ lies in the interior of the unit interval. This result contrasts with Rosenberg (2003), where lineage classes correspond to species tree leaves and the $T \to \infty$ probability of E_S is 1. In our scenario, because multiple lineage classes are permitted at a leaf, a nonzero limit can be below 1.

If $\mathcal{T}^*_{SC}$ contains C lineages but is not a leaf, however, then complete coalescence in every branch implies that some proper subset of S lineages must coalesce with C lineages before all of the S lineages can coalesce with each other. In this case, the limiting monophyly probability is 0.

Finite, Nonzero T

The extreme cases assist in understanding the behavior of the probability of E_S for intermediate T. We enumerate the possible situations based on $\mathcal{T}^*_{SC}$, continuing to assume that relative branch lengths are fixed and that a changing tree height changes all branch lengths proportionally.

If $\mathcal{T}^*_{SC}$ contains no C lineages, then decreasing the tree height decreases the probability of monophyly by decreasing the time during which S lineages are able to coalesce with only themselves, eventually approaching a minimum $f(s,c)$ achieved at $T = 0$. Similarly, increasing T increases the monophyly probability toward 1 as $T \rightarrow \infty$.

If $\mathcal{T}^*_{SC}$ contains C lineages and is a leaf, then decreasing the tree height decreases the monophyly probability by decreasing the time before more C lineages are added to the population that contains the S lineages. Shrinking the tree also increases the expected number of additional C lineages introduced at species merging events, further decreasing the monophyly probability. The minimal probability of monophyly therefore occurs at $T = 0$. Similarly, increasing the tree height increases the probability of monophyly, approaching a maximal value as $T \rightarrow \infty$. Consequently, in this case, like in the previous case, the probability also increases monotonically in T.

If $\mathcal{T}^*_{SC}$ contains C lineages and is not a leaf, then the minimal probability of monophyly, approached as $T \rightarrow \infty$, is 0. As we will see in numerical examples, however, monotonicity of the monophyly probability with T is not guaranteed, and different initial sample sizes on the same species tree can generate different behavior.

Effect of Relative Branch Lengths

Next, to investigate the behavior of the monophyly probability as T increases, we devise a simple three-species, two-parameter scenario, subdividing the tree height T by a parameter r. We calculate the probability of E_S for different sample-size conditions, varying r and T.

Fig. 6.3 shows the species tree and its resulting monophyly probabilities for four representative initial conditions. For each lineage class, S and C, the four cases place one or more lineage pairs into the three species, using different placements across the four cases. The cases include scenarios in which at least one species contains both S and C lineages (B, D, E), in which one (C) or both lineage classes span multiple species

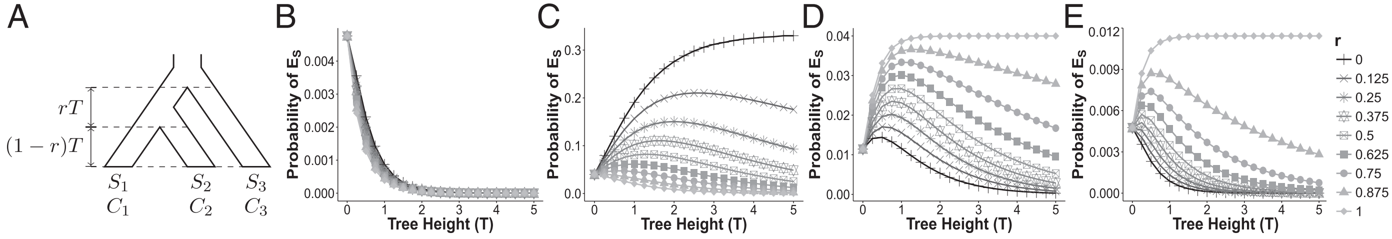


FIGURE 6.3 The effect on monophyly probabilities of changing two branch lengths in relation to each other. (*A*) Model species tree. If the branch length coefficient r is 0, then the tree has a polytomy, and if $r = 1$, then the tree reduces to a two-species tree. (*B–E*) The probabilities of E_S (Eq. 6.7) for monophyly of S for the tree in *A* under different scenarios: (*B*) $(S_1 S_2 S_3)$ = (2,0,2), $(C_1 C_2 C_3)$ = (2,2,2). (*C*) $(S_1 S_2 S_3)$ = (2,0,2), $(C_1 C_2 C_3)$ = (0,2,0). (*D*) $(S_1 S_2 S_3)$ = (2,2,0), $(C_1 C_2 C_3)$ = (2,0,2). (*E*) $(S_1 S_2 S_3)$ = (2,2,0), $(C_1 C_2 C_3)$ = (2,2,2).

(B, D, E), and in which the species containing *S* lineages are not monophyletic in the species tree (B, C).

The four cases (Fig. 6.3*B*–*E*) illustrate differences in the pattern of increase or decrease in the monophyly probability with changes in *r* at fixed tree height *T* (*Supporting Information*). In most cases with fixed *r*, the probability decreases to 0 with increasing *T*, although in some boundary cases with $r = 0$ and $r = 1$ that change the case for the limiting behavior with *T* (see above on $T \rightarrow \infty$), it approaches a positive value strictly within the unit interval. These scenarios highlight the fact that depending on the relative branch lengths and distribution of lineage classes across species, the monophyly probability can be monotonically increasing in *T*, monotonically decreasing, or not monotonic at all.

Effect of Pooling

Our next scenario simulates the difference between separating and pooling distinct species when computing monophyly probabilities, recalling that tests with more than two species have until now required the pooling of multiple clades (Carstens and Richards, 2007; Kubatko et al., 2011).

We consider four species trees with equal height and 12 lineages (Fig. 6.4). Six class-*C* lineages appear in one species descended from the root. The other six—the *S* lineage class—are evenly divided between one, two, three, or six other leaves. If we interpret the seven-leaf tree in Fig. 6.4*D* to be the "true" species tree, then the other trees represent pooling schemes, the two-leaf tree (Fig. 6.4*A*) being the only one possible to analyze using previous results.

Fig. 6.4*E*–*J* displays the probabilities of all possible monophyly events for each tree. For each event, pooling does not affect the extreme cases $T = 0$ and $T \rightarrow \infty$. For intermediate *T*, the monophyly probability for the *S* lineages decreases as pooling is reduced from the case in which the six class-*S* lineages are treated as belonging to a single species to the case in which each lineage is in its own species (Fig. 6.4*E*); the monophyly probability for *C* remains largely unchanged (Fig. 6.4*F*). As pooling is reduced, the probability of monophyly of only *S* and not *C* decreases (Fig. 6.4*G*), and that of only *C* and not *S* increases (Fig. 6.4*H*). The reciprocal monophyly probability decreases (Fig. 6.4*I*) and the probability of no monophyly increases (Fig. 6.4*J*).

In this scenario, the *S* and *C* lineages meet only at the species tree root, and the monophyly probabilities are determined by the numbers of lineages that reach the root. Coalescence is faster with more nonisolated lineages; pooling species together results in more coalescence events and fewer *S* lineages entering the root, increasing the probability of monophyly of both *S* and *C* lineages as well as the reciprocal monophyly prob-

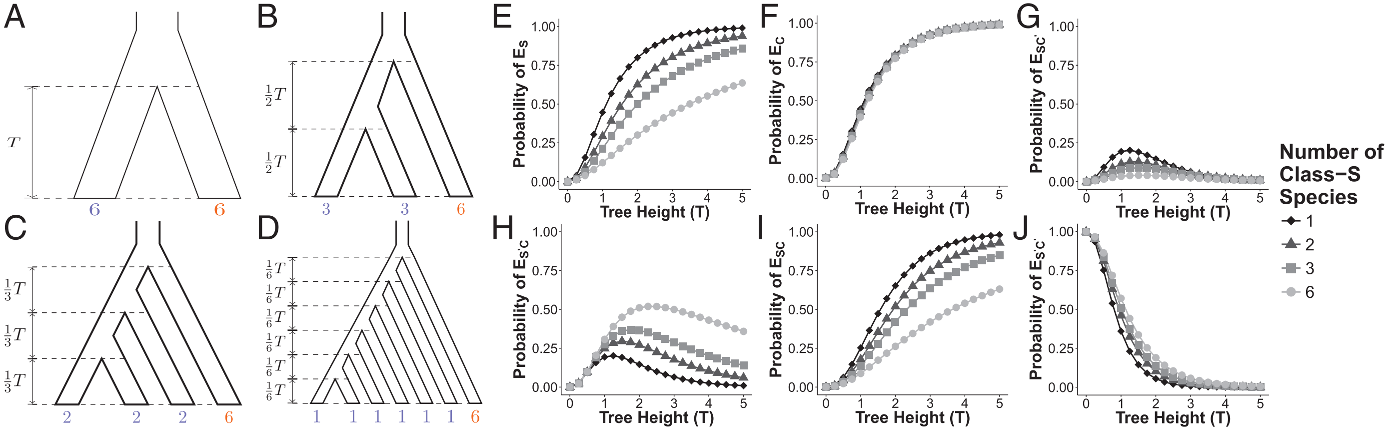


FIGURE 6.4 The effect on monophyly probabilities of pooling lineages from separate species. (*A–D*) Model species trees. Labels record numbers of input lineages (*S* in blue, *C* in orange). (*E–J*) Probabilities of monophyly events. The trajectories represent species trees with six class-*S* lineages evenly distributed over one (*A*), two (*B*), three (*C*), and six (*D*) species. (*E*) E_S (Eq. 6.7). (*F*) E_C (Eq. 6.8). (*G*) $E_{SC'}$ (Eq. 6.10). (*H*) $E_{S'C}$ (Eq. 6.11). (*I*) E_{SC} (Eq. 6.9). (*J*) $E_{S'C'}$ (Eq. 6.12).

ability (Fig. 6.4*E*, *F*, and *I*). Decreasing the number of S lineages at the root decreases the number of coalescences needed to produce E_S above the root, decreasing the chance of an interclass coalescence, whereas decreasing the number of S lineages does not change the number of coalescences necessary to produce E_C and has a smaller effect on its probability (cf. Fig. 6.4*E* and *F*). The probability for E_{SC} closely follows that of E_S, as production of reciprocal monophyly is limited by the monophyly of the individual classes.

As can be seen from the increase in probability for E_S as pooling is increased (Fig. 6.4*E*), the correct monophyly probability for clades that have been pooled tends to be lower than that obtained under a model where the pooled clades are treated as a single clade. The monophyly probability will likely be overestimated if populations are pooled.

Application to Data

To illustrate the empirical use of Eq. 6.7 and to test if our theoretical results reasonably replicate patterns in real data, we perform an analysis of monophyly frequencies using *Zea mays* maize and teosinte genomic data (Chia et al., 2012).

Hufford et al. (2012) analyzed 75 individuals from the data of Chia et al. (2012), considering four groups: teosinte varieties var. *parviglumis* ("parviglumis") and var. *mexicana* ("mexicana") and domesticated maize landraces ("landraces") and improved lines ("improved"). Modifying the estimated tree of individuals from figure 1 in Hufford et al. (2012) to make a model "species" tree the leaves of which are the four groups (Fig. 6.5*A*), we compute theoretical monophyly probabilities for each of the groups via Eq. 6.7. We also estimate the empirical frequency of monophyly for each group by randomly sampling individuals from each group, constructing multiple gene trees per sample from SNP blocks, and averaging frequencies of monophyly in the gene trees over the random samples. This procedure employs 100 unique random samples of eight individuals from the Hufford et al. subset, each containing two individuals from each of the four groups. Finally, we compare the observed and theoretical monophyly frequencies.

The monophyly frequencies appear in Fig. 6.5*B* and are summarized in *Table S2*. The theoretical frequencies predict the observations reasonably well. For each clade, especially parviglumis and mexicana, the mean observed monophyly frequency over 100 samples closely coincides with the theoretical monophyly probability (Fig. 6.5*B*). Although the theoretical probability is noticeably below the mean for the improved and landrace clades and above the mean for parviglumis and mexicana, it lies well inside the observed distributions.

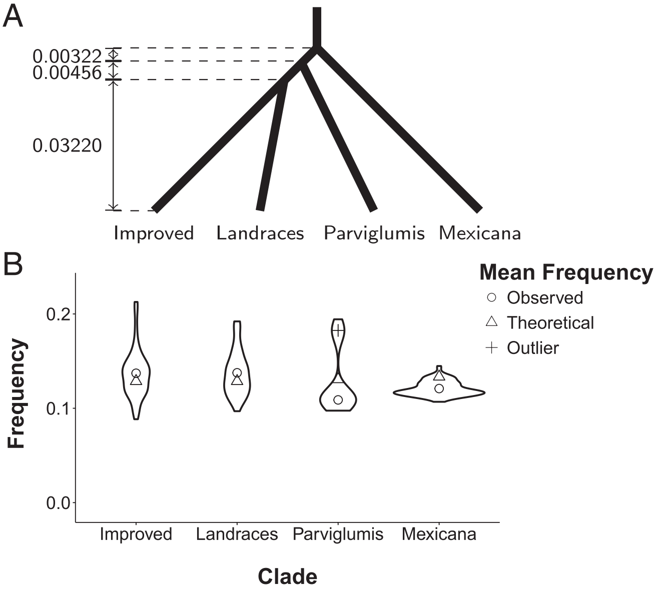


FIGURE 6.5 Monophyly frequencies in maize and teosinte. (*A*) Model species tree. (*B*) Violin-plot distributions across lineage subsamples of monophyly frequencies for four clades. Means of the observed distributions (excluding outliers for the improved and parviglumis clades) appear as circles and theoretical values appear as triangles. Outliers appear for a single point at frequency ~0.43 in the improved clade and for several points at frequency >0.17 in the parviglumis clade, with the cross indicating the mean of the parviglumis outliers (*Supporting Information*).

Eq. 6.7 relies on a model with selectively neutral loci and constant population size; a deviation from theoretical probabilities could suggest a violation of one of the model assumptions. Domestication imposes strong selection and population bottlenecks (Wang et al., 1999; Innan and Kim, 2004; Wright et al., 2005), factors that violate our model in a manner that would increase monophyly frequencies. Excess empirical monophyly in the improved and landrace clades (Fig. 6.*5B*, *Table S2*) is thus compatible with domestication in the history of these domesticated groups.

DISCUSSION

Extending a past computation (Rosenberg, 2003) from 2 to *n* species, we have obtained a general algorithm for the probability of any monophyly event of two lineage classes in a species tree of any size. In our generalization, unlike in previous calculations, no restriction exists on the class labeling of lineages, so that monophyly probabilities can be computed on samples aggregated across multiple species. We have uncovered behaviors absent in the two-species case, including nonmonotonicity of the monophyly probability in the tree height and positive limiting probabilities below 1. Both phenomena occur in scenarios newly possible to include in monophyly calculations, in which the lineage set whose monophyly is of interest spans multiple species, or in which lineages of at least one species span both classes.

We have used a pruning algorithm similar to other species tree computations (Efromovich and Kubatko, 2008; RoyChoudhury et al., 2008; Bryant et al., 2012; RoyChoudhury and Thompson, 2012; Stadler and Degnan, 2012; Wu, 2012) that evaluate a quantity at a parent node in terms of corresponding values for daughter nodes. In previous applications of this idea, the states recorded at a node are generally simpler than our input and output states. For example, in evaluating the time to the MRCA (Efromovich and Kubatko, 2008), they are one-dimensional; our approach instead tracks lineage classes as three variables, accommodating complex transitions that occur at interclass coalescences.

Previous work on monophyly probabilities has been limited to small numbers of species (Rosenberg, 2002, 2003; Degnan, 2010; Zhu et al., 2011; Eldon and Degnan, 2012). This limitation has forced investigators to either group multiple species together into a single clade (Carstens and Richards, 2007; Kubatko et al., 2011)—a choice that our tree-pooling experiment shows can overestimate monophyly probabilities—or to consider pairwise comparisons when multispecies analyses would be preferable (Baker et al., 2009; Neilson and Stepien, 2009; Bergsten et al., 2012). By identifying a bias that occurs when pooling distinct species in monophyly probability computations, our experiment suggests that pooling should be avoided when possible. Our results allow researchers to move beyond such simplifications by performing monophyly calculations in larger species groups.

One application of our results is to extend a test of a null hypothesis that an observed monophyletic pattern is due to chance alone (Rosenberg, 2007). This test has been available only in situations with species-specific lineages and two-species trees; it can now be extended to arbitrary trees and non-species-specific lineages. The results also provide a step toward computations for monophyly events on three or more lineage groups considered jointly.

As an empirical demonstration, we analyzed data from maize and teosinte, calculating theoretical and observed monophyly frequencies in four groups. The empirical frequencies generally match the predictions; frequencies exceeding predicted values in the domesticated species may reflect the fact that domestication bottlenecks and strong selection can violate our model in a manner that increases the likelihood of monophyly.

We note that our *Z. mays* results should be viewed with caution. We assumed a model of instantaneous divergence events without incorporating the subsequent gene flow that likely occurred in this system (Hufford et al., 2012). Furthermore, our model species tree contains uncertainty; however, we do not expect a bias in any specific direction to have resulted from its construction. Perhaps more seriously, we generated the model tree from the same study whose data we used for constructing gene trees. However, considerations of monophyly were irrelevant in producing the model tree, so that construction of the model did not guarantee the agreement we obtained between theoretical and observed monophyly.

The maize analysis illustrates how our framework can be used to study monophyly in multispecies genomic data. The formulas derived here allow for greater flexibility in studies of monophyly and its relationship to species trees, contributing to a more comprehensive toolkit for phylogeographic, systematic, and evolutionary studies.

MATERIALS AND METHODS

Maize Species Tree

We used maize HapMap V2 SNP data from www.panzea.org/#!genotypes/cctl (Chia et al., 2012) consisting of 55 million SNPs and small indels from 103 *Z. mays* inbred lines. To construct Fig. 6.5*A*, we determined relative branch lengths from figure 1 in Hufford et al. (2012). We chose a tree height of 0.04, measured in units of N generations, where N is the haploid population size, noting that a ~10,000-year domestication time (Hufford et al., 2012) translates via conversion factors calculated from figure 7 in Ross-Ibarra et al. (2009) (top panel, T_D column) to 0.036 units of N generations. We chose our root as the root of the Hufford et al. ingroup tree (second node from left in figure 1 of Hufford et al. (2012), call it x), our Parviglumis/Domesticated node as the MRCA of all domesticated lineages and parviglumis lineages TIL01, TIL03, TIL11, and TIL14 [$y = x_{LLLLL}$ in figure 1 of Hufford et al. (2012), oriented so that L is “down” rather than “left”], and our Landrace/Improved node as the MRCA of all domesticated lineages [y_L in figure 1 of Hufford et al. (2012)].

Maize Samples

We chose 100 samples of four lineage pairs, selecting randomly among 29 improved, 12 landrace, 8 parviglumis, and 2 mexicana individuals. We chose pairs within groups so that the Hufford et al. tree, a genomewide tree of individuals, restricted to each eight-lineage sample would display the model species tree in Fig. 6.5*A*, irrespective of which lineage in a pair was chosen to represent its group (*Supporting Information*).

Maize Gene Trees

The maize genome has ~2.3 × 10^9 bp (Schnable et al., 2009), with linkage disequilibrium (LD) decay at ~1,500 bp (Remington et al., 2001). For simplicity and to accommodate large quantities of missing data, despite genomewide variation in recombination rate and SNP density, we fixed a single block size for analyses throughout the genome. With ~5 × 10^7 SNPs in the dataset, SNP density per "LD block" is 32.6, which we round to 30. We divided the SNPs into nonoverlapping 30-SNP blocks and used every hundredth block in a concatenated genome starting from chromosome 1, resulting in ~6,000–7,000 gene trees per sample after removing blocks monomorphic in the sample and gene trees polytomic for the sample. We concatenated SNPs within blocks, computed blockwise Hamming distance matrices, and obtained gene trees using the *hclust* UPGMA (unweighted pair group method with arithmetic mean) clustering function in the R *stats* package. SNPs with missing data for a lineage pair were excluded in distance calculations.

Software Implementation

The *Monophyler* software package implementing Eqs. 6.7, 6.8, and 6.9 can be found at rosenberglab.stanford.edu/monophyler.html.

ACKNOWLEDGMENTS

We thank Jeff Ross-Ibarra for assistance with the maize data and John Rhodes and two reviewers for comments on a draft of the manuscript. We acknowledge support from NIH Grant R01 GM117590, NSF Grant DBI-1458059, a New Zealand Marsden grant, and a Stanford Graduate Fellowship.

7

Phylogeographic Model Selection Leads to Insight into the Evolutionary History of Four-Eyed Frogs

MARIA TEREZA C. THOMÉ* AND BRYAN C. CARSTENS[†][‡]

Phylogeographic research investigates biodiversity at the interface between populations and species, in a temporal and geographic context. Phylogeography has benefited from analytical approaches that allow empiricists to estimate parameters of interest from the genetic data (e.g., $\theta = 4Ne\mu$, population divergence, gene flow), and the widespread availability of genomic data allow such parameters to be estimated with greater precision. However, the actual inferences made by phylogeographers remain dependent on qualitative interpretations derived from these parameters' values and as such may be subject to overinterpretation and confirmation bias. Here we argue in favor of using an objective approach to phylogeographic inference that proceeds by calculating the probability of multiple demographic models given the data and the subsequent ranking of these models using information theory. We illustrate this approach by investigating the diversification of two sister species of four-eyed frogs of northeastern Brazil using single nucleotide polymorphisms obtained via restriction-associated digest sequencing. We estimate the composite likelihood of the observed data given nine demographic models and then rank these models using the Akaike information criterion. We demonstrate that estimating parameters under a model that is a poor fit to the data is likely to produce values that lead to spurious

---

*Departamento de Zoologia, Instituto de Biociências, Universidade Estadual Paulista, Campus Rio Claro, 13506900 Rio Claro, SP, Brazil; and [†]Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH 43210. [‡]To whom correspondence should be addressed. Email: carstens.12@osu.edu.

phylogeographic inferences. Our results strongly imply that identifying which parameters to estimate from a given system is a key step in the process of phylogeographic inference and is at least as important as being able to generate precise estimates of these parameters. They also illustrate that the incorporation of model uncertainty should be a component of phylogeographic hypothesis tests.

In biological populations with interbreeding individuals, allele frequencies will inevitably change with time, in both stochastic and systematic manners, through neutral and adaptive processes. These processes—genetic drift, gene flow, mutation, recombination, and natural selection—constitute observable phenomena that lead directly to population structure, population divergence, and eventually speciation. Phylogeography is ideally situated to investigate systems where the microevolutionary processes that act within gene pools begin to form macroevolutionary patterns and has been described as the bridge between population genetics and phylogenetics (Avise et al., 1987). The power of the discipline comes from the consideration of geographic origin of individuals and populations along the continuum between populations and species (Knowles, 2004; Hickerson et al., 2010).

Phylogeographic research has progressed through several stages since Avise et al. (1987) introduced the term. Initial studies were based on information that can be gathered from the genetic data under few assumptions, for example by calculating summary statistics or estimating gene trees. Inferences were then derived from qualitative interpretations about what that information implied about the evolutionary history of the system (e.g., Demesure et al., 1996; Bernatchez and Wilson, 1998). This approach has been criticized as being prone to overinterpretation, because researchers are inclined to propose more detailed and complex historical scenarios than are actually supported by the data (Knowles and Maddison, 2002). The general response to such criticisms has been the widespread adoption of model-based methods to analyze phylogeographic data, particularly models that incorporate coalescent theory (Kingman, 1982) to estimate parameters of interest under a formal framework. Model-based methods of phylogeographic inference clearly represent an advance to the field, but making inferences from these parameter estimates still forces researchers to make subjective decisions. Despite the potential complexity of the demographic models, the actual process of phylogeographic inference remains largely analogous to that of earlier investigations: The relative influence of evolutionary processes is derived from the magnitude of numeric values estimated for parameters that measure what the researchers believe to be important evolutionary processes. For example, subjective decisions regarding estimated rates of gene flow are commonly used to determine whether populations are reproductively isolated from their sister taxa (e.g., Dolman and Moritz, 2006) or conspecifics (e.g., Runemark et al., 2012).

Once efficient algorithms and computational power became available, researchers applied model-based methods to phylogeographic research with little hesitation (but see Templeton, 2010), with models implemented in software packages being particularly popular. For example, the paper describing a popular method that estimates temporal divergence with gene flow has been cited in more than 500 studies to date (Hey, 2010). Simulation-based techniques are also commonly applied to empirical systems, either to test competing hypotheses such as introgression and lineage sorting (e.g., Reid et al., 2012; Debiasse et al., 2014; Grummer et al., 2015) or to test phylogeographic hypotheses against a null model (e.g., Knowles, 2001; DeChaine and Martin, 2005; Smith et al., 2011). Such methods have been widely adopted by the phylogeographic community because model-based methods offer a path toward estimating putatively relevant parameters, and because the models themselves can be tailored to the particulars of a given system (e.g., Knowles, 2009b; Beaumont MA et al., 2010). Phylogeographic inferences are more transparent when based on parameters estimated under these models, and arguably less subjective. However, simply using a complex demographic model to analyze genetic data is not a guarantee that phylogeographic inferences will be correct.

In the cognitive sciences, researchers have long been mindful of confirmation bias, the tendency to interpret novel information in a manner consistent with preconceived ideas (Nickerson, 1998). People tend to seek out information that supports their preexisting beliefs and are unlikely to consider contradictory information. Particularly problematic is the primacy effect, in which the information that is learned first effectively has more emphasis than information that is obtained at a later date (Nickerson, 1998). Confirmation bias is likely prevalent in phylogeographic research (Carstens et al., 2013), influencing phylogeographic inference by shaping the very questions that are asked by researchers. For example, if initial investigations into a given system used gene trees and phylogenetic thinking, researchers may not consider population processes such as gene flow as being potentially important, and choose to estimate divergence times under a species tree model, which may not actually fit the data (e.g., Reid et al., 2014). Researchers working in temperate systems in the Northern Hemisphere may assume that postglacial expansion is an important process and choose to estimate effective population size under growth models (e.g., Kuhner, 2006), whereas those working on focal taxa that inhabit island systems are likely to consider dispersal to be a key process shaping allele frequencies, and estimate effective population sizes under migration models (e.g., Beerli and Felsenstein, 2001). Such assumptions will guide choices about which models and software should be used to analyze the data and might also bias their interpretation of the values of parameters estimated under these models. Objective assessment of model fit should be an important component of phylogeographic research, particularly in systems where there is little preexisting information about the demographic history.

WHAT IF THE PHYLOGEOGRAPHIC MODEL IS WRONG?

There is a great asymmetry in terms of the amount of available background information between model and nonmodel systems. In the extreme case of *Homo sapiens*, the analytical models used for data analysis are informed by the academic output of entire disciplines (e.g., anthropology) as well as thousands of previous genetic investigations. In contrast, the average phylogeographer likely knows very little about the focal organism before an investigation, save what can be inferred from its taxonomy and general habitat. This asymmetry is exacerbated for researchers interested in tropical diversity, which account for the vast majority of organisms: Chances are that even the most basic natural history traits (area of occurrence, density, feeding habitats, maturation age, and reproductive mode) are unknown to science. Given this paucity of information, how should researchers determine which models to use in data analysis?

In their review of statistical methods in phylogeography, Nielsen and Beaumont (2009) argue strongly that population parameters should be estimated under appropriate models to avoid bias in the parameter estimates: “A clear limitation of any model-based method is that the model might be wrong. In fact, the real complexity of the demography of natural populations is unlikely to be captured by any simple model we could propose. In some cases, this may not affect inferences much, but in other cases it will.” If phylogeographic inferences are largely derived from parameter estimates made under complex models, then such inferences are implicitly conditioned on the statistical fit of the model used to estimate these parameters to the empirical data collected from the focal system. To date, there has been too little attention devoted to methods for assessing the statistical fit of phylogeographic models to the data.

STATISTICAL FRAMEWORKS FOR PHYLOGEOGRAPHY

Phylogeographic research is a historical discipline rather than an experimental one, and evolutionary history cannot be replicated. Because the experimental controls used in classical hypothesis testing are not available (e.g., Neyman and Pearson, 1933), testing hypotheses, even with parametric simulation (e.g., Knowles, 2001; Carstens et al., 2004), forces phylogeography to conform to a statistical framework that may not be suited to historical research (Cleland, 2001). A more promising strategy for phylogeographic data analysis is to proceed by identifying which of many possible models of historical demography offer the best statistical fit to the observed data, rather than testing null hypotheses, where rejection only tells us that the model representing the hypothesis is a poor fit to the data. If the goal of phylogeography is to infer the evolutionary history of the focal taxon, then ranking a set of models that represent alternative evolutionary scenarios provides a rigorous tool for inference because it will help researchers to avoid confirmation bias. Because the parameters in each model correspond to various evolutionary processes, the relative influence of particular evolutionary processes to the empirical system can be assessed by considering the set of parameters included in the model that offers the best fit to the data. Model selection is a useful framework for phylogeographic inference because it offers an approach that accounts for the uncertainty in the models used to analyze the data.

MODEL SELECTION IN BAYESIAN AND INFORMATION THEORETIC FRAMEWORKS

Fagundes et al. (2007) provided a compelling example of phylogeographic research using model selection in a Bayesian framework, using approximate Bayesian computation (ABC) to evaluate alternative models of human demographic history. Inspired by this work, many researchers have applied a similar approach to a wide range of nonmodel systems (e.g., Tsai and Carstens, 2013; Espindola et al., 2014; Jamamillo-Correa et al., 2015; Peres et al., 2015; Vera-Escalona et al., 2015). However, as with any approach to data analysis, phylogeographic model choice using ABC has limitations, and decisions about which models to include in the comparison set can be challenging. Because ABC loses power to differentiate among models as the number of models in the comparison set increases (Pelletier and Carstens, 2014), one cannot easily evaluate large numbers of models. Fagundes et al. (2007) had the advantage of working in a model system where they could identify three types of models to test based on the results of hundreds of previous investigations, but the lack of similar information in nonmodel systems increases the odds of erroneous model choice and faulty phylogeographic inference.

A solution to evaluating a large number of models representing many possible demographic histories is to use information theory (Burnham and Anderson, 1998) to rank models. Information theory relies on the estimation of the Kullback–Leibler (Kullback and Leibler, 1951) information of a given model using the Akaike information criterion (AIC) (Akaike, 1973), and the subsequent ranking of all models in the comparison set. The model ranking is achieved by calculating the difference between the AIC score of a particular model and the best model in the set (e.g., $\Delta_i = \mathrm{AIC}_i - \min_{\mathrm{AIC}}$), and subsequent transformation to model likelihoods (w_i) by normalizing AIC differences across the set of R models such that they sum to 1.0 [$w_i = \exp(-1/2\Delta_i)/\sum_{r=1}^{R}\exp(-1/2\Delta_r)$; see Burnham and Anderson, 1998]. A reasonable interpretation of these model probabilities is that they correspond to posterior probabilities under a uniform prior distribution (Burnham and Anderson, 1998). Information theory is commonly used to select models of DNA nucleotide substitution for analyses of sequence data (as in the software ModelTest; Posada and Crandall, 1998), and has been effectively used to compare among a large number of models in this context. To date, information theoretic approaches have been used in phylogeography to choose the best of several isolation-with-migration models (e.g., Koopman and Carstens, 2010; Rittmeyer and Austin, 2015), to evaluate models of postglacial expansion and colonization (Carstens et al., 2013), and to evaluate models of source-sink migration (Beerli and Palczewski, 2010; Barrow et al., 2015). In this chapter, we briefly illustrate its application using data from the four-eyed frogs of northeastern Brazil.

CASE STUDY: THE *PLEURODEMA* SYSTEM IN THE BRAZILIAN CAATINGA

Pleurodema alium and *Pleurodema diplolister* are sister species of four-eyed frogs that inhabit the Caatinga in northeastern Brazil (Faivovich et al., 2012). The Caatinga is a widespread xeric biome, surrounded by the extensive mesic environments of the Amazon, Cerrado, and Atlantic Rainforest. Its climate is highly seasonal and unpredictable, with severe droughts and rainless years. As is typical of amphibians from xeric habitats, *Pleurodema* persist throughout most of the year by burrowing underground, becoming active only after seasonal heavy rains create ephemeral pools for breeding. Even though the life cycle in *Pleurodema* depends on precipitation, these frogs cannot maintain populations in more mesic biomes and its distribution is restricted to the Caatinga xeric habitat.

Floristically, the Caatinga is one of the isolated nuclei in the Seasonally Dry Tropical Forests (SDTFs) of South America. The history of the SDTFs is debated, with some evidence suggesting that they were formerly continuous and recently fragmented [during the Last Glacial Maximum (LGM); Prado and Gibbs, 1993], and other evidence favoring an older (Tertiary) fragmentation (Pennington et al., 2000). Environmental niche modeling results in contrasting maps ranging from a largely continuous to a fragmented Caatinga, depending on the approach used (Werneck et al., 2011; Collevatti et al., 2013). Regardless of the broader continental trends of the SDTFs, there is abundant geologic evidence that the Caatinga has been recurrently invaded (and at least partially replaced) by mesic forest throughout its history (de Oliveira et al., 1999; Auler et al., 2004).

P. alium and *P. diplolister* were recently the subject of phylogeographic investigation. Thomé et al. (2016) collected >350 samples, sequenced the mitochondrial cytochrome oxidase I (COI) gene, and genotyped 12 microsatellite loci. Using these data, they were able to confirm that the species were distinct at the genetic level (both at COI and microsatellite markers), and that they have partly sympatric distributions: *P. alium* is restricted to the southern Caatinga, whereas *P. diplolister* is widespread in the biome, occurring also in pockets of Caatinga embedded within the Cerrado (Fig. 7.1). The population genetic structure within the broadly distributed *P. diplolister* reflected the distribution of its sister species, in that the *P. diplolister* samples that were sympatric with *P. alium* formed a separate genetic cluster.

Given the available information, a wide range of evolutionary processes (and therefore parameters) could be incorporated into a demographic model of *P. alium* and *P. diplolister*. Temporal divergence likely represents an important component, supported by the deep divergence in the COI data (Thomé et al., 2016). Effective population sizes are likely to differ between species, because *P. diplolister* has a much larger geographic range than *P. alium*, and probably a corresponding difference in census population size. Although range size and effective population size are not necessarily correlated, the difference in geographic range provides justification for allowing for the possibility of differences in effective population size among species, so long as we assume that the mutation rate does not vary between species. In addition to the processes of temporal divergence and different population sizes, other evolutionary processes could be important: population size change within species (such as population bottlenecks or exponential population growth), gene flow, and/or natural selection.

We specified nine demographic models for analysis, which were designed to represent a range of demographic histories. All models included lineage divergence between the sister taxa *P. alium* and *P. diploister* and some combination of the following demographic processes: population expansion or contraction, population bottlenecks, gene flow, and population-specific θ values (Fig. 7.2). There are hundreds of ways that the divergence of two species from a common ancestor could be parameterized (see Pelletier and Carstens, 2014); here, we hope to specify models that span the range of possible models but include those that we believe to be plausible (e.g., we do not include *n*-island models that lack temporal divergence, because we consider divergence time to be an essential parameter to include in any model that contains sister species).

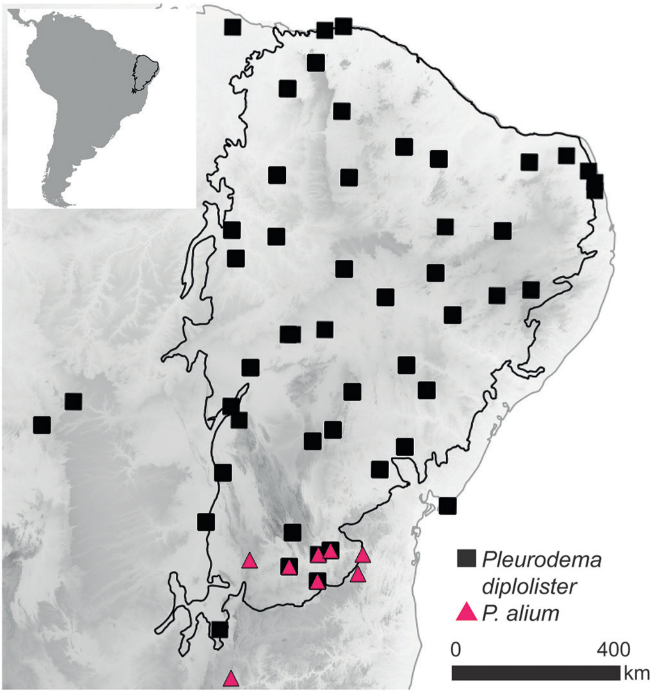


FIGURE 7.1 Map of the sampling localities. The outline of the Caatinga is shown on an elevation map of northeastern Brazil, where darker shading corresponds to higher elevation. *P. diplolister* localities are marked with a dark square, *P. alium* localities with a triangle.

Sampling and Molecular Protocols

We sampled 183 individuals of *Pleurodema* from 55 locations in the core, isolates, or peripheral regions of the Caatinga, comprising most of its distribution in the Caatinga biome (see Thomé et al., 2016). SNPs were collected via genomewide sampling using restriction enzymes (double-digest RADseq; Peterson et al., 2012). DNA digestion and barcode ligation were performed individually for each sample using 300 ng of freshly extracted DNA, the restriction enzymes Sbf1-HF and MspI, the ligation enzyme Ligase T4, and eight different barcoded Illumina adaptors. The digestion–ligation reactions were then pooled in groups of eight and purified with Agencourt AMPure beads, and PCR (12 cycles) was used to amplify the fragments containing barcodes using six different Illumina

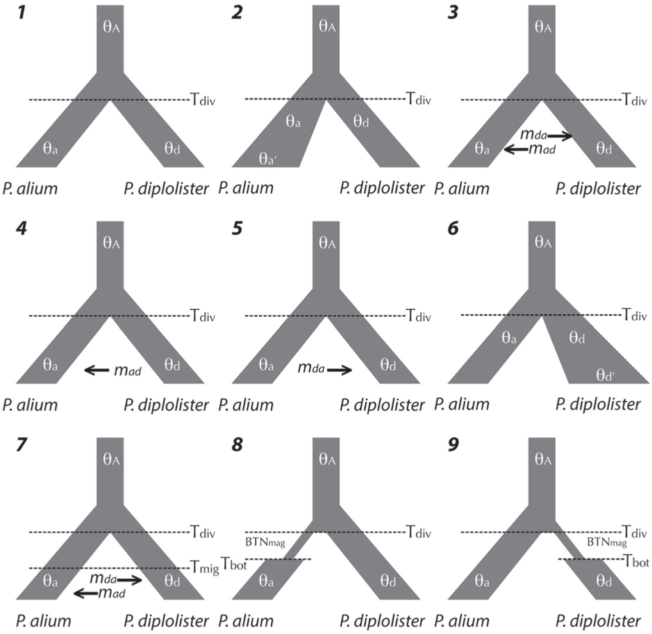


FIGURE 7.2 Nine demographic models used in model selection are shown. Parameter abbreviations include genetic diversity of *P. alium* and *P. diplolister* (θ_a, θ_d), ancestral genetic diversity (θ_A), the timing of population divergence (Tdiv), migration between diverging lineages (m_{ad}, m_{da}), the magnitude of the population bottleneck (BTN_{mag}), the timing of migration (T_{mig}), and bottlenecks (T_{bot}).

indexed primers and Phusion DNA polymerase. PCR products were quantified with Qubit Fluorometric Quantitation (Invitrogen), equimolar quantities of six groups containing eight samples each were pooled, and 250-bp to 500-bp fragments were selected using a Blue Pippin Prep. The fragment sizes were confirmed with an Agilent 2100 Bioanalyzer (Agilent), and 100-bp, single-end, sequencing reactions were conducted using an Illumina HiSEq 2000 at Beckman Coulter Genomics.

Data Processing

Illumina outputs from *Pleurodema* samples were processed using the pyRAD pipeline (Eaton, 2014). Except for the initial demultiplexing step, which was conducted separately on each library, we processed data for all samples together with the following parameter specifications: 10× minimal coverage, four or fewer unknown bases per sequence, minimum similarity of 0.90, a maximum ratio of shared polymorphisms of 20%, and a minimum coverage taxon of 70%. The number of reads that passed quality control was plotted against the number of loci obtained in each sample to establish a minimum number of reads for a sample to be considered. Because the number of loci stabilizes above 300,000 reads, we eliminated the 18 samples that were below this threshold before conducting a final SNP calling step in the remaining 165 samples. This scheme yielded 6,027 alignments containing SNPs.

Missing Data

After excluding SNPs that were possibly under selection (*Supporting Information*[1]), our dataset consisted of 5,810 sequenced regions containing one or more SNPs. However, every region was not sequenced in each sample. Population-level data collected using RADseq and related protocols typically consist of data matrices with some degree of missing data (e.g., Rubin et al., 2012; Wagner et al., 2013), and these missing data can lead to biased estimates of effective population size and other parameters (Arnold et al., 2013; Gautier et al., 2013). Missing data are likely to be particularly problematic for analytical methods that rely on estimates of allele frequencies because rare alleles may be undercounted. However, it is not clear how to best conduct analyses in a manner that accounts for the missing data. Missing data might be related to mutations in the recognition site of the enzymes, and removing all individuals that contain missing data about a certain threshold would be equal to removing the most divergent individuals, which could artificially homogenize the dataset and dramatically change the estimates of the number of rare alleles. Alternatively, removing all loci that contain missing data will dramatically reduce the size of any observed RADseq dataset and negate some of the advantages of collecting such data in the first place. Because we analyze our data using a method that relies on estimates of the population site frequency spectra (discussed below), it is important to account for missing data in a manner that does not bias our estimate of these frequencies. To accomplish this, we choose SNPs (one per locus) and individuals at random from our full data and then replicated this downsampling 10 times using a Python script provided by Jordan D. Satler, The Ohio State University, Columbus, OH (*Supporting Information*). After the downsampling procedure, our replicate data matrixes consisted of approximately one-third of the total SNPs in one-half of the individuals and enabled us to calculate confidence intervals by comparing estimated parameters across replicates.

---

[1] Supporting information for this chapter, which includes *Table S1*, is available online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1601064113/-/DCSupplemental.

Model Selection

We estimated the composite likelihood of the probability of the observed data given the specified model using fastsimcoal2 (FSC2) (Excoffier et al., 2013). FSC2 estimates parameters specified by the user (including $\theta = 4Ne\mu$, population size change, gene flow, and population divergence) from the site frequency spectrum (SFS). Demographic processes will influence the site frequency distributions; for example, gene flow will produce an abundance of shared SNPs, population bottlenecks will result in a reduction of genetic diversity and thus fewer low-frequency SNPs, and so on. After the demographic model is specified, FSC2 selects initial parameter values at random from a range specified by the user and simulates data using the demographic model and parameter values. Composite likelihoods are calculated following Nielsen (2000), who demonstrated that there is a relationship between the branch lengths of the genealogy and the probability of observing a SNP of a certain frequency distribution. Parameter optimization was conducted using the Brent algorithm implemented in FSC2, which identifies parameter values that maximize the likelihood estimate of the observed SFS given the demographic model. Finally, the maximized likelihood observed across all iterations is used in model comparison.

Using FSC2, the analysis of each of the 10 downsampled datasets was replicated 50 times (Excoffier et al., 2013). The individual run settings of each replicate included 100,000 simulations for the calculation of the composite likelihood and 50 cycles of the Brent algorithm (for parameter optimization). FSC2 analyses were conducted using massively parallel computing resources provided by the Ohio Supercomputer Center. After the maximum likelihood was estimated for each model in every replicate, we calculated the AIC scores and converted to model probabilities as above. This transformation allows us to measure the probability of each model given the observed data across replicates (e.g., *Table S1*), which we interpret as a measure of the degree of support for a particular model following Anderson (2008).

RESULTS AND DISCUSSION

The results of the FSC2 analysis were consistent in the sense that only three models, all isolation with migration, have any appreciable model probability (i.e., >0.001; *Table S1*). The model with ongoing gene flow from *P. diplolister* to *P. alium* has the highest model probability. The secondary contact model and the model asymmetric gene flow between *P. diplolister* and *P. alium* have similar log-likelihoods given the data to the best model but lower AIC scores due to having additional parameters. Additionally, parameter estimates suggest that these models may be more similar than they seem (Table 7.1). For example, in the secondary contact model (i.e., model 7) parameter estimates of the time that gene flow begins are closer to the divergence of these species from their common ancestor than to the present, and in model 3 (i.e., the model with asymmetric gene flow) the rate of gene flow from *P. alium* to *P. diplolister* is estimated to be much lower than the rate of migration in the opposite direction (although these estimates are not perfectly comparable because the duration of gene flow is not the same under these models). Because of the similarity in parameters estimated by these models, our phylogeographic inferences are based on model-averaged parameter values (i.e., the value of a given parameter estimated under a particular model weighted by the model probability of that model, averaged across models that share the particular parameter; Table 7.1).

There are several striking features of the divergence with gene flow models. Assuming a mutation rate of 2.1×10^{-9} substitutions per site per generation (Gottscho et al., 2014) to convert parameter estimates, the ancestral effective population size (averaged across replicates and models) was estimated to be small (~12,500 individuals). *P. alium* and *P. diplolister* began to diverge from their common ancestor during the last glacial cycle of the Pleistocene (~58,900 y BP) but continued to exchange alleles via migration. The rate of migration into each species from the other was not equal; roughly 10 times as many *P. diplolister* migrants entered the *P. alium* gene pool than the reverse ($2\ Nm_{da} = 0.78$; $2\ Nm_{ad} = 0.07$). Finally, whereas the current effective population size of each species is estimated to be larger than the ancestral population, current effective population sizes in *P. diplolister* are substantially larger than in *P. alium* ($N_d = 1.34 \times 10^6$; $N_a = 6.9 \times 10^4$), consistent with differences in their geographic ranges.

Perhaps the most surprising result from our analysis is how much parameter estimates depend on the model used to estimate the parameters. For example, divergence time is estimated to be two orders of magnitude more ancient when estimated under model 6 (~3,280,000 y BP) than under the best-ranked model (Table 7.1), whereas the ancestral effective population size was estimated to be much smaller (2.65×10^2). Given the lack of previous estimates for these parameters in this system, there would be little reason to be suspicious of these values absent an assessment of model fit. This example illustrates the importance of performing phylogeographic model selection before any attempt to make inferences about the evolutionary history of a system, especially those based on parameter estimates.

TABLE 7.1 Comparison of Parameter Estimated Using FSC2 Under Four Models

Model (w_i)	$N_{\text{ancestral}}$	N_{alium}	$N_{diplolister}$	T_{div}	$2Nm_{12}$	$2Nm_{21}$	N_{found}	T_{event}	G_{exp}
3 (0.21)	1.48×10^4	6.86×10^4	1.34×10^6	5.86×10^4	0.069	0.904	—	—	—
4 (0.56)	1.43×10^4	$6.98 \times 10v$	1.33×10^6	5.88×10^4	0.072	—	—	—	—
7 (0.23)	5.59×10^3	6.92×10^4	1.36×10^6	5.93×10^4	0.078	0.738	—	2.97×10^4	—
6 (0.00)	2.65×10^2	8.20×10^4	2.3×10^6	3.28×10^6	—	—	73	1.09×10^4	-4.6×10^{-5}
Model average	1.25×10^4	6.94×10^4	1.34×10^6	5.89×10^4	0.073	0.783	—	—	—
Lower confidence interval	1.12×10^4	6.61×10^4	1.31×10^6	5.75×10^4	0.063	0.643	—	—	—
Upper confidence interval	1.37×10^4	7.26×10^4	1.37×10^6	6.02×10^4	0.083	0.887	—	—	—

NOTES: Shown for all models are estimates of population sizes ($N_{\text{ancestral}}$, N_{alium}, $N_{diplolister}$), population divergence (T_{div}), and gene flow (2Nm). Additional parameters estimated under model 6 are the size of the founding population (N_{found}), the time that expansion begins (T_{event}), and the magnitude of population size change (G_{exp}). The additional parameter estimated under model 7 is the time that gene flow begins (T_{event}). The model probability of each model is shown in parentheses after the model number. Also shown for six parameters are the model averaged estimate values and the upper and lower 90% confidence interval. All parameters were converted to real units assuming a mutation rate of 2.1×10^{-9}. See *Table S1* for additional information regarding the results from all models.

Phylogeographic Inferences

There are several advantages to basing phylogeographic inferences on the results of model selection exercises. Such analyses allow researchers to identify which evolutionary processes have shaped genetic diversity. In *Pleurodema*, the divergence of the sister taxa *P. alium* and *P. diplolister* is occurring despite ongoing gene flow. This inference stems directly from results of the model selection exercise: All of the models that have good AIC scores and thus receive any appreciable support include some gene flow between these species. This inference is not based on the magnitude of the parameter estimates, but solely on the inclusion of the gene flow parameters in the highest-ranked models. In addition, the results of the model selection analysis prevent us from overinterpreting our data (*sensu* Knowles and Maddison, 2002). In *Pleurodema*, previously collected evidence suggested that population expansion could represent an important feature of this system (Thomé et al., 2016), but none of the population size change or bottleneck models offered a good fit to the empirical data. As much as we expected expansion to be a dominant force shaping these data, there is no evidence for the influence of this process in the SNP dataset. We attribute this discrepancy to one of two causes. It could be that there is an actual difference in the signal between the SNP data analyzed here and the microsatellite and COI data analyzed by Thomé et al. (2016). Each of these markers evolves at a different rate and thus will be informative at different timescales. Thus, it is possible that faster markers perform better in detecting demographic expansions as recent as 4,240 y BP (de Oliveira et al., 1999). However, because these analyses differed in the number of individuals included (approximately three times as many in the microsatellite analysis as here), as well as in details of each analysis, this difference could result from some combination of these differences.

What factors may have caused the initial divergence of *P. alium* and *P. diplolister*? Results from analyses of environmental (climatic) niche modeling provide two important clues. First, the environmental niche of *P. alium* does not differ from that of *P. diplolister* (see Box 7.1). This makes it unlikely that these species are undergoing adaptive diversification, a result that is supported by an outlier loci analysis (e.g., a BayeScan analysis detects only 14 out of 6,027 loci as being potentially under selection; *Supporting Information*). Second, species distribution modeling supports the hypothesis of a dynamic distribution for the Caatinga, as the predicted distribution of these species has changed over the last 130,000 years, including being notably smaller at the mid-Holocene, and somewhat reduced at the LGM (Fig. 7.3). These historical distributions are at odds with previous paleomodeling of the SDTFs but consistent with the palynological record, which indicates that the present-day distribution of the Caatinga established very recently in the late Holocene (4,240 y BP; de Oliveira et

BOX 7-1
Environmental Niche Models

We gathered 51 georeferenced occurrence points (2 for *P. alium* only, 44 for *P. diplolister* only, and 5 for both species) from sequenced samples collected in the core area of the Caatinga at a minimum distance of 8 km between points. We extracted climate information from 19 layers of bioclimatic variables available at the WorldClim website and used principal component analysis of occurrence data to compare their niches (Broennimann et al., 2012). Niche overlap was high ($D = 0.95$) and the hypothesis of niche equivalency could not be rejected ($P = 0.99$). The niches of the two species are more similar than would be by chance ($P = 0.0198$). To estimate past distributions we constructed correlative maps of potential distribution with the maximum entropy algorithm (Phillips et al., 2006) and projected the model to past environmental conditions of the mid-Holocene (6,000 y BP) LGM at 21,000 y BP (MIROC4m general circulation model, Pliocene Model Intercomparison Project), and last interglacial (LIG) at 120,000 y BP (Otto-Bliesner et al., 2006). The study area encompasses current and putative past Caatinga distributions according to previous studies (Werneck et al., 2011; Collevatti et al., 2013). We selected eight uncorrelated variables (Pearson correlation < 0.7) downloaded from Bioclim at 2.5 arc minutes resolution: mean diurnal range, isothermality, temperature seasonality, annual precipitation, precipitation of driest month, precipitation seasonality, precipitation of warmest quarter, and precipitation of coldest quarter. We used random training-test percentages (70% of observations for model training and 30% for model testing), the auto features function, and the default regularization multiplier. The high mean value for the area under the receiver operating characteristics curve (AUC = 0.960, SD = 0.007, $n = 100$) indicates that the model performance was satisfactory. The most important variable was annual precipitation (evaluated with 100 iterations).

al., 1999). The dynamic range of these species supports the idea that these lineages have been periodically fragmented, possibly isolated, with secondary contact inhibiting the formation of reproductive isolation.

New Data, Better Methods, and Improved Inferences from Nonmodel Organisms

One of the pressing issues facing the discipline of phylogeography in the past was the limited amount of genetic data that could be collected from most systems, and the poor quality of parameter estimates that

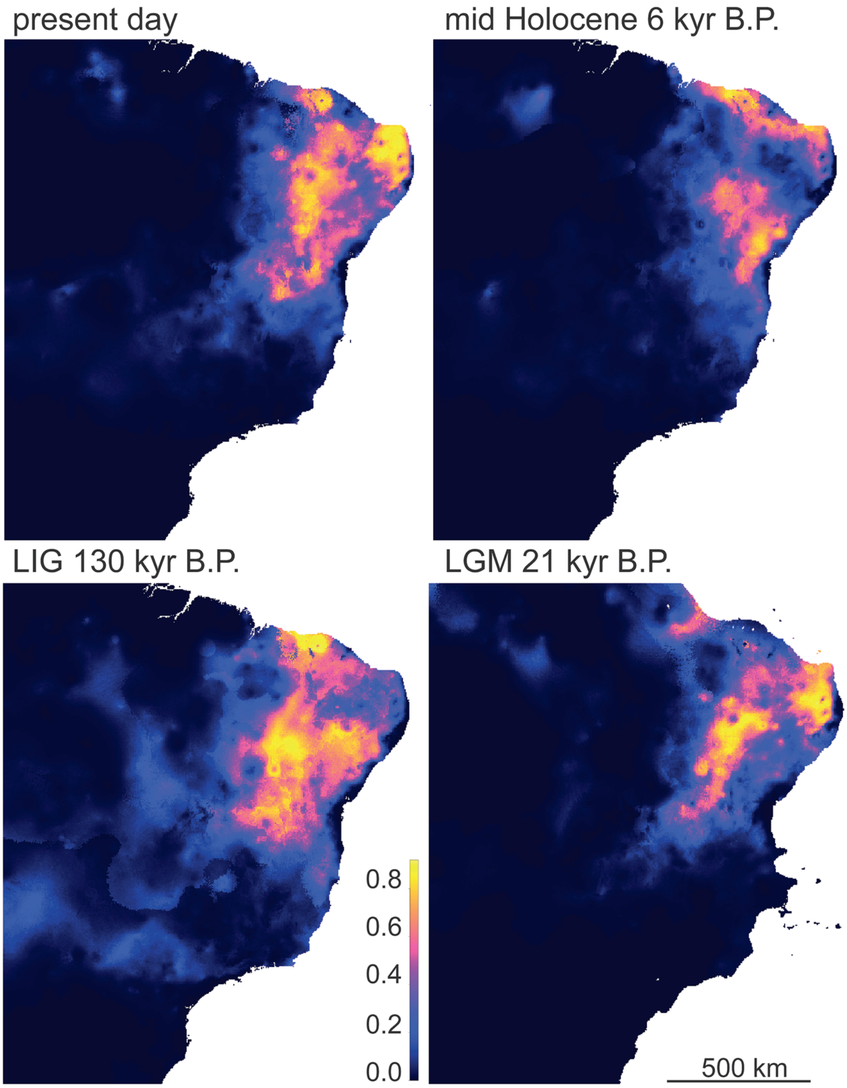


FIGURE 7.3 Projections of suitable habitat for *P. alium* and *P. diplolister*. Shown clockwise from upper left are estimates of the current ecological niche, as well as projections of this niche onto past conditions of the mid-Holocene, the LGM, and the LIG.

resulted from analysis of these data (Edwards and Beerli, 2000; Brumfield et al., 2003; Felsenstein, 2006). In the last decade, advances in sequencing technology have led to dramatic improvements in the amount of data that can be collected from nonmodel systems (McCormack et al., 2013; Garrick et al., 2015). Given modest levels of funding, researchers can now collect more data from any system than are likely required to accurately estimate parameters of interest (e.g., Felsenstein, 2006; Carling and Brumfield, 2007). With next-generation datasets, phylogeography is well positioned to address a more important question: Which parameters are important to estimate in a given system? Whereas many of the methods applied by phylogeographic investigations were developed initially for the analysis of data from model systems (e.g., Excoffier et al., 2013), scientists working in nonmodel systems have been forced to confront the question of model fit, and in response they are developing creative solutions to identifying models that fit a particular system.

Some approaches to model selection are built into the framework of existing analytical methods. For example, IMa (Hey and Nielsen, 2007), which implements a divergence with gene flow model, can be used to conduct model selection using either likelihood ratio tests (e.g., Hey and Nielsen, 2007) or information theoretic approaches (Carstens et al., 2009). Similarly Migrate-*n* (Beerli and Palczewski, 2010), which implements an *n*-island model, can be used to select among many migration models (Beerli and Palczewski, 2010; Barrow et al., 2015). In addition, there are a number of approaches to species delimitation that incorporate model selection. These include methods that identify the optimal species delimitation using likelihood ratio tests (Knowles and Carstens, 2007), reversible-jump Markov chain Monte Carlo (Yang and Rannala, 2010; Solís-Lemus et al., 2015), information theory (Ence and Carstens, 2011), ABC (Camargo et al., 2012), and marginalized likelihoods (Leaché et al., 2014). Methods for analyzing comparative phylogeographic data are also under active development, including the use of hierarchical Bayesian models to test simultaneous divergence (Hickerson et al., 2007; Oaks et al., 2013) or simultaneous population expansion (Chan et al., 2014; Xue and Hickerson, 2015).

Although methods that implement model selection are extremely useful, they lack the flexibility of simulation-based approaches, which provide researchers with the capacity to customize their models to the particular details of nearly any empirical systems. ABC continues to be a useful approach to model selection, particularly when implemented in computational environments such as R (e.g., Csilléry et al., 2010) that can be easily used by researchers. Other methods are available that calculate the probability of SNP data. In addition to FSC2, used here, model selection can be conducted using diffusion approximation in the software dadi (Gutenkunst et al., 2009).

CONCLUSIONS

Testing the statistical fit of our models given the data enabled us to address a major limitation of model-based phylogeography (Beaumont MA et al., 2010). By deriving our phylogeographic inferences from parameters estimated under suitable models, we avoided confirmation bias and overinterpretation. Parameter estimation was of central importance to our phylogeographic inference process, but only after we made an objective determination about which parameters to estimate. Perhaps the greatest advantage of this approach to phylogeography is that while the inferences themselves do not rely solely on parameter estimates, the parameters that are estimated via model averaging are likely to be more representative of the actual population values. It is incumbent on researchers who do not conduct model selection as part of their phylogeographic investigations to ask whether their phylogeographic inferences are based on a model of historical demography that is appropriate for their empirical system.

ACKNOWLEDGMENTS

We thank Célio F. B. Haddad, Miguel T. Rodrigues, José Pombal, Jr., and Marcelo Nápoli for donation of samples; ICMBio for the collecting permit (30512); and Francisco Brusquetti for help in the field. We also thank members of the B.C.C. laboratory and two reviewers for comments that improved this manuscript prior to publication. Financial support was provided by Fundação Grupo Boticário de Proteção à Natureza Grant 0909_20112 and São Paulo Research Foundation Grants 2012/50255-2, 2011/51392-0, and 2013/09088-8. Computational resources were provided by the Ohio Supercomputer Center. Files used in the analysis have been deposited at DRYAD (doi:10.5061/dryad.8m6j3).

8

Toward a Paradigm Shift in Comparative Phylogeography Driven by Trait-Based Hypotheses

ANNA PAPADOPOULOU*† AND L. LACEY KNOWLES*‡

For three decades, comparative phylogeography has conceptually and methodologically relied on the concordance criterion for providing insights into the historical/biogeographic processes driving population genetic structure and divergence. Here we discuss how this emphasis, and the corresponding lack of methods for extracting information about biotic/intrinsic contributions to patterns of genetic variation, may bias our general understanding of the factors driving genetic structure. Specifically, this emphasis has promoted a tendency to attribute discordant phylogeographic patterns to the idiosyncracies of history, as well as an adherence to generic null expectations of concordance with reduced predictive power. We advocate that it is time for a paradigm shift in comparative phylogeography, especially given the limited utility of the concordance criterion as genomic data provide ever-increasing levels of resolution. Instead of adhering to the concordance-discordance dichotomy, comparative phylogeography needs to emphasize the contribution of taxon-specific traits that will determine whether concordance is a meaningful criterion for evaluating hypotheses or may predict discordant phylogeographic structure. Through reference to some case studies we illustrate how refined hypotheses based on taxon-specific

---

*Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, Ann Arbor, MI 48109; and †Department of Integrative Ecology, Estación Biológica de Doñana, Consejo Superior de Investigaciones Cientificas, 41092 Seville, Spain. ‡To whom correspondence should be addressed. Email: knowlesl@umich.edu.

traits can provide improved predictive frameworks to forecast species responses to climatic change or biogeographic barriers while gaining unique insights about the taxa themselves and their interactions with their environment. We outline a potential avenue toward a synthetic comparative phylogeographic paradigm that includes addressing some important conceptual and methodological challenges related to study design and application of model-based approaches for evaluating support of trait-based hypotheses under the proposed paradigm.

Paralleling its critical role in identifying biogeographic phenomena by common structuring of genetic variation across taxa, comparative phylogeography has the potential to offer unprecedented insights about the taxa themselves. We here make an argument that such a paradigm shift, in which the contribution of biotic attributes is a focal point, as opposed to an emphasis on abiotic factors, is not only a valuable endeavor in itself, but it is also critical to (*i*) understanding the relative contributions of abiotic/extrinsic (e.g., geographic, geological, or climatic history) vs. biotic/intrinsic (e.g., ecological or life history traits; hereafter referred to as biotic) factors in structuring genetic variation and divergence across taxa, as well as (*ii*) determining the relative prevalence of deterministic processes versus stochasticity in the evolutionary history of taxa, and we argue that (*iii*) the key to meaningful insights is not just a function of the statistical support we might have for a model, but it also depends on the creativity of biologists to identify hypotheses that are relevant to the structuring of genetic variation. That is, the development of biologically informed model-based approaches and the application of refined hypotheses based on taxon-specific traits provide a needed balance in comparative phylogeography for addressing fundamental questions in ecology and evolution given that the field has traditionally focused (both conceptually and methodologically) on the concordance criterion as a means for identifying biogeographic phenomena. We illustrate the unique insights that a comparative phylogeographic framework can offer about the taxa themselves through reference to some case studies in which the study design provides a predictive framework for interpreting phylogeographic concordance. We present a variety of examples to highlight some of the interesting ecological and evolutionary questions that can be addressed using refined hypotheses based on species-specific traits, but also to emphasize the broader significance of this proposed paradigm shift to ecology and evolution.

HISTORICAL EMPHASIS ON CONCORDANCE

Comparative phylogeography emerged three decades ago as an integrative approach to historical biogeography (Arbogast and Kenagy, 2001;

Zink, 2002), and has since offered invaluable insights into the factors that shape spatial and temporal patterns of genetic variation. An inherent emphasis on concordance in patterns of genetic variation as a criterion for evaluating hypotheses has accompanied the development of the field since its infancy. Specifically, phylogeographic congruence among several codistributed taxa has provided key evidence of the impact of biogeographic barriers, geological events, or past environmental change on today's distribution of genetic variation (Avise, 2000; Hewitt, 2004). Such contributions are exemplified in classical comparative phylogeographic work, such as the study of coastal taxa from the southeast of the United States, where genetic surveys revealed concordant phylogeographic discontinuities between Gulf of Mexico and Atlantic populations in 12 of 19 species (including terrestrial, freshwater, and marine taxa), consistent with the vicariant separation of populations by Pleistocene sea level change (Avise, 1992). Such highly influential studies clearly demonstrated the strength of the concordance criterion, and their legacy is evident in that tests of concordance became the central paradigm in comparative phylogeographic analyses.

With concordance as a metric for testing expectations for patterns of genetic variation, comparative phylogeographic analyses have been used to address a broad array of ecological and evolutionary questions, from inferring Quaternary refugia and assigning conservation priorities (Marske et al., 2012; Dauby et al., 2014) to evaluating the stability of ecological associations between interacting organisms (Stone et al., 2012; Ornelas et al., 2013). Commonly applied tests include the evaluation of spatially concordant phylogeographic breaks (Carstens et al., 2005; Soltis et al., 2006), temporally synchronous population divergence (Carnaval et al., 2009; Fouquet et al., 2012), or concerted demographic expansion/contraction (Smith et al., 2011; Chan et al., 2014) among members of an ecological community. Methodological advances of the last decade, especially coalescent-based tools for hypothesis testing and parameter estimation (Knowles, 2009b; Hickerson et al., 2010) that include statistical assessment of concordance across taxa, such as the widely used hierarchical approximate Bayesian computation (hABC) (Hickerson et al., 2006b; Huang et al., 2011), also promoted the utility of the concordance criterion. At the same time, such statistical frameworks have been critical for overcoming past perceptions of phylogeography as primarily a descriptive field that lacked rigor (Hey and Machado, 2003).

LIMITATIONS OF STRICT ADHERENCE TO CONCORDANCE CRITERIA

Although the phylogeographic concordance criterion, enhanced by methodological advances that provide statistical rigor to inferences, has undoubtedly made an enormous contribution to our understanding of the historical/extrinsic factors structuring genetic variation, it has also had several consequences that have arguably limited the potential insights from comparative phylogeographic analyses. For example, it has promoted a tendency to disregard discordance as uninteresting and attribute lineage-specific patterns to stochastic effects. Likewise, with the concordance criterion as a key basis for inferences, a focus on abiotic factors, and consequently on the spatial and temporal scales where the geologic/historical events are likely to have impacted codistributed taxa similarly, came to predominate the field (Avise, 2008). Such an emphasis has contributed to an imbalance in the perceived relative importance of extrinsic factors in structuring genetic variation (as opposed to the influence of intrinsic factors) (Zamudio et al., Chapter 11, this volume). Moreover, the adherence to the concordance criterion results in tests that are often too generic and thus have reduced predictive power (Papadopoulou and Knowles, 2015b), where the rejection of concordance is often met with ad hoc explanations for the observed discordance. Last, without a correspondingly well-developed statistical framework for making predictions about discordance (Massatti and Knowles, 2016), extracting information about biotic contributions to patterns of genetic variation among codistributed species also becomes challenging. This limitation is especially critical given that if species-specific traits actually have a significant impact, then by definition, discordance will characterize genetic variation among codistributed species. Below we expand the evidence about the limitations of a strict adherence to a concordance criterion in comparative phylogeography, especially in respect to its broader impact on ecological and evolutionary thought.

Interpreting Discordance

As a consequence of the tendency to attribute discordance and lineage-specific patterns to stochastic effects (Kropf et al., 2003; Marske et al., 2012; Ornelas et al., 2013), deterministic processes that cause discordant patterns among codistributed taxa (Massatti and Knowles, 2014) may be disregarded. Invoking the idiosyncratic nature of history makes discordant phylogeographic structure uninteresting in a comparative framework (i.e., it implies there are no conclusions to be drawn about deterministic processes from discordant patterns of genetic variation), limiting the generalizations comparative phylogeography might provide about the influence of biotic factors on genetic structure. Likewise, although the "concerted vs. individualistic responses" dichotomy can offer important insights (e.g., about the stability of ecological communities through time) (Stewart et al., 2010), by considering trait-mediated responses as "individualistic," we might miss an opportunity for drawing lessons from comparative phylogeographic study about fundamental ecological and evolutionary principles. For example, when evaluating the effect of Pleistocene glaciations on the evolutionary history of codistributed montane plant taxa, spatially discordant phylogeographic patterns are often deemed inconclusive (Kropf et al., 2003), whereas they could provide insights into the interaction between species' ecology and climatic change (Massatti and Knowles, 2014). However, the potential insights into trait-mediated effects of climate change go unrealized without a conceptual (and methodological) framework for discordant genetic structure, which is significantly underdeveloped compared with that for phylogeographic concordance.

Imbalance in the Relative Contribution of Abiotic vs. Biotic Factors

The inherent emphasis on phylogeographic concordance that has dominated the field since its infancy has additionally created an unbalanced focus on the spatial and temporal scales where the impact of geologic/historical events is likely to predominate. It has been demonstrated in a range of different systems that contrasting conclusions might be drawn about the prevalence of concordant or discordant patterns depending on the spatial and temporal scale in which taxa are studied. For example, comparative phylogeography of two alpine butterfly species in the Rocky Mountains (DeChaine and Martin, 2005) or two saproxylic springtail species in southeastern Australia (Garrick et al., 2008) revealed shared responses to Pleistocene climatic change across broad spatial scales, but species-specific patterns at finer spatial, and recent temporal, scales. Similarly, comparisons among three symbiotic fungal species associated with the mountain pine beetle in western Canada (Roe et al., 2011) identified congruent phylogeographic patterns at a broad geographical scale vs. incongruent structuring at a finer scale, which was linked to species-specific traits such as transmission mode or environmental tolerance. By emphasizing concordance, most phylogeographic studies tend to adhere to broader temporal and spatial scales where concordant patterns are more prevalent, whereas contemporary effects across local scales are treated separately within a landscape genetics framework (Manel et al., 2003). Although we might expect that processes at fine scales should predict patterns at broader scales, the links between mechanism (e.g., traits that limit or promote movement) and pattern (phylogeographic signal) may not be evident when relying on distinct analytical frameworks. Without a shared framework, the relative influence of extrinsic vs. intrinsic factors in structuring genetic variation, as well as the influence of contemporary vs. past processes, remain unknown, and difficult to separate from the scale of a study (Rissler, Chapter 16, this volume).

Reduced Predictive Power of Generic Expectations of Concordance

Null expectations of concordance among taxa with disparate traits are prevalent in the comparative phylogeographic literature (Taberlet et al., 1998; Riddle et al., 2000; Stone et al., 2012; Oaks et al., 2013; Ornelas et al., 2013), even though it is well documented that certain ecological or life history traits can greatly influence genetic structure (Hamrick and Godt, 1996; Alvarez N et al., 2009), especially with respect to species' responses to biogeographic barriers and climatic change (Stewart et al., 2010; Papadopoulou et al., 2009). A main concern with such generic hypotheses of global phylogeographic concordance is their reduced predictive power (Paz et al., 2015); that is, rejection of concordance leads to rather inconclusive statements that do not offer particularly meaningful insights (Papadopoulou and Knowles, 2015b). Refinement of the expectation for concordance is needed if concordance itself is to be a meaningful metric for making inferences about the processes structuring genetic variation (Box 8.1). In other words, support (or the lack there of) for a process may simply be a function of how the concordance criterion is applied (i.e., if it is used as a generic expectation without regard to species-specific traits), as opposed to whether or not a process may drive divergence (Box 8.1).

Methodogical Developments Reinforce the Knowledge Gap Created by Adherence to Concordance

Given that interpreting discordance has never been a focus in the field, currently available and newly developed comparative phylogeographic methods (Hickerson et al., 2006b; Oaks, 2014; Xue and Hickerson, 2015) are designed to quantify congruence, rather than to gain insights from discordant patterns. As a consequence, the standard comparative phylogeographic toolbox indirectly encourages users to emphasize congruence and disregard discordance as uninteresting. For example, when applying coalescent-based hypothesis testing using hABC to assess simultaneous population divergence across a set of disparate codistributed taxa (Hickerson et al., 2006b; Huang et al., 2011), the rejection of a global model of temporal congruence will unavoidably emphasize idiosyncratic aspects of history, if the test is not pursued under an appropriate study design allowing for an improved predictive framework based on taxon attributes (Papadopoulou and Knowles, 2015b). Similarly, some common practices in hABC implementations adopt computational shortcuts that also limit the interpretation of discordance. For example, the default option of resorting the summary statistics vector (e.g., in msBayes) (Hickerson et al., 2006b; Huang et al., 2011) reduces the number of simulations required for hABC analyses but comes at the expense of losing information critical to interpretation (Oaks, 2014; Papadopoulou and Knowles, 2015b). This lack of methodological attention for testing predictions of discordance therefore reinforces the current knowledge gap about biotic contributions to patterns of genetic variation.

INSIGHTS GAINED FROM DISCORDANCE HIGHLIGHT THE PROMISE OF A PARADIGM SHIFT

Although there is a historically pervasive trend to attribute phylogeographic discordance to stochastic effects, a number of empirical studies have attempted to provide deterministic explanations based on taxon-specific traits. Morphological or life-history traits directly related to dispersal potential, such as body size, wing development, reproductive mode, or larval type (Whiteman et al., 2007; Papadopoulou et al., 2009; Fouquet et al., 2012; Paz et al., 2015), have been commonly used to interpret contrasting responses to biogeographic barriers. In some cases, habitat (or microhabitat) preference has also been associated with differences in spatial genetic structure, especially when (micro)habitats differ in their long-term stability (Abellán et al., 2009; Papadopoulou et al., 2009; Papadopoulou and Knowles, 2015b) or their responses to environmental change (Beavis et al., 2011; Massatti and Knowles, 2014). Alternatively, discordance in spatial genetic structure may be linked to the degree of ecological specialization (e.g., in habitat, diet, or host) (Rocha et al., 2002; Whiteman et al., 2007; Dellicour et al., 2015) and differential responses to climate change among taxa may reflect differences in thermal traits (Moussalli et al., 2009; Moritz et al., 2012).

The range of empirical systems and suite of traits that predict species-specific patterns of genetic variation highlight the potential insights phylogeographic discordance could provide about biotic/intrinsic factors structuring genetic variation. However, this aspect still remains marginalized in comparative phylogeographic research. Specifically, species-specific or trait-based explanations are typically applied in an ad hoc manner (i.e., after discordance is observed), in contrast to the more synthetic perspective applied to inferences about historical abiotic factors in which predictions are made a priori and tested within a well-developed statistical framework. Instead, the primary context for statistical tests of trait-based hypotheses has been in comparative phylogenetics, where certain clade-specific ecological traits have been linked with increased diversification rates (Phillimore et al., 2006; Rabosky et al., 2007). The success of such inferences in macroevolutionary treatments highlights how important the biotic influence in species divergence may be. However, these tests do not provide explicit links with the underlying microevolutionary divergence processes themselves. Comparative phylogeography has the potential to bridge this gap, providing insights about the role of ecological and life

BOX 8.1
Refined vs. Generic Hypotheses in hABC to Evaluate Climate-Driven Diversification

hABC analyses are widely used in comparative phylogeography to evaluate scenarios of climate-driven diversification by testing for simultaneous divergence among population pairs of codistributed taxa (Carnaval et al., 2009; Bell et al., 2012; Hope et al., 2013). Here we highlight two case studies that applied hABC to island taxa to assess the hypothesized "species-pump" action of rising and falling sea levels under the Pleistocene Aggregate Island Complex (PAIC) (Brown and Diesmo, 2009) model of diversification, which is expected to produce a pattern of temporally clustered divergence times coinciding with high sea level periods (Oaks et al., 2013) (Fig. 8.1*A*). Consider the case in which initial hABC analyses of 22 taxon pairs spanning five orders of terrestrial vertebrates sampled from six island pairs across the Philippine archipelago fit a model of simultaneous divergence, seemingly providing support for the "species-pump" diversification hypothesis (Oaks et al., 2013). However, when the authors performed a suite of simulation-based power analyses and discovered inherent bias in inferences because of improper priors on demographic and divergence-time parameters in the msBayes program (Hickerson et al., 2014; Oaks et al., 2014), reanalysis showed divergence times were not clustered. The merit of identifying methodological biases not withstanding, equivalent scrutiny should be given to the metric used to evaluate support for the species-pump diversification hypothesis, that is, the concordance criterion. Failure to identify clustered divergences across taxa with highly disparate dispersal capabilities and ecological traits that were sampled from different islands is not surprising and is arguably too generic of an expectation for drawing meaningful conclusions about the role of island connectivity cycles in driving diversification.

Instead of applying a generic expectation of concordance, refined hypotheses based on species-specific traits might be considered. For example, Papadopoulou and Knowles (2015b) generated refined expectations of concordance (Fig. 8.1*B*) to test the species-pump diversification hypothesis. Specifically, hypotheses of concordance were restricted to closely related darkling beetles species based on (*i*) bathymetric data that identified islands where sea level shifts would have produced similar degrees of connectivity and (*ii*) ecological traits that would have mediated the effects of sea level shifts (namely, preferences for either sand or soil; two habitats that differ in their stability over time). When hABC was applied on the full set of 13 taxa, there was moderate support for a model with two divergence events, but with considerable uncertainty and variation in divergence time estimates, as expected due to habitat-mediated differences in local extinction rate and dispersal propensity. However,

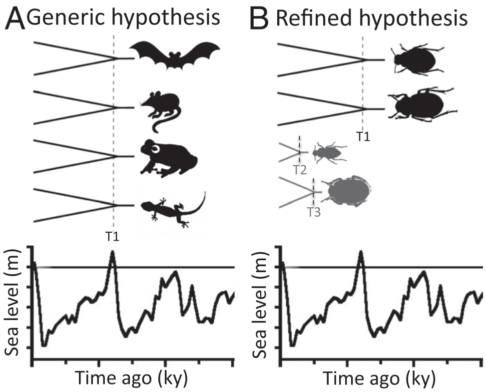


FIGURE 8.1 Generic vs. refined hypotheses to test the role of sea level changes in driving diversification of island taxa. (*A*) A generic hypothesis assumes simultaneous divergence across taxa with disparate ecological traits and dispersal abilities. (*B*) A refined hypothesis limits the expectations of concordance to taxa subject to similar levels of population connectivity and persistence based on their habitat preference (top beetles in black) and excludes closely related taxa subject to high local extinction rates that would supersede divergence driven by past sea level change (lower beetles in gray). Temporal concordance in population divergence among ecologically similar taxa provides support for the role of island connectivity cycles in driving diversification. At the same time, discordance between sets of similar taxa with different habitat preferences identifies habitat stability as a key factor structuring genetic variation in this island system (Papadopoulou and Knowles, 2015b).

when the expectations of concordance in the hABC analyses were refined based on habitat association, a model with a single divergence event was supported across all six stable-habitat specialists, demonstrating that the support for sea level shifts as drivers of divergence is a function of the null model used in the hABC analyses. This example illustrates the reduced predictive power of generic hypotheses of concordance in hABC analyses in comparison with a refined hypothesis based on taxon-specific traits. Moreover, together the examples highlight how a bias in our understanding of the factors driving divergence may not be limited to algorithmic issues associated with comparative phylogeographic analyses, but may also be introduced by how we apply the concordance criterion to make inferences about the factors driving divergence.

history traits in driving divergence at the microevolutionary level. However, this will require a shift from the primarily historical focus and strict adherence to concordance.

TOWARD A TRAIT-BASED COMPARATIVE PHYLOGEOGRAPHY

As highlighted above, case studies provide useful insights for focal taxa and demonstrate the effect of species-specific traits on patterns of genetic variation. We are not discounting the important contributions of such work (Zamudio et al., Chapter 11, this volume), which actually motivates the paradigm shift proposed here. Likewise, our attention to the limitations of the concordance criterion should not be viewed as a general call for abandoning the approach. To the contrary, our intention is to call for a conceptual and methodological shift, in which the broader significance of biological traits in shaping patterns of genetic variation can be revealed from comparative phylogeography, paralleling the field's contribution to understanding the role of historical abiotic factors (Avise, 2008). As phylogeography is being transformed by the unprecedented resolution of genomewide data for demographic inference and fine-scale population structure (Papadopoulou and Knowles, 2015a; Xue and Hickerson, 2015; Knowles et al., 2016), thus removing previous data-related constraints (Garrick et al., 2015), there is an opportunity for the field to offer unprecedented insights into the interactions of taxa with their environment and to make equally important contributions to the field of ecology as it has traditionally made to historical biogeography. Specifically, the multitaxon comparative framework and the historical perspective offered by comparative phylogeography, when combined with a thorough consideration of ecological and life-history traits, provides the potential to disentangle the relative contributions of historical vs. contemporary and abiotic vs. biotic factors to the structuring of genetic variation. However, there are both conceptual and methodological challenges to this endeavor. Here we focus on how (*i*) the study design is critical to providing insights about the biotic factors that might structure genetic variation and (*ii*) performing analyses in a model-based framework is essential for evaluating the fit of data to species-specific predictions under biologically informed hypotheses. Below we describe a synthetic model-based comparative phylogeographic paradigm that considers both biotic and abiotic factors structuring genetic variation. We fully acknowledge that this is not the only approach that might be applied. In fact, the full potential of this paradigm shift will not be realized without additional conceptual and methodological development from a diversity of researchers, in much the same way that comparative phylogeography evolved gradually to provide a robust framework for understanding the role of abiotic factors in structuring genetic variation.

The Importance of the Study Design

The insights provided by testing the role of biotic factors in structuring genetic variation are a direct function of a researcher's thorough knowledge of a study system, including both the thoughtful selection of hypotheses and an appropriate experimental design that can actually capture the processes of biological interest. Different approaches might be used to design comparative phylogeographic studies for testing biologically informed hypotheses. If the goal is to address questions about the possible role of a specific ecological trait in structuring genetic variation, it might be useful to focus on taxa that differ only in the targeted ecological trait (Papadopoulou and Knowles, 2015b; Massatti and Knowles, 2016). With such a study design, refined hypotheses of concordance based on taxon-specific traits can be used, thereby providing a better predictive framework for comparative phylogeographic tests (Box 8.1). This approach can be applied to any set of carefully selected taxa, although for more distantly related taxa, differences in addition to the targeted ecological traits could complicate interpretations. Alternatively, study designs for testing the role of biotic factors in structuring genetic variation might be based upon predictions of phylogeographic discordance. For example, species-specific predictions for patterns of genetic variation (Knowles and Alvarado-Serrano, 2010) might be used to evaluate how the response to climate change may differ in a deterministic manner depending upon key ecological traits (Box 8.2), rather than interpreting the lack of concordance among taxa as a rejection of shared glacial refugia (Knowles et al., 2016). Of course when designing a study, especially for model-based tests (Knowles, 2009b), it is imperative that the biological processes of interest are represented by the suite of hypotheses tested, as there is no way to rule out additional hypotheses that have not been evaluated. Phylogeographic analyses do not identify "the evolutionary history" of taxa per se (Knowles and Maddison, 2002); instead the insights of model-based approaches rest with demonstrating the relative support of one hypothesis relative to others where each of the different hypotheses relevant to evaluating a question are represented in the study (Massatti and Knowles, 2016).

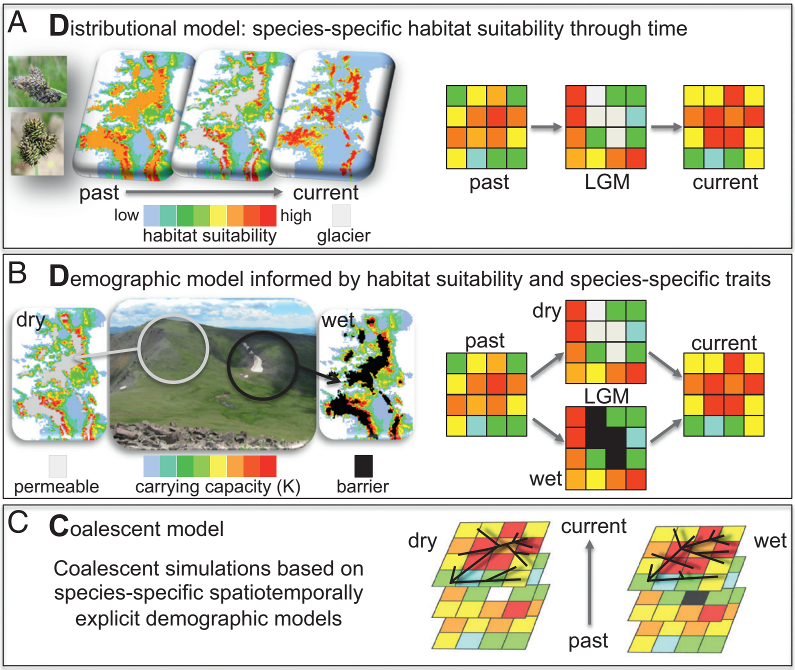
A Model-Based Framework for Generating Species-Specific Predictions

The proposed paradigm shift in comparative phylogeography will require an appropriate statistical framework to accommodate and interpret phylogeographic discordance. Here we focus exclusively on model-based inference, following in the tradition of statistical phylogeography (Knowles, 2009b), given that our objective is to make inferences about causation. The utility of correlative approaches for generating hypotheses should not be overlooked, but neither should their limitations for infer-

BOX 8.2
iDDC as a Framework for Generating Species-Specific Predictions Under Biologically Informed Hypotheses

The iDDC modeling approach (He et al., 2013) generates species-specific expectations for patterns of genetic variation under a broad range of biologically informed hypotheses, which can then be tested against the empirical genetic data using ABC (Beaumont et al., 2002). In particular, it can be used to evaluate questions that focus on spatial or temporal changes in habitat suitability or to accommodate habitat heterogeneity and generate more realistic patterns of genetic variation than a simple diffusion process (e.g., when evaluating models about population expansion or colonization dynamics; Fig. 8.2).

A range of taxon-specific attributes can be translated into a suite of alternative models for model selection (e.g., isolation by distance, isolation by environment, or heterogeneity in the stability of habitats over time) (He et al., 2013). Alternatively, within a comparative phylogeographic framework,



this approach can be used to generate contrasting expectations for patterns of genetic variation among species that differ in species-specific ecological traits (e.g., models in which a geographic region remains habitable or acts as a barrier) (Massatti and Knowles, 2016). In this context (Fig. 8.2), the iDDC approach can be used to make predictions about discordance (i.e., to evaluate whether trait-mediated deterministic processes contribute to observed discordance in patterns of genetic variation), making it an especially useful tool in comparative phylogeography.

FIGURE 8.2 The iDDC procedure generates species-specific predictions for patterns of genetic variation, which when coupled with approximate Bayesian computation (ABC), can be used to test biologically informed hypotheses about the effect of taxon-specific traits on phylogeographic structure. For example, it has been used to address questions on how the effects of climate change differ between wet- vs. dry-adapted montane sedges from the Rocky Mountains (Massatti and Knowles, 2016). The two closely related and codistributed Carex species have very similar distributional models (*A*), based on environmental niche modeling (Massatti and Knowles, 2014). However, when considering microhabitat affinity, different carrying capacities are predicted across the glaciated area at LGM (*B*), and coalescent simulations produce distinct expectations for patterns of genetic variation across the spatiotemporally dynamic landscape (*C*). Model comparisons using ABC supported a barrier model (i.e., zero carrying capacity in the glaciated region at LGM) for the wet-adapted species, which was presumably displaced during glacial periods due to accumulation of snow in wet microhabitats (e.g., drainages). On the contrary, a permeable model (i.e., allowing for nonzero carrying capacity in the glaciated region) is supported for the dry-adapted species, which could have persisted in situ within the glaciated areas, as drier microhabitats (e.g., ridges) remained relatively free of persistent snow (Massatti and Knowles, 2016). Photos reproduced from Massatti and Knowles (2014) with permission of the publisher.

ring processes from patterns of genetic variation. For example, general linear models (GLMs) may be used to assess the relative importance of individual variables as predictors of spatial genetic variation (Paz et al., 2015). However, correlative approaches may misidentify the impact of candidate factors, especially when multiple processes can produce similar patterns (e.g., when multiple predictor variables are correlated), or when certain assumptions about the association between genetic variation and a predictor variable are violated. Therefore, tests of the contribution of taxon-specific traits to the structuring of genetic variation should be based on model-based approaches that provide explicit links between processes and patterns of genetic variation (He et al., 2013). In particular, testing hypotheses about the intrinsic factors that might structure genetic variation requires generating species-specific predictions (Box 8.2). Recently developed methodologies integrating distributional, demographic, and coalescent modeling (iDDC) (He et al., 2013) with ABC (Beaumont et al., 2002) provide a framework for evaluating the relative probability of alternative hypotheses based on species-specific expectations under a diverse array of models that can accommodate differences in population dynamics over space (e.g., habitat heterogeneity) or time (e.g., shifting climatic conditions), but also differences in taxon-specific demographic processes because of disparate natural histories or ecologies (Box 8.2). With the analytical tools to test more complex models (Excoffier et al., 2013), and the power to distinguish among alternative models with genomic data, biologically informed hypotheses are poised to provide insights into questions that could not be addressed before.

Methodological Challenges for Trait-Based Comparative Phylogeography

As with any model-based approach for phylogeographic inference, there are important methodological and conceptual challenges with such tests of taxon-specific traits. For example, it is not trivial to identify from an essentially infinite set of hypotheses, those that will be the focus of statistical evaluation (Knowles, 2009b). Even though it might be possible to compare large sets of hypotheses without using any a priori biological knowledge of the study system (Pelletier and Carstens, 2014; O'Meara et al., 2015), that might not be particularly meaningful. The insights provided by selecting one model over another will be limited if the models only vary in nonkey parameters that do not make any important difference for their biological interpretation (Massatti and Knowles, 2016). In other words, without knowing what the question is, it is not clear what insights can be provided by biologically "naïve" hypotheses (i.e., models chosen without a biologically motivated hypothesis). There are of course situations where distinguishing among alternative hypotheses via some model selection procedure (Pudlo et al., 2016) is not the primary goal, but the aim is instead to estimate a demographic parameter, such as divergence time, taking into account the uncertainty about the actual demographic history (Thomé et al., 2016). In such cases, the model itself is actually a nuisance parameter, and therefore, considering an array of models with the inclusion/exclusion of other demographic factors might be desirable so that the robustness to a variety of potential histories is accounted for.

Decisions about what biologically informed hypotheses to test need to be coupled with appropriate methods for assessing the fit of the data to the models; that is, even if one model is more probable than another, they might both fit the data poorly (Wegmann et al., 2006). Practical challenges with such model-based tests may include computational limitations for complex models (e.g., using parameters from spatially explicit demographic models to inform the coalescent process for thousands of independent loci unavoidably requires substantial amounts of computational resources) or methodological constraints related with the use of ABC (e.g., model posterior probabilities poorly evaluated by ABC techniques) (Pudlo et al., 2016). Likewise, potential problems might be introduced when translating biological hypotheses into models for statistical tests. Specifically, the support (or lack thereof) for alternative hypotheses might reflect differences in the extent to which taxa conform to assumptions used to generate species-specific predictions, rather than key differences in the models themselves (Massatti and Knowles, 2016).

CONCLUSIONS

After three decades of comparative phylogeography providing invaluable insights into the biogeographic factors structuring genetic variation, now is the time for the field to consider a shift that has the potential to offer equivalent insights about the biotic component: the taxa themselves and the lessons they can provide on ecological and evolutionary thought. Advances in genomic datasets will no doubt push comparative phylogeography away from the concordance-discordance dichotomy, as with increasing amounts of data, what appeared as congruent patterns of divergence or phylogeographic structure may reflect the lack of resolution or power to detect discordance (Knowles, 2009b; Moritz et al., 2016). As such, a shift toward a trait-based paradigm in comparative phylogeography may become more broadly relevant, even if the focus stays on inferences about historical/abiotic factors. That is, with the increased resolution of genomic data for resolving incongruent aspects of genetic structure that were not detectable in the past, a model-based framework for extracting the signature of shared responses to a barrier, as opposed to the taxon-specific contributions, may become critical. We outlined a proposal of refined hypotheses based on species-specific traits to improve the predictive power of comparative phylogeography, as well as aspects in need of methodological development for this paradigm shift to realize its full potential. Building on the foundations of phylogeography that underlie the field's prominence in ecology and evolution, its integrative approach of drawing on input from multiple disciplines, the biologically informed hypotheses that are central to trait-based comparative phylogeography will continue to be an integral part of the field's continued success in the future.

ACKNOWLEDGMENTS

We thank Qixin He, Rob Massatti, and Jeet Sukumaran for helpful discussions, three anonymous reviewers for constructive comments, and the organizers, Francisco Ayala and John Avise, and the participants of the symposium whose work helped motivate our work. This work was funded by National Science Foundation Grant DEB 1118815 (to L.L.K.).

9

Reticulation, Divergence, and the Phylogeography–Phylogenetics Continuum

SCOTT V. EDWARDS,*§ SALLY POTTER,†‡
C. JONATHAN SCHMITT,* JASON G. BRAGG,†‡
AND CRAIG MORITZ†‡

Phylogeography, and its extensions into comparative phylogeography, have their roots in the layering of gene trees across geography, a paradigm that was greatly facilitated by the nonrecombining, fast evolution provided by animal mtDNA. As phylogeography moves into the era of next-generation sequencing, the specter of reticulation at several levels—within loci and genomes in the form of recombination and across populations and species in the form of introgression—has raised its head with a prominence even greater than glimpsed during the nuclear gene PCR era. Here we explore the theme of reticulation in comparative phylogeography, speciation analysis, and phylogenomics, and ask how the centrality of gene trees has fared in the next-generation era. To frame these issues, we first provide a snapshot of multilocus phylogeographic studies across the Carpentarian Barrier, a prominent biogeographic barrier dividing faunas spanning the monsoon tropics in northern Australia. We find that divergence across this barrier is evident in most species, but is heterogeneous in time and demographic history, often reflecting the taxonomic distinctness of lineages spanning it. We then discuss a variety of forces generating reticulate patterns in phylogeography, including

---

*Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138; †Research School of Biology, The Australian National University, Acton, ACT 2601, Australia; and ‡Centre for Biodiversity Analysis, Acton, ACT 2601, Australia. §To whom correspondence should be addressed. Email: sedwards@fas.harvard.edu.

introgression, contact zones, and the potential selection-driven outliers on next-generation molecular markers. We emphasize the continued need for demographic models incorporating reticulation at the level of genomes and populations, and conclude that gene trees, whether explicit or implicit, should continue to play a role in the future of phylogeography.

Phylogeography is being revolutionized by a whole-genome perspective driven by next-generation sequencing (NGS) in combination with development of coalescent-based methods of analysis within and among species. The classical phylogeographic foundation from which genome-scale phylogeography has grown was established in the decades spanning the early 1980s' emphasis on animal mtDNA (Avise et al., 1987; Moritz et al., 1987) to the mid-2000s, just before the first genomewide surveys of genetic variation in humans (Novembre et al., 2008; Novembre and Di Rienzo, 2009). By the early 2000s, phylogeographic surveys of nonmodel species typically included a handful of loci, mostly using methods that facilitated a locus-by-locus phylogeographic analysis (Dolman and Moritz, 2006; Brito and Edwards, 2009). There are now a growing number of studies realizing a distant goal of phylogeography, geographically informed whole-genome resequencing (Jones et al., 2012; Wallberg et al., 2014), as well as many more sampling subgenomes through varied approaches (Potter et al., 2016; McCormack et al., 2012; Smith et al., 2014a; Harvey and Brumfield, 2015). With the expansion to genomewide analyses afforded by NGS, phylogeographic analysis has necessarily expanded its analytical toolkit.

The increasingly routine analysis of genome-scale data has blurred the disciplinary boundaries between phylogeography and its sister discipline, population genetics, and has allowed phylogeography to contribute to endeavors such as scans for selection and association mapping (Edwards et al., 2015). Indeed, with burgeoning data and increasing applications of related analytical tools, such as site-frequency spectra and coalescent simulations, we can ask whether and how phylogeography is now distinct from population genetics (Edwards et al., 2015). We contend that there is still value in the original conception of phylogeography as a bridge between population biology and phylogenetics (Avise et al., 1987) (Fig. 9.1). This bridge can be thought of across geography and time, as is often the case with practitioners, or across gradients of migration rates and linkage disequilibrium (Slatkin, 2008), with the former decreasing and the latter increasing from the population to phylogenetic scale. That phylogeography sits centrally in this process-oriented space emphasizes the importance of understanding interactions between reticulation (gene flow/introgression and recombination), drift, and protracted isolation.

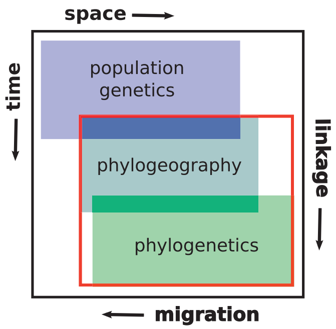


FIGURE 9.1 Diagram classifying the disciplines of population genetics, phylogeography, and phylogenetics. Traditionally, we think of these respective disciplines as being concerned with variation among organisms arising over short, intermediate, and long temporal (and often spatial) scales. Increasingly, with large quantities of data, there are opportunities to classify studies according to the way different processes are inferred to have shaped datasets. For example, it is likely that migration among populations is common in "population genetics" datasets and rarer in phylogenetics. Similarly, recombination is likely to reduce the detectable effects of linkage in population genetic datasets, such that the effects of linkage likely lead to larger haplotype blocks in studies at the "phylogeographic" scale. In this way, different studies might form, and next-generation methods might facilitate, a continuum from population genetics to phylogenetics. In this review, we focus on studies spanning the part of this continuum spanning phylogeography and shallow phylogenetics, indicated by the red box.

This combination of processes sets phylogeography apart from traditional population genetics and phylogenetics.

Scanning entire genomes of closely related organisms has unleashed a level of heterogeneity of signals that was largely of theoretical interest in the PCR era. This genomic heterogeneity is profoundly influencing our basic concepts of phylogeography and phylogenetics, and indeed our views of speciation processes. It is now routine to encounter a diversity of gene trees across the genome that is often as large as the number of loci surveyed (Song et al., 2012; Fontaine et al., 2015). Aside from variation induced by the coalescent process within and across species, we are only beginning to understand how such gene tree heterogeneity arises (Fontaine et al., 2015; Nater et al., 2015). Recognition of this heterogeneity has driven the development of phylogenetic methods for accommodating such conditional independence of gene trees, so-called "species tree" methods (Liu et al., 2009a, 2015; Knowles and Kubatko, 2010; Edwards, 2016). For phylogeographic analyses, at the transition from population structure to phylogenetic divergence, incomplete lineage sorting (ILS) is prevalent where populations have been separated for less than $4N_e$ generations, where N_e is the effective population size (Neigel and Avise, 1986; Rosenberg, 2002). Another increasingly evident source of heterogeneity is introgression among species (Fontaine et al., 2015) (Fig. 9.2), the converse of the deep phylogeographic structure often observed in low-dispersal taxa. Such reticulation has long been recognized in plants, or in microbial systems, where horizontal gene transfer is an established paradigm. Increasingly, zoologists are also finding evidence for extensive movement of genes between phenotypically divergent taxa (Novick et al., 2009; Lamichhaney et al., 2015), including nonsister species. Such observations have increased attention to models of "speciation-with-gene-flow" (Feder et al., 2012). The new genome-scale analyses are causing evolutionary biologists to reevaluate the very nature of species (Mallet, 2008; Mallet et al., 2016), which, in some cases, appear to maintain phenotypic distinctiveness despite extensive gene flow across most of the genome (Mavárez et al., 2006; Heliconius Genome Consortium, 2012; Soria-Carrasco et al., 2014; Malinsky et al., 2015), and to recognize introgression as an important source of adaptive traits in a variety of study systems (Nosil and Crespi, 2004; Rheindt and Edwards, 2011; Nosil and Feder, 2012). Analytically, evidence of introgression among species is driving the emergence of network models of diversification (Nakhleh, 2013). Clearly genome-scale biology and the abundant reticulations across the "Tree of Life" are turning much of evolutionary biology, including phylogeography, on its head.

In this chapter, we explore the themes of reticulation and the genomics of speciation as key processes across the phylogeography–phylogenetics continuum. Reticulation presents challenges to many concepts and methods in phylogeography and speciation that were not as evident when we were locus-poor in the PCR era. Reticulation in the form of recombination causes gene trees to depart from a strictly bifurcating pattern, hence posing challenges for some methods of reconstructing evolutionary history (Lanier and Knowles, 2012). Recombination has also long been known to play a central role in speciation (Butlin, 2005), and the suppression of recombination, such as occurs in chromosomal inversions, can dramatically reduce the local genomic rate of gene flow (Joron et al., 2011; Nishikawa et al., 2015). Reticulation via gene flow is to be expected among intraspecific lineages that are the classical domain of phylogeography (Fig. 9.2). Finally, reticulation is becoming increasingly conspicuous at the level of diverging species and adaptive radiations (Rheindt and Edwards, 2011; Ellegren et al., 2012; Yeaman, 2013; Hermansen et al., 2014; Delmore et al., 2015;

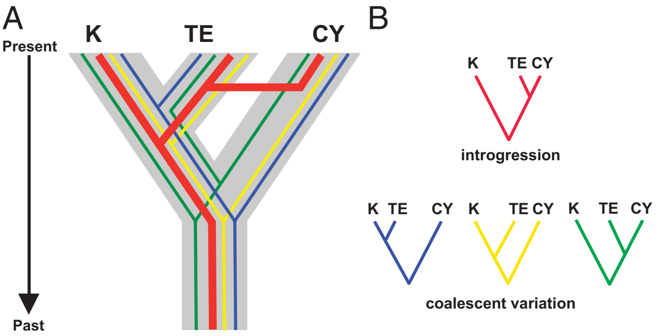


FIGURE 9.2 Sources of gene tree heterogeneity among diverging lineages. The three lineages KIM (K), TE, and CY are representative of northern Australia populations (Fig. 9.3). (*A*) These lineages are arranged by true evolutionary relationships depicted in gray: the "species" coalescent. (*B*) Within the species tree are gene trees colored in blue, yellow, red, and green. These gene trees represent independent coalescent histories and highlight various sources of gene tree discordance. Blue, green, and yellow gene trees highlight variation due to differing mutation rates and stochastic coalescent histories (including ILS), and red depicts effects of introgression among nonsister lineages. The MSC model allows for estimation of species trees, given mutational and coalescent variance. However, introgression, if extensive, can yield an incorrect species tree (e.g., [K, (TE,CY)] in this case) using most available methods. Additionally, distinguishing introgression from ILS in gene trees can be challenging and mostly relies on branch lengths in gene trees (as in IM models).

Lamichhaney et al., 2015), and is causing biologists to consider "ephemeral species," with frequent lineage mergers, not only among sister taxa but between lineages with any geographic co-occurrence (Rosenblum et al., 2012). The goal, however imperfect it is now realized, is to develop and apply models that integrate phylogeography, demography, and genome evolution in ways that will allow more nuanced interpretation of myriad interacting evolutionary processes from patterns of genomic diversity (Cutter, 2013). Acknowledging and modeling reticulation at various levels in the hierarchy of life will be an important part of reaching this goal.

KEY PROCESSES OF DIVERGENCE AND RETICULATION IN NATURE

Comparative Phylogeography Across the Australian Monsoonal Tropics

In essence, comparative phylogeography is about establishing commonalities of spatial patterns of genetic and gene tree diversity across codistributed species (Bermingham and Moritz, 1998; Avise, 2000). Combined with population genetic (coalescent) and spatial modeling (Knowles, 2009b), this effort has yielded insights into biogeographic history, such as locations of refugia and expansion areas (Hewitt, 2011) and the varying effects of ecological or physical dispersal barriers. In a comparative setting, such studies can identify how landscape features and regional climatic variation have interacted with the varying ecologies of species to shape current diversity (Carnaval et al., 2014) and how these interactions can influence speciation processes (Moritz et al., 2009; Ellegren et al., 2012).

To explore divergence vs. reticulation processes in a comparative context, we focus on phylogeographic data for ecologically diverse species from northern Australia, a vast stretch of monsoonal savanna and woodlands with interspersed ancient sandstone plateaus (Bowman et al., 2010) (Fig. 9.3). That this rich tropical fauna is biogeographically structured has long been known (Keast, 1961) and formalized using cladistic biogeography by Cracraft (1986), who recognized a basal dichotomy across the treeless "Carpentarian Barrier" (CB) separating the Kimberley (KIM) and Top End (TE) faunas from the faunas in Cape York (CY) and the eastern Australian forest (EF), as well as New Guinea. KIM and TE are, in turn, separated by hot, low-relief, and relatively dry regions associated with several smaller barriers (Eldridge et al., 2011), collectively referred to here as the Kimberley-Top End Barrier (KTEB). Early sequence-based phylogeographic studies in babblers [*Pomatostomus* (Edwards, 1993)] reported deep divergence across the CB relative to divergence within the eastern and western regions of the continent. Subsequent multilocus analyses of several avian systems revealed mostly Pleistocene divergences across the CB for congeners (Jennings and Edwards, 2005; Balakrishnan et al., 2010; Toon et al., 2010), as well as within species (Lee and Edwards, 2008; Kearns et al., 2011). Some studies examining divergence across the region have discovered clines (Rollins et al., 2012) or complex reticulate patterns in the form of introgression; for example, in butcherbirds (*Cracticus*), populations east of the KTEB are introgressed with mtDNA from populations of arid-adapted species to the south that expanded during the Last Glacial Maximum, whereas populations west of the KTEB are not introgressed (Kearns et al., 2014).

Fewer multilocus phylogeographic studies have been conducted for mammals across the monsoonal tropics, yet some common themes

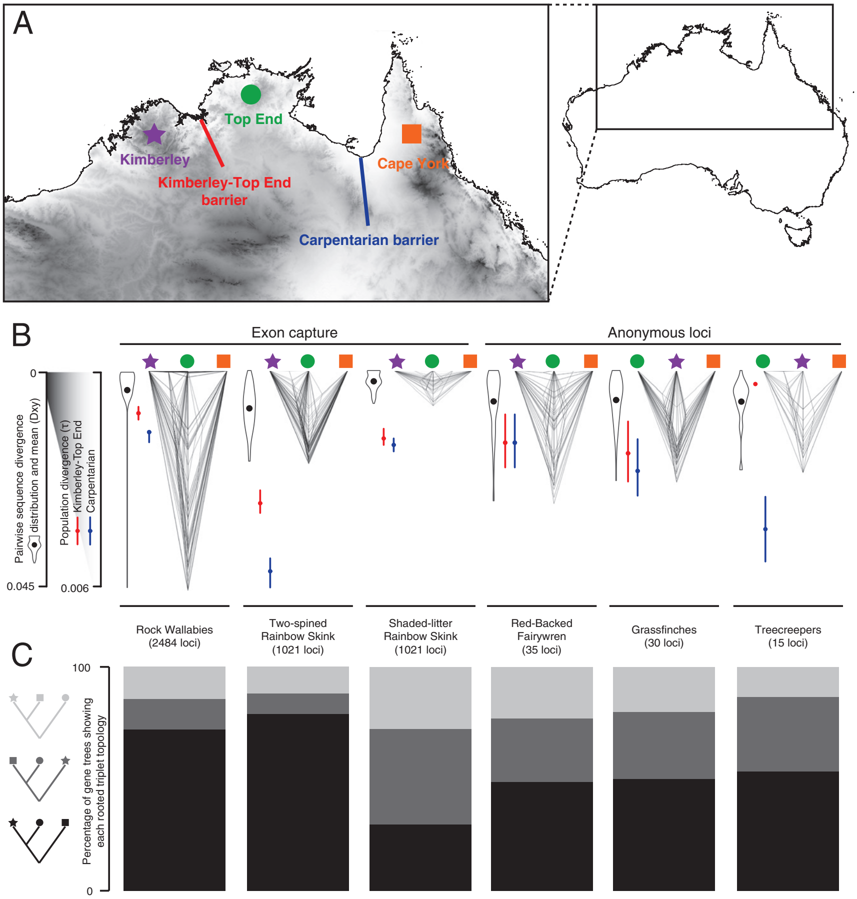


FIGURE 9.3 Gene tree heterogeneity in multilocus phylogeographic datasets of birds (red-backed fairywren, *Malurus melanocephalus*; *Poephila* grassfinches; *Climacteris* treecreepers), skinks (two-spined rainbow skink, *C. amax*; shaded-litter rainbow skink, *C. munda*), and mammals (*Petrogale* rock-wallabies) across northern Australia. (*A*) Map of northern Australia showing the KTEB and CB that separate the KIM, TE, and CY faunas. (*B*) Cloudograms illustrate topological and branch length variation of gene trees. Violin plots represent the distribution of pairwise sequence divergences across the CB, and black dots indicate mean pairwise sequence divergence, or D_{xy}. Red dots and lines are estimates and 95% confidence intervals of population divergence across the KTEB, whereas green dots and lines are estimates and 95% confidence intervals of population divergence across the CB. (*C*) Distribution of rooted triplets shows that gene trees exhibiting deeper divergence times across the CB than the KTEB are the most frequent in all taxa except the shaded-litter rainbow skink. Additional details are provided in *Supporting Information Text*.

emerge. Rock-wallabies (*Petrogale*), specialists of rocky habitats as the name implies, are strongly structured across the disjunct sandstone plateaus of the region, with deeper divergences across the CB than across the KTEB (Potter et al., 2012a,b) (Fig. 9.3). Other macropod species have different degrees of geographic and genetic discontinuity across the CB, suggesting the species' ecology has played a key role in their ability to adapt and persist across this region. The antilopine wallaroo (*Macropus antilopinus*), a savanna-woodland specialist, has a disjunct distribution but shallow divergence on either side of the CB grasslands, suggesting recent gene flow or range expansion. By comparison, a more ecologically generalized and widespread congener, the common wallaroo (*Macropus robustus*), has a more continuous distribution but substantial divergence across the CB (Eldridge et al., 2014). Preliminary analyses of other mammals suggest deep divergences across the CB, but require more expansive molecular study (Malekian et al., 2010; Aplin et al., 2015).

Low-dispersal species, such as lizards and frogs, have been the subject of a burst of recent multilocus phylogeographic studies across the region, including early applications of NGS in this context. Relative to birds and mammals, these taxa exhibit phylogeographic structure at a finer spatial scale, often with cryptic species and greater phylogenetic depth among regions, possibly reflecting a combination of lower dispersal and higher localized persistence through cycles of harsh climate. Deep structure across the CB, and often also the KTEB, is observed across phylogenetically and ecologically diverse reptiles, including species complexes of agamid lizards (Melville et al., 2011), rainbow skinks (Dolman and Hugall, 2008; Potter et al., 2016), several species complexes of geckos (Oliver et al., 2014; Moritz et al., 2016), and toadlet frogs (Catullo et al., 2014). In many cases, the divergence across the CB appears at deep phylogenetic scales rather than within species. For example, *Carlia* rainbow skinks have radiated across the KIM and TE, yet these taxa diverged from the eastern species of *Carlia* in the mid-Miocene (Dolman and Hugall, 2008). Analyses of ~2,000 exons for the two-spined rainbow skink (*Carlia amax*) inferred recent population expansion from western KIM across the KTEB to the western TE, emphasizing that the KTEB is a more porous filter than the CB (Potter et al., 2016). Studies of low-dispersal taxa (Catullo et al., 2014; Moritz et al., 2016; Potter et al., 2016) are also revealing congruent patterns at a finer scale than the major barriers envisioned by Cracraft (1986). These congruent patterns include deep structuring between offshore islands and mainland populations and an unexpected north-south split from the TE to the northern desert region (Catullo et al., 2014; Moritz et al., 2016; Potter et al., 2016). Closely related northern desert taxa often have ranges that are more widespread than those ranges across the savannas and sandstones to the north, and sometimes with evidence of broad-scale introgression

(Catullo et al., 2014; Moritz et al., 2016; Potter et al., 2016). For example, in contrast to strong phylogeographic structuring within *C. amax*, an arid-adapted congener, *Carlia munda* (shaded-litter rainbow skink) includes a single widespread clade from the west coast across the northern desert to the east coast (Fig. 9.3).

Gene Tree Heterogeneity Across the CB

The complex landscapes and dynamic climate history across this region have resulted in a combination of often strongly vicariant processes across the CB and a mix of divergence and dispersal or introgression across the KTEB. Given that gene tree heterogeneity arises from both ILS and gene flow between populations (Fig. 9.2), we can expect to see a more dominant phylogenetic signal across the CB in which the deepest split for a majority of gene trees spans the CB, with fewer loci having their deepest split across the KTEB, or between the CY/EF and TE (Fig. 9.3). We explored this hypothesis for exemplar avian, mammal, and lizard taxa for which we had multilocus sequence data spanning these geographic regions (Fig. 9.3 and *Tables S1* and *S2*[1]). As expected, among four-tip gene trees (one allele sampled for the KIM, TE, and CY/EF plus outgroup), we found diverse gene tree distributions across the region, with gene trees exhibiting deeper divergence times across the CB than the KTEB being the most frequent (Fig. 9.3 and Table S3). An exception to this pattern is *C. munda*, the more arid-adapted lizard, in which the dominant gene tree is one in which the TE and CY alleles are sisters, implying a more isolation-by-distance than vicariance model (Cracraft, 1986). Analyzing the larger datasets in which these simple gene trees are embedded with coalescent models (Rannala and Yang, 2003) uniformly suggests deeper population divergence and speciation across the CB than across the KTEB, although these divergences are quite close temporally in several cases (Fig. 9.3 and *Fig. S1*). Although our sample sizes are small, the analysis also suggests that the highest genetic diversity currently segregating within each complex varies among regions; in fairy wrens and wallabies, the highest diversity is in the CY/EF, whereas diversity is similar among regions in *C. amax*. Finally, the analysis does not support a key prediction of the simplest vicariance scenario: that the effective population sizes of descendant lineages are smaller than the sizes of the ancestral populations inhabiting the area before the vicariant event. Our estimates of ancestral N_e for at least four of the six species groups are smaller than for contemporary lineages

---

[1] Supporting information for this chapter, which includes additional text, *Figs. S1* and *S2*, and *Tables S1*, *S2*, *S3*, and *S4*, is available online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1601066113/-/DCSupplemental.

in the KIM, TE, or CY. A challenge with our brief analysis of comparative phylogeography across the monsoon tropics of Australia is the diversity of markers, which prevents easy comparison across groups due to differences in substitution rate. Use of a common set of markers, such as provided by various forms of target capture (Brumfield et al., 2003; Lemmon and Lemmon, 2012; Jones and Good, 2016), whether coding or noncoding, will be an important focus of future research.

Reticulation Driven by Ecology and Introgression

Our comparative phylogeographic analysis for northern Australia highlights the complex mix of divergence and reticulation and diverse spatial and temporal scales of phylogeographic structure that can emerge. Much of this heterogeneity appears to relate to differences among species in their capacity to persist or disperse across the landscape as climates oscillated over the Quaternary. Whereas it is convenient to focus on common patterns of divergence, in a classic vicariance mindset, closer attention to differing outcomes of reticulation, such as we have seen when comparing the results for *C. munda* with the results for other taxa spanning the CB, will yield more insight into speciation processes (Harrison and Larson, 2014).

Following secondary contact, genetically distinct populations can form "tension zones," maintained over time by a balance between dispersal and selection against hybrids (Barton and Hewitt, 1985), progressively merge via introgression [i.e., ephemeral taxa (Cutter, 2013)], or overlap while maintaining their integrity (Fig. 9.4). A special case of introgression occurs when an expanding lineage overrides a static (relictual) one, but is itself invaded by genes from the resident population due to sequential founder events during the spatial expansion (Ray and Excoffier, 2009). Over time, introgressed chromosome segments will recombine between lineages, leading to a mosaic of coalescent histories within and across loci. These reticulation events can manifest at two scales: in genetic clines, for single-nucleotide polymorphisms (SNPs) at the contact zone(s) themselves, and in lineage-scale migration, as estimated from allopatric populations using isolation-migration (IM) models (Fig. 9.4). Genome-scale data are enabling new approaches (reviewed in Payseur and Rieseberg, 2016), including genomic clines (Gompert and Buerkle, 2009) and analyses of lengths of introgressed haplotype blocks (Pool and Nielsen, 2009). Given estimates of recombination rate, the length of immigrant haplotypes can, in principle, be used to estimate the timing of recent introgression events at a lineage scale, a parameter that has proved difficult to infer from IM models (Sousa et al., 2011).

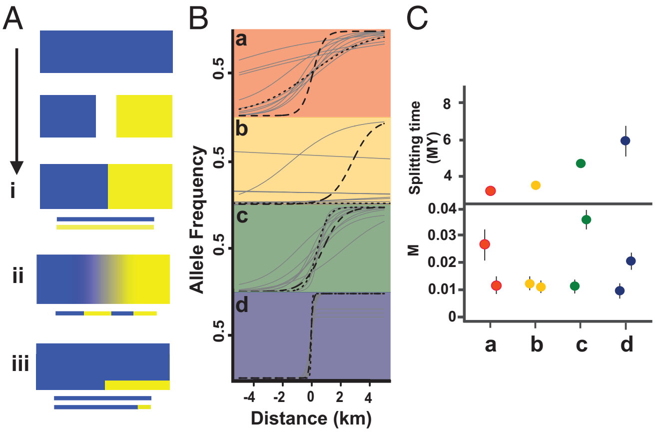


FIGURE 9.4 Contrasting processes and views of introgression. (*A*) Progression over time from population splitting, divergence in isolation, and secondary contact, with alternate outcomes: (*i*) tension zone, (*ii*) merging, and (*iii*) overriding of expanding population (blue) over the resident population with introgression from yellow → blue for some genes. (*B* and *C*) Contrasting perspectives on introgression among cryptic lineages of Australian Wet Tropics lizards at the local scale (*B*, contact zone) vs. lineage-scale estimates from IM analyses (modified from Singhal and Moritz, 2013). Note decreasing introgression at contact zones with increasing divergence time of lineage pairs, but no corresponding signal of decreasing migration at the lineage scale.

In the context of comparative phylogeography, insights into reticulation processes can be gleaned by comparing outcomes for taxa with varying ecologies and lineage divergence times across a common geographic and paleoenvironmental setting. Suture zones can be useful for this purpose, where multiple taxa have co-occurring contact zones (Dasmahapatra et al., 2010). The fauna endemic to the rainforest of northeastern Australia are a case in point. Climate-driven fluctuations of rainforest-dependent taxa on mountaintops have resulted in spatially concentrated contact zones between morphologically indistinguishable but genetically distinct lineages (Moritz et al., 2009). A comparative analysis of clines and genetic disequilibria across different contact zones (Singhal and Moritz, 2013) revealed less introgression and stronger genetic disequilibrium between more divergent lineage pairs, showing that reproductive isolation between these phenotypically cryptic lineages scales with divergence time (Fig. 9.4). However, at the lineage scale, levels of gene flow inferred from IM analyses of comparative transcriptomes are generally low and do not scale with divergence time (Fig. 9.4). These contrasting patterns remind us that estimates of gene flow are often averaged over the entire divergence history.

The Nexus of Comparative Phylogeography and Speciation Genomics

So how will a fully genomic perspective enrich our understanding of the nexus between comparative phylogeography and speciation? A plethora of recent whole-genome comparisons among sister taxa reveal fascinating, but complex, heterogeneity of divergence across the genome (reviewed in Payseur and Rieseberg, 2016). The most common outcome among recently diverged taxa is stronger differentiation on X and Z sex chromosomes than autosomes and scattered "islands" of high divergence against a background of low divergence. Islands of divergence were initially taken as suggesting locations of incompatible genes in the context of ongoing gene flow (Wu, 2001). However, it is also possible that they reflect varying levels of background selection in the absence of gene flow (Payseur and Nachman, 2005; Cruickshank and Hahn, 2014), leading to reinterpretation of some high-profile examples (Burri et al., 2015).

A key factor emerging from these studies and earlier scans of intraspecific diversity is the strong effect of recombination rate variation on the spatial patterning of genomic diversity, mediated most strongly by hitchhiking (Charlesworth et al., 1993). Thus, we expect to see reduced within-lineage diversity in regions of low recombination, with a corresponding increase in divergence using measures that are sensitive to levels of within-lineage diversity [e.g., Wright's fixation index F_{st} (Nachman and Payseur, 2012)]. Paradoxically, it has also been proposed that low-recombination regions, as might occur within chromosomal inversions or near centromeres, will accumulate locally adapted alleles, thereby contributing to genetic incompatibility between lineages (Kirkpatrick, 2010). Empirical evidence for this proposal is mixed, but there are some positive examples (Lemmon et al., 2012; Lemmon and Lemmon, 2013; Zhou et al., 2014). Finally, genome comparisons among closely related taxa have also highlighted introgression of adaptive alleles from one lineage to another (Rheindt and Edwards, 2011; Hedrick, 2013), an old concept reborn (Lewontin and Birch, 1966). Such alleles can readily flow across contact zones even if there is strong hybrid breakdown. Analytical challenges aside, such cases point to the exciting prospect of understanding how adaptive evolution influences divergence and reticulation among lineages.

One limitation of many of the above analyses of genomic divergence during speciation is that the historical biogeographic and environmental setting of isolation and reconnection of diverging lineages over time is often not well established (Payseur and Rieseberg, 2016). Understanding these processes is the core business of phylogeography, and closer interaction between analyses of historical biogeography and speciation genomics can be expected to bear fruit. Conversely, in the genomic era, comparative phylogeographers will not just have to master details of environmental history, species' ecology, and the plethora of methods for NGS and demographic inference but will also have to comprehend effects of selection and recombination rate variation across the genome. This challenge is exciting and will serve to strengthen further the link between population genomics and phylogenetics.

RECONSTRUCTING PROCESSES OF DIVERGENCE AND RETICULATION

Evolution of Molecular Markers in the Next-Generation Era

Having outlined some of the key processes of divergence and reticulation observable with genomic data, we now ask: How do we reconstruct phylogeographic history in the era of NGS? What new practical and analytical challenges do the increased detail afforded by NGS bring to phylogeographic reconstruction? The glimpse of comparative phylogeography across the CB in northern Australia makes clear the implications of one of the key components of any effort at reconstructing demographic history, namely, how we select molecular markers and the need for easy comparison across datasets. As a start, we may ask: Has NGS finally liberated phylogeographers from the constraints of marker choice, allowing unfettered access to the most appropriate markers for the questions being asked? Which combinations of markers may promote the further integration of phylogeography and phylogenetics? In these still-early days of next-generation phylogeography, marker choice is still constrained somewhat by technical and resource considerations, and will remain so until whole-genome sequencing of individuals or at least exemplars of the clades being studied becomes routine. The emergence of several widely used NGS platforms and marker suites in the past few years illustrates this point. For example, the flanking regions of ultraconserved elements (UCEs) have been promoted as suitable for phylogeographic questions, with the advantage that they are variable and their presence in can be predicted in uncharacterized genomes (McCormack et al., 2013). Another comparison (Brandley et al., 2015) found similar phylogeographic resolution between exons (drawn randomly from transcriptomes) vs. anchored hybrid enrichment (AHE) loci, which mostly target conserved exons (Lemmon and Lemmon, 2013; Lemmon et al., 2012). Exon capture has been effectively used to study diverging lineages of both vertebrates and invertebrates (e.g., Zhou et al., 2014; Hugall et al., 2016; Potter et al., 2016) and is particularly appropriate for retrieving genomic data from museum specimens (Rowe et al., 2011; Bi et al., 2013; Jones and Good, 2016).

Arguably, most UCE, AHE, or exon capture loci that have been used thus far for next-generation phylogeography are under mild or even strong purifying selection. Such selection is not necessarily a problem; after all, much of the animal mitochondrial genome, despite its high variability, is under purifying selection. However, purifying selection will likely reduce variation and bias the site-frequency spectrum toward low-frequency variants in a manner similar to, but less extreme than, selective sweeps, making gene trees compressed toward the tips (Nielsen, 2005). There is also clear evidence that loci in the vicinity of exons exhibit reduced levels of ILS compared with anonymous genomic regions (Scally et al., 2012). So long as researchers frame their findings within the context of the diversity of loci found throughout the genome, exons and UCEs are likely to remain a powerful force in phylogeography. The pervasiveness of natural selection, particularly for species with large effective population sizes (Corbett-Detig et al., 2015), is, however, a force with which phylogeographers have not yet fully come to grips. One wonders whether any of the loci used in phylogeography in the next-generation era are genuinely neutral.

The approach using restriction-site-associated DNA sequences, or RAD-seq, is a popular application of NGS to phylogeography, and yields large but sometimes patchy matrices of relatively short and mostly noncoding loci (Andrews et al., 2014a), which are often analyzed in the form of SNPs. Such markers can be powerful measures of phylogeographic structure and, in some cases, seem relatively free of strong selection (Dierickx et al., 2015). Within-locus recombination is irrelevant to SNPs, whereas recombination may pose challenges for analysis of the longer loci such as are generated by target capture and AHE. RAD-seq loci are less amenable to the type of gene tree building that has characterized phylogeography (Hare, 2001; Brumfield et al., 2003; Knowles, 2009a), but few NGS loci of any kind yield highly resolved gene trees when applied on a phylogeographic scale. The power of next-generation methods lies primarily in generating more independent loci, although the phylogenetic informativeness of individual loci also plays a role, especially in species tree reconstruction (Liu L et al., 2015). Going forward, it will be important to compare the behavior and informativeness of different types of markers and genomic compartments explicitly in phylogeographic settings (Fontaine et al., 2015; Nater et al., 2015).

The ideal phylogeographic marker in the next-generation era presumably depends on the questions being asked and the temporal and taxonomic scales over which comparisons are made. Whereas introns and anonymous loci were popular sequence-based markers in the PCR era (Hare et al., 1996), and continue to be captured by various NGS approaches (Lemmon and Lemmon, 2012; Barrow et al., 2015; Perez et al., 2016), targeting of such unconstrained sequence-based markers has made few inroads in the next-generation era, presumably because compared with exons (Bragg et al., 2016), such loci are difficult to predict, and therefore capture, in unknown genomes using probes from other species. Herein lies a conflict between the ease of retrieving markers and their variability within species: Until whole-genome sequencing of phylogeographic exemplars allows us to design probes that are optimal for a given species, and consistent across taxa, the practicalities of easily and cheaply capturing large numbers of loci may tend the field toward conserved loci. Ultimately, phylogeographers should embrace a diversity of marker types even within individual studies, not only to allow the phylogeographic history of different marker types to illuminate each other but also to study genomic diversity and history in an unbiased way that facilitates the discovery of genomic loci underlying adaptation.

Insight into Processes of Reticulation from Gene Tree Outliers

Gene tree outliers, like F_{st} outliers, may be important indicators of nonneutral or locus-specific processes in the genome. We used a newly proposed gene tree outlier approach, KDEtrees (Weyenberg et al., 2014), to explore the behavior of gene tree distributions in empirical phylogeographic and low-level phylogenetic datasets of several marker types (*Fig. S2* and *Table S4*). KDEtrees appears effective at identifying loci that result from horizontal gene transfer or are clear outliers, such as gene trees generated by a species tree different from the majority. However, it is unclear how KDEtrees behaves when confronted with loci influenced largely by demographic processes, or how the number of outliers varies by marker type. Although our sample size is small, our analysis of eight datasets (*Fig. S2* and *Table S2*) suggests that many phylogeographic and transcriptome datasets harbor surprisingly few gene tree outliers, and so conform well to expected distribution based on overall patterns and levels of divergence. For example, given our chosen level of sensitivity ($\lambda = 1.5$), for which we expect roughly 5% of gene trees to exhibit outlier behavior simply by chance, none of the datasets we analyzed contains a significant number of outliers. In the future, the KDEtrees approach, and other methods (Gruenstaeudl et al., 2015), should be a useful tool to explore gene tree heterogeneity within and between datasets.

METHODS FOR DETECTING RETICULATION: RECOMBINATION AND INTROGRESSION

Reticulation and Phylogenetic Networks

As we have seen, as phylogeography and speciation studies begin to probe the genomes of diverse species on a large scale, reticulation, in the forms of introgression and recombination, appears much more common than previously supposed. Accordingly, a major challenge going forward is to incorporate reticulation as a standard component of phylogeographic analysis. Many computational methods targeted at the phylogeography–phylogenetics continuum necessarily ignore some kinds of reticulation. Key examples include models to estimate species trees from multiple unlinked loci using the multispecies coalescent (MSC) model (Edwards, 2016). MSC methods ignore two fundamental aspects of reticulation: recombination within loci and postspeciation hybridization. Some MSC methods (Liu et al., 2009b) are known via simulations and theoretical arguments to be robust to reticulations, such as introgression, particularly when datasets are large and when introgression is confined to a subset of loci. However, other MSC methods are not robust to such model violations (Edwards, 2016). It is not surprising that phylogeographers are among the most comfortable working with MSC methods because of the similarity in assumptions they apply to multilocus datasets. At the same time, due to their familiarity with reticulating lineages, phylogeographers are the most likely to identify shortcomings arising from the inherent simplifications of standard MSC models.

Conceptually, networks subsume trees; networks are trees with reticulation (figure 1 of Edwards et al., 2016). Genome-scale evidence for introgression is renewing enthusiasm for coalescent phylogenetic models that allow for hybridization between diverging lineages. Several recent phylogenomic datasets, including those datasets analyzing human populations as well as distantly related lineages of birds or mammals (Hallström and Janke, 2010; Jarvis et al., 2014), have noted signals for reticulation in the form of ancient interlineage hybridization. Phylogenomic network models based on the MultiSpecies Network Coalescent (MSNC) (Park and Nakhleh, 2012; Nakhleh, 2013; Stenz et al., 2015) are likely to be an important new tool for phylogeneticists in general and phylogeographers in particular. Early studies suggest that application of the MSNC to genome sequences from diverging species will yield new insights into complex evolutionary histories of divergence and reticulation (Wen et al., 2016).

Better insight into the presence of reticulation at the level of populations need not involve computationally intensive algorithms. For example, application of simpler SNP-based tests of introgression and admixture [e.g., the "ABBA-BABA" test (Durand et al., 2011)] will help flag phylogeographic scenarios that may be more complex than originally envisioned. Although they have yet to make inroads into the phylogeography of nonmodel species, a suite of recently developed drift (*F*) statistics, related to but distinct from Wright's *F*-statistics, provide simple and powerful metrics to test various models of population history, such as whether populations are related in a treelike fashion [Reich et al., 2009; Patterson et al., 2012; reviewed by Peter (2016)]. Tools for model selection in phylogeography (Hickerson et al., 2010; Ray et al., 2010; Carstens et al., 2013; Excoffier et al., 2013; Tsai and Carstens, 2013; Chan et al., 2014; Xue and Hickerson, 2015) will also be critical for determining whether reticulation at the population level is an important part of the demographic history under study.

Capturing Heterogeneity with the Sequentially Markovian Coalescent

Recombination within loci violates assumptions of most phylogenomic analyses, whether informed by the MSC or not. The departure from the assumptions of the MSC could be particularly acute for datasets consisting of sequences from long loci relative to the distance over which linkage disequilibrium decays, which can be <1 kb in many organisms. The one simulation study exploring effects of intralocus recombination (without introgression) on the performance of species tree methods (Lanier and Knowles, 2012) found little effect, and then only on very short trees, as is typical of phylogeographic datasets. However, Potter et al. (2016) observed incongruent and less resolved species trees among lineages of *C. amax* (as in Fig. 9.3) when using full-length exons compared with the longest nonrecombining segments of these loci. Still, phylogeography is no stranger to intralocus recombination. Several phylogeographic models have been adapted to incorporate recombination (Becquet and Przeworski, 2007; Naduvilezhath et al., 2011), using information from the joint site frequency spectrum among loci and other data. If justified, such models could be adapted to MSC and MSNC methods to allow for intralocus recombination. Additionally, several postgenome phylogeographic models have emerged that incorporate recombination via the sequentially Markovian coalescent, a new approach that models the coalescent site-by-site along the genome, exploiting the variation in site patterns among linked SNPs (Marjoram and Wall, 2006; Li and Durbin, 2011; Wang et al., 2014). Sequentially Markovian coalescent models have obvious applications in traditional species tree methods and may alleviate lingering concerns about recombination. A final means of addressing the issue of intralocus recombination in phylogeography is by using SNP data, which obviates intralocus recombination. Phylogeographic models using SNP data have been available for a number of years (e.g., Beerli and Felsenstein, 1999; Beerli and Palczewski, 2010), and several MSC methods (Bryant et al., 2012; Chifman and Kubatko, 2014) now use linked or unlinked SNP data to estimate phylogenetic trees without explicitly estimating constituent gene trees. It remains to be seen whether the limited genealogical information in SNPs is compensated for by the large number of SNPs that can be collected in typical phylogenomic datasets.

CONCLUSION

Phylogeography has come a long way from its origins of analyzing single gene trees across geography (Avise et al., 1987). Sophisticated statistical inference, integration with spatial modeling, model choice, parameter estimation, and now access to sequence or SNP data for thousands of loci have all enriched the field tremendously. It will be interesting to see how closely future phylogeographers adhere to its conceptual roots, the "mitochondrial DNA bridge," as mirrored in gene trees empirically derived from nuclear sequence data. On the one hand, extensive reticulation in the form of recombination and the convenience of analyzing large numbers of unlinked SNPs with rapid parametric tests may be ushering in an era of phylogeography beyond gene trees, or, at the very least, an era that acknowledges them only implicitly, via connections with coalescent theory (Degnan and Rosenberg, 2009; McVean, 2009). On the other hand, some of the currently popular methods of locus capture are showing promise for capturing genetic diversity in the form of gene trees, even if weakly resolved at lower taxonomic levels. What seems clear is that next-generation approaches are pushing phylogeography toward a future dominated by SNPs or sequence data for thousands of loci, a positive development that we believe will help bridge the phylogeography–phylogenetics continuum as envisaged by Avise et al. (1987).

In the future, we can expect integration of phylogeography with increased understanding of genome organization and parameters, such as variation in recombination rate across the genome (Cutter, 2013). We can also expect continuing integration of phylogeography with speciation biology and with analyses of adaptive variation, including phenotype-genotype associations (Edwards et al., 2015). Almost by definition, phylogeography will retain its distinctions from sister disciplines like population genetics by its emphasis on broad geographic sampling and the natural history origin of the questions it seeks to answer. One of the thrills of phylogeographic research is the ability of researchers to absorb cutting-edge technologies that now put whole-genome variation within our grasp, yet also retain the exploratory field spirit that has motivated the discipline since its inception. With such a detailed view of genomic variation across geography, reticulation is likely to be omnipresent, pushing phylogeography to reinvent itself, question its foundations, and strive for new syntheses.

ACKNOWLEDGMENTS

We thank Mark Eldridge (Australian Museum Research Institute) for access to tissues and data of rock-wallabies; Ke Bi and Sonal Singhal for assistance with *Carlia* data collection and preparation of Fig. 9.4; Brant Faircloth, John McCormack, and Robb Brumfield for assistance in acquisition of online datasets; and John Wakeley, Rudy Yoshida, and Alan Lemmon for helpful discussion. S.V.E.'s research is supported by the U.S. National Science Foundation and Harvard University. S.P., J.G.B., and C.M. are supported by grants from the Australian Research Council.

Part III

COMPARATIVE PHYLOGEOGRAPHY IN A TAXONOMIC SENSE

Because nucleic acids are the universal hereditary material of life on Earth, molecular approaches to phylogeography can in principle be applied to creatures ranging from bacteria to whales and humans. The chapters in this section will further illustrate the catholic nature of comparative phylogeography by describing a few chosen examples spanning a wide array of disparate taxonomic groups.

Although microbes have not been a traditional focus of phylogeographic research (mostly for logistical or technical reasons), Michaeline Nelson and colleagues (Chapter 10) show how a trait-based approach can enlighten our understanding of microbial biogeography on a global scale. The genetic traits they surveyed involve different pathways for nitrogen cycling in soil microbes sampled from around the world. By analyzing the spatial and taxonomic distributions of such microbial traits, the authors develop an argument for how comparative phylogeographic methods could be highly relevant to predicting the responses of biogeochemical cycles to future climatic and other environmental changes.

Kelly Zamudio and colleagues (Chapter 11) make a similar case for vertebrate animals by emphasizing the need to integrate trait (phenotypic) data with molecular genetic data in future phylogeographic appraisals. The authors argue that because phenotypes are the actual targets of selection, they should be assessed in conjunction with other phylogeographic data to illuminate the ecological and evolutionary forces shaping biodiversity. Ever since the time of Alfred Russel Wallace (the founder of biogeography in the mid-1800s), biogeographic appraisals have relied primarily on the spatial distributions of phenotypic attributes. Thus, if the integrative prescriptions of Zamudio et al. are to be followed, the field of comparative phylogeography may in effect partially return to its intellectual biogeographic roots even while simultaneously expanding the latter in new genealogical directions. Bowen and colleagues (Chapter 1) and Papadopoulou and Knowles (Chapter 8) likewise endorsed the links between biogeography defined with organismal traits and phylogeography founded on gene genealogies.

In Chapter 12, Anne Yoder and colleagues focus on comparative genetic patterns in one particular taxonomic assemblage: congeneric mouse lemurs on the island of Madagascar. Apart from demonstrating the profound phylogeographic ramifications of altered landscapes across recent geological and human timescales, this survey is instructive because it exemplifies the types of comparative phylogeographic analyses that have been applied to many other animal taxa in various regions around the globe, and because it involves a single large island (as opposed for example to an oceanic, continental, or archipelago setting).

In a phylogeographic study of North American bison, Peter Heintzman and colleagues (Chapter 13) advocate for a greater emphasis on paleoecology as a major driving force shaping contemporary phylogeographic patterns. Their example involves the purported role of an "Ice-Free Corridor" as a plausible route for dispersal (of humans and other animals) between Beringia and more southerly areas of North America. Without knowledge of such paleoecological conditions in times past, the authors argue that phylogeographers will remain handicapped in their attempts to interpret the spatial genetic architecture of extant taxa.

Whereas mitochondrial (mt) DNA has been the traditional molecular workhorse of phylogeographic appraisals in animals, chloroplast (cp) DNA has performed the corresponding role in plant studies. In Chapter 14, Victoria Sork and coauthors summarize evolutionary lessons learned from comparative analyses of phylogeographic patterns displayed by diverse plant taxa in California. Most plants have several biological features that differ from those of most (not all) animals and that presumably have major phylogeographic and demographic relevance: a sessile growth form; high reproductive output; leptokurtic dispersal through pollen and seeds; intimate interactions with local environmental conditions; and the potential for individual longevity. Sork et al. provide empirical examples of how these botanical features apparently have impacted phylogeographic patterns in "plants" as diverse as oak trees and lichens.

Unsurprisingly, perhaps no single species has received greater phylogeographic attention than *Homo sapiens*. In the final chapter of this section of the book, Alexander Harcourt (Chapter 15) provides a panoramic overview of the phylogeography of our species, beginning with origins on the African continent some 200,000 years ago and continuing with the diaspora(s) of our ancestors from that landmass to eventually occupy the entire world. Included in this chapter are the deduced evolutionary dates and specific migration routes employed by our ancestors to "people the planet," as well as discussions of the patterns of cultural diversity associated with this phylogeographic history. Harcourt's concise summary of a vast literature on human origins should be of considerable interest to anthropologists and general audiences alike.

10

Global Biogeography of Microbial Nitrogen-Cycling Traits in Soil

MICHAELINE B. NELSON,* ADAM C. MARTINY,*† AND
JENNIFER B. H. MARTINY*‡

Microorganisms drive much of the Earth's nitrogen (N) cycle, but we still lack a global overview of the abundance and composition of the microorganisms carrying out soil N processes. To address this gap, we characterized the biogeography of microbial N traits, defined as eight N-cycling pathways, using publicly available soil metagenomes. The relative frequency of N pathways varied consistently across soils, such that the frequencies of the individual N pathways were positively correlated across the soil samples. Habitat type, soil carbon, and soil N largely explained the total N pathway frequency in a sample. In contrast, we could not identify major drivers of the taxonomic composition of the N functional groups. Further, the dominant genera encoding a pathway were generally similar among habitat types. The soil samples also revealed an unexpectedly high frequency of bacteria carrying the pathways required for dissimilatory nitrate reduction to ammonium, a little-studied N process in soil. Finally, phylogenetic analysis showed that some microbial groups seem to be N-cycling specialists or generalists. For instance, taxa within the Deltaproteobacteria encoded all eight N pathways, whereas those within the Cyanobacteria primarily encoded three pathways. Overall, this trait-based approach provides a baseline for investigating the relationship between microbial diversity and N cycling across global soils.

---

*Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697; and †Department of Earth System Science, University of California, Irvine, CA 92697.
‡To whom correspondence should be addressed. Email: jmartiny@uci.edu.

A grand challenge for this century is to predict how environmental change will alter global biogeochemical cycles. The field of biogeography has an important role to play in this effort (Violle et al., 2014). Environmental change is altering the distribution of biodiversity, which in turn is a key driver of biogeochemical processes (Naeem and Wright, 2003; Cardinale et al., 2012). Historically, biogeography has viewed biodiversity through a taxonomic lens, primarily resolving species distributions. However, a focus on traits—particularly those involved in ecosystem processes—may offer a clearer link between biodiversity patterns and biogeochemistry (Diaz and Cabido, 2001; McGill et al., 2006; Reichstein et al., 2014).

These ideas are particularly relevant for microorganisms. Microbes catalyze most of the biological transformations of the major elements of life (Falkowski et al., 2008), and because of their sheer abundance they account for a large pool of elements in living matter (Whitman et al., 1998). Furthermore, like plants and animals, microbial taxonomic composition varies over space (Martiny et al., 2006; Fierer et al., 2012), and this variation can influence ecosystem processes (van der Heijden et al., 2008; Strickland et al., 2009; Schimel and Schaeffer, 2012; Reed and Martiny, 2013). Thus, a consideration of microbial traits should improve efforts to connect biogeographic patterns and ecosystem processes (Green JL et al., 2008).

Here, we provide a first characterization of the global biogeographic patterns of microbial nitrogen (N) cycling traits in soil. Microbially driven transformations regulate biologically available N through exchange with the atmosphere (via N fixation and denitrification) and loss by nitrate leaching. They also influence the forms of N available for plant uptake. At the same time, human activities have altered, and continue to alter, the N cycle by increasing the amount of reactive N in the biosphere (Vitousek et al., 1997; Fowler et al., 2013). At local scales, N addition consistently shifts microbial composition in soils and other ecosystems (Allison and Martiny, 2008; Ramirez et al., 2012). The distribution of microbial traits might therefore be relevant for understanding current and future N cycling.

The taxonomic composition of soil microorganisms is correlated with spatial variation in climate, plant diversity, pH, disturbance, and many other factors (Lauber et al., 2009; Philippot et al., 2009; Tedersoo et al., 2014; Prober et al., 2015). These biogeographic patterns help to identify factors that select on the entire suite of microbial traits. In this study, we reverse this direction of inquiry. We first characterize the patterns and drivers of just a handful of traits associated with N cycling and then ask which taxa comprise these functional groups.

To quantify the abundance and composition of N-cycling traits, we analyzed ~2.4 billion short-read sequences from 365 soil metagenomes sampled from around the globe. From this dataset, we identified sequences that indicate the potential for a microorganism to perform one of eight N pathways that convert inorganic N to other inorganic forms or microbial biomass. We then quantified the frequency and taxonomic association of microorganisms carrying these pathways in each sample. If a gene from a pathway was detected, we assumed the presence of the entire pathway in the organism. To compare the frequencies among the N pathways, we standardized for the number of genes (2–20) in each pathway. Although metagenomic sequences provide a measure of a community's trait diversity (Barberán et al., 2012), the presence of a trait does not indicate how it is being used in the community. Thus, we cannot determine whether genes in the N pathways are expressed or the rate at which N is being transformed. However, assaying traits based on metagenomic sequences are parallel to other trait metrics used to describe an organism's functional potential, such as nutrient uptake affinity or temperature optimum for growth.

The global N trait dataset allowed us to address four main questions. First, what are the overall frequencies of the different N pathways in soil? We expected the frequencies to vary greatly by pathway. Indeed, the ability to perform nitrification is restricted to few microbial taxa, whereas ammonia assimilation is probably present in almost all taxa. Second, what drives variation in the frequencies of N pathways among soil samples? We hypothesized that N pathway frequencies would vary primarily by habitat type, which reflects major differences in plant communities and therefore N inputs into soils. Third, what are the main taxa encoding each N pathway? Surprisingly little is known about the dominant lineages encoding N-cycling traits across global soils. We therefore expected to find previously unrecognized, prominent players, particularly for the less-studied pathways such as dissimilatory nitrate to ammonium (DNRA). Finally, what underlies compositional variation among soil samples in microorganisms encoding N pathways? We hypothesized that the taxa responsible for each pathway would vary greatly by habitat type, because the habitat would select for specialized taxa. We further predicted that soil pH—previously identified as an important driver of soil composition (Rousk et al., 2010; Tsiknia et al., 2015)—would also influence compositional variation within microorganisms encoding N-cycling traits.

RESULTS

Metagenomic data from surface soil samples were retrieved from the metagenomics analysis server (MG-RAST) (Meyer et al., 2008). After curating the samples for sequence and metadata quality, the final 365 samples represented 118 unique locations from 10 distinct habitat types covering natural and human-dominated systems (Fig. 10.1 and

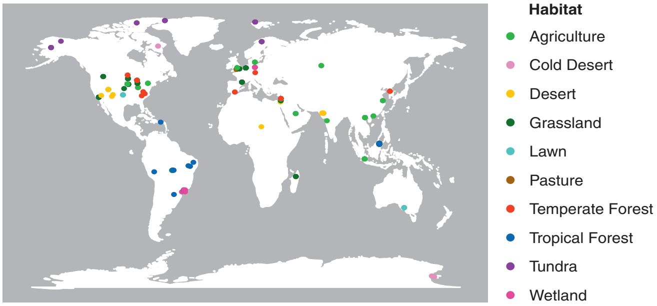


FIGURE 10.1 The locations ($n = 118$) sampled to create the soil metagenomic libraries ($n = 365$) used in this analysis. The samples represent 10 distinct habitats including agriculture ($n = 19$), cold desert ($n = 6$), desert ($n = 15$), grassland ($n = 14$), lawn ($n = 4$), pasture ($n = 2$), temperate forest ($n = 12$), tropical forest ($n = 34$), tundra ($n = 7$), and wetland ($n = 5$).

Dataset S1[1]). Sequencing depth varied greatly among the samples but was not overtly biased toward any particular habitat type (*Fig. S1*). To standardize for sequencing depth, we report the abundance of each N pathway as its frequency in a sample. The trends observed were similar whether pathway frequency was normalized as the number detected per annotated sequence or per marker gene (based on 30 conserved, single-copy genes) (*Fig. S1*).

Bacteria dominated the metagenomic libraries, comprising 95% of all sequences, followed by 3% for Fungi and only 2% for Archaea. The fraction of fungal sequences in metagenomic libraries is known to be lower than their contribution to soil microbial biomass (Fierer et al., 2012). We therefore concentrate our analyses on Bacteria and Archaea and report only general trends for Fungi. For instance, the proportion of total sequences of Bacteria, Archaea, and Fungi varied across habitat type (*G*-test of independence; $P << 0.001$) (*Fig. S2*). Archaea ranged from 0.9 to 11% of all sequences by habitat, with the highest percentage detected in deserts. The ratio of fungal to bacterial sequences was particularly high in temperate forest soil, as previously observed (Fierer et al., 2009).

---

[1] Supporting information for this chapter, which includes *Dataset S1*, *Figs. S1 through S6*, and *Table S1*, is available online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1601070113/-/DCSupplemental.

Frequency of Soil N Pathways

On average, 0.5% of all annotated sequences in a soil sample were associated with one of the eight N pathways (Fig. 10.2*A*), or an average of 3.3 and 4.7 N pathways per marker gene for Bacteria and Archaea, respectively. The frequency of the individual pathways varied by several orders of magnitude (one-way ANOVA $P < 0.001$; $F = 74.21$, df = 7) (Fig. 10.2*B*). Bacteria and Archaea displayed similar trends in their relative frequency of N pathways except for the absence of the dissimilatory nitrite reduction to ammonium pathway in Archaea. Fungal sequences were only associated with assimilatory pathways, including ammonia assimilation, assimilatory nitrate to nitrite, and assimilatory nitrite to ammonium.

Across all domains, the most common pathway was ammonia assimilation (Fig. 10.2*B*). For instance, among the Bacteria, an average of 280 ammonia assimilation pathways was detected for every million annotated bacterial sequences. In comparison, nitrification and N fixation were the least common pathways and detected only 6.1 and 4.6 times per million sequences, respectively. Notably, the relatively unstudied dissimilatory nitrite reduction to ammonium pathway was slightly more common than these two pathways, detected on average 9.3 times per million sequences.

Across all soil samples, N pathway frequencies were overwhelmingly positively correlated for both the Bacteria and Archaea (Fig. 10.3*A* and *B*). To examine differences in pathways beyond the trends shared by all, we calculated the residuals of the frequency of each pathway regressed against the frequency of all N pathways in a sample. This residual variation was also significantly correlated among many of the N pathways (Fig. 10.3*C* and *D*). For instance, denitrification was highly positively correlated with dissimilatory nitrate reduction to nitrite within both Bacteria and Archaea ($R^2 = 0.86$ and 0.97, respectively; $P \leq 0.001$). This relationship is expected, because dissimilatory nitrate reduction to nitrite is the first step of the complete denitrification process; however, we separated the two steps here, because nitrate reduction to nitrate is also the first step in DNRA (Yoon et al., 2015). Similarly, we separated DNRA into its two pathways: dissimilatory nitrate reduction to nitrite and dissimilatory nitrite reduction to ammonium (Fig. 10.2*A*). Among Bacteria, the assimilatory nitrite to ammonium pathway residual was negatively correlated with all other pathways. Likewise, the residual frequency of the ammonia assimilation pathway was negatively correlated with all other N pathways in both Bacteria and Archaea. N fixation generally showed weak or no correlation with other pathways.

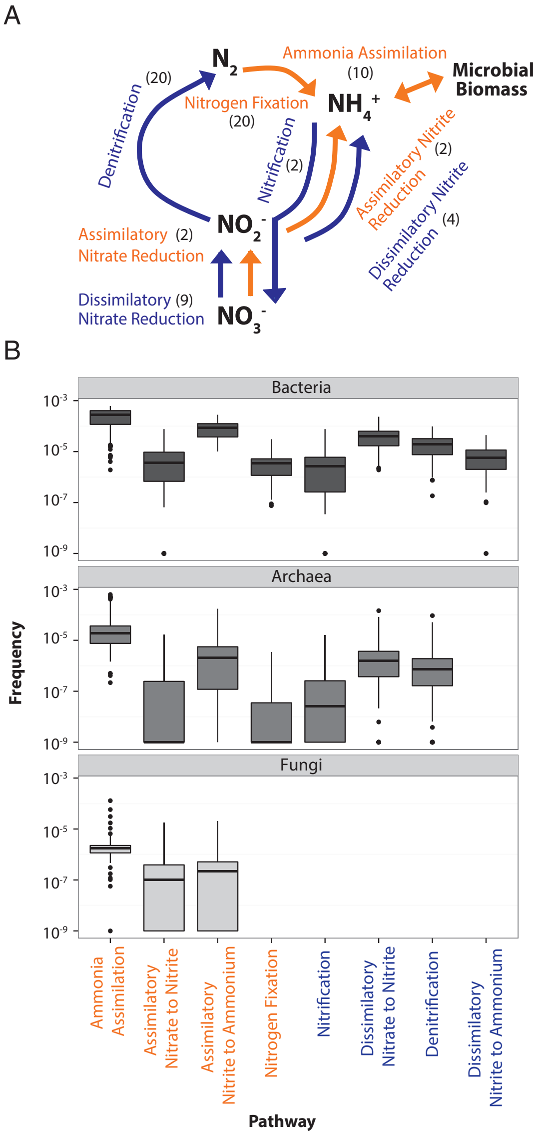


FIGURE 10.2 N pathways and their frequencies. (*A*) N pathways considered in this study. The numbers in parentheses are the number of genes targeted for each pathway. Assimilatory pathways are in orange and dissimilatory pathways in blue. (*B*) Box plot of the frequency of each N pathway in a metagenomic library for Bacteria, Archaea, and Fungi. To compare across domains, frequencies are calculated as per annotated sequence in each domain. The upper and lower bounds of boxes correspond to the 25th and 75th percentiles, with a median line shown. Whiskers represent 1.5*IQR (interquartile range). Dots represent outliers.

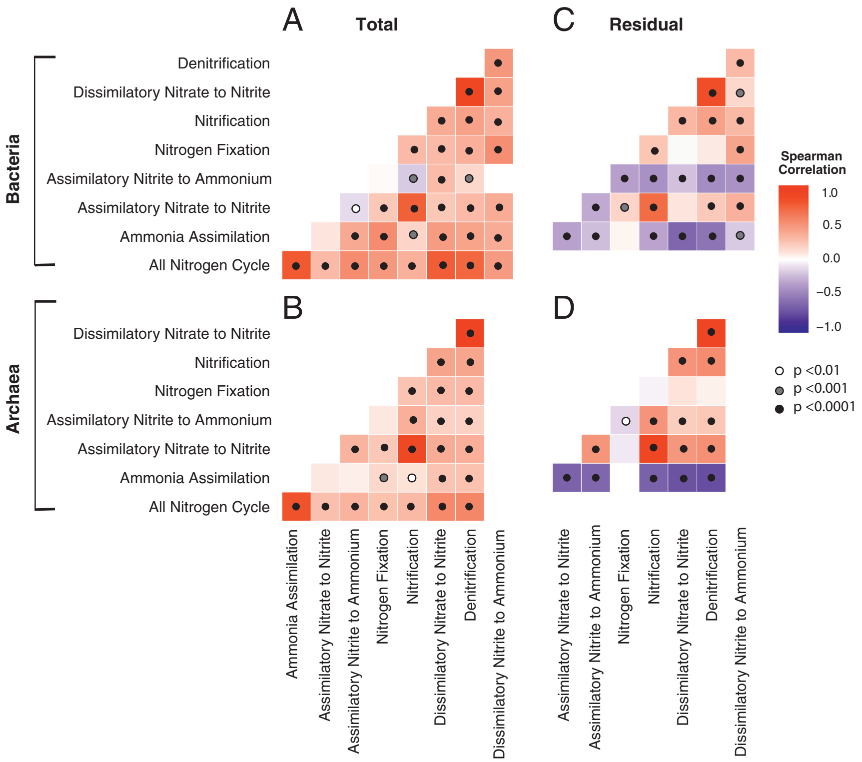


FIGURE 10.3 The relationships between N pathway frequencies. Correlations between N pathways encoded by Bacteria (*A*) and Archaea (*B*) across the samples. (*C* and *D*) Correlations between the residuals of each pathway regressed against the total frequency of all N pathways.

Drivers of N Pathway Frequencies

The frequency of all N-cycling traits (summing across all pathways) varied greatly among soil samples, and initial analyses revealed broad biogeographic patterns. On average, the highest frequencies of total N pathways were detected in tropical forest and human-dominated (pasture, lawn, and agriculture) soils, whereas the lowest frequency was observed in cold deserts (*Fig. S3*). Total N pathway frequency also tended to decrease with increasing latitude ($R^2 = 0.22$, $P < 0.05$; *Fig. S4*).

To disentangle the drivers behind these patterns, we performed a multivariate regression analysis including habitat type and environmental parameters known to influence microbial abundance and composition (Fierer and Jackson, 2006; Bru et al., 2011). Local measurements were not available for most samples; instead, we estimated these variables from secondary sources. For Bacteria, the regression model explained a large and significant proportion of the variability in the frequency of total N pathways ($R^2 = 0.58$, $P < 0.001$; Table 10.1). Habitat type contributed most to this model, both directly (positively related to total N pathways) and through interactions with soil carbon and N. The regression model for Archaea explained less variability in total N pathway frequency than for Bacteria ($R^2 = 0.43$, $P < 0.001$; Table 10.1). An interactive effect between carbon and N contributed the most to the model, and habitat was only important through an interactive effect with temperature.

We next examined the drivers of individual N pathway frequencies. Due to high covariance between pathways (Fig. 10.3*A* and *B*), we fitted regression models to the total-frequency-corrected residuals for each pathway. These models varied greatly in their ability to explain this additional variation (Table 10.1). For example, the models for the N fixation pathway explained 80% and 63% of the variation among samples in Bacteria and Archaea, respectively ($P < 0.001$). In contrast, the same parameters did not explain any variation in the frequency of the dissimilatory nitrite reduction to ammonium pathway in Bacteria.

Among the significant models, habitat type was an important predictor of the individual pathway frequencies (Table 10.1). Habitat also interacted with other factors including precipitation, temperature, and soil N to influence the frequency of some pathways. For instance, denitrification frequency increased with temperature in deserts but decreased with temperature in tropical forests. Similarly, ammonia assimilation frequency increased with soil N in temperate forests but decreased with soil N in tropical forests. Soil carbon, which seemed to be a primary driver of total N pathway frequency, did not explain differences in the frequency of individual pathways in Bacteria. Including estimates of N deposition in these models only improved the denitrification model (R^2 increased from 0.41 to 0.48); denitrification frequency increased with increasing N deposition.

The models for individual pathway frequencies in Archaea generally explained less variation than those for Bacteria, perhaps due to the lower number of sequences per sample (*Dataset S1*). However, for the significant models, the individual N pathways were often best explained by the same parameters as the Bacteria. For instance, habitat type and habitat by temperature were the most important predictors of N fixation frequency within both domains. Likewise, habitat, habitat by precipitation, and habitat by temperature contributed to the variation in assimilatory nitrate to nitrite frequency in both Archaea and Bacteria.

Taxonomic and Phylogenetic Distribution of N Pathways

A diverse range of microorganisms, encompassing 402 bacterial and 53 archaeal genera, encoded the N pathways. We first investigated the association of pathways within the same genera (Fig. 10.4 and *Fig. S5*). All genera for which we detected over 10 sequences carried the ammonia assimilation pathway. Genera carrying the pathway to complete the second half of denitrification also generally carried the first half of the pathway, dissimilatory nitrate to nitrite reduction. The same genera carrying these denitrification pathways sometimes, but not always, carried the dissimilatory nitrite reduction to ammonium pathway, or the second part of the complete DNRA process (Fig. 10.4 and *Fig. S5*). Some genera within the Gamma-, Delta-, and Epsilonproteobacteria (e.g., *Edwardsiella*, *Wolinella*, and *Anaeromyxobacter*) contained all three pathways. Indeed, denitrification and DNRA has recently been shown to be present and functional in the same bacteria (Mania et al., 2014; Yoon et al., 2015). We also detected genera that only carried the dissimilatory nitrite to ammonium pathway (in addition to ammonia assimilation), as was the case for five genera within the phylum Bacteriodetes.

More broadly, soil genera, and the phyla they fall into, varied in their degree of pathway specialization. Genera within the Cyanobacteria seemed to be specialists, carrying primarily the assimilatory nitrite to ammonium and N fixation pathways. In contrast, genera within the Deltaproteobacteria seemed to be N-cycling generalists, harboring up to six pathways (in addition to ammonia assimilation). Note, however, that these patterns do not distinguish between whether these genera are made up of generalists that encode many pathways or multiple specialists that encode specific pathways.

Focusing on each pathway individually revealed the most prominent taxa carrying the pathway across all soil samples. Here we consider two contrasting pathways, both in terms of their taxonomic distribution and the degree to which they have been studied. First, the abundance of the N fixation pathway in the soil samples was distributed broadly among both Archaea and Bacteria (Fig. 10.4 and *Fig. S5*). The most abundant N fixers detected were concentrated within the phylum Proteobacteria, with notable exceptions among the Chlorobi, Firmicutes, and Cyanobacteria (Fig. 10.5*A*). Most sequences were closely related to N-fixing genera that might be predicted to be common in soil, such as *Bradyrhizobium* and *Burkholderia*. Other abundant genera were less expected. For example, *Azoarcus* is an organism studied for its abilities to degrade soil contaminants (Sun and Cupples, 2012), and *Pectobacterium* (Gammaproteobacteria) is known primarily as a plant pathogen (Ma et al., 2007). Indeed, although it is known that *Pectobacterium* encodes the suite of N fixation genes, it remains unclear whether they are functional (Toth et al., 2015).

TABLE 10.1 Variation Explained by the Environmental Variables in the Regression Models of the Frequency of All (Total) and Individual N Pathways

	Individual Pathways (Residuals)			
Environmental Variables	Total	Ammonia Assimilation	Assimilatory Nitrate to Nitrite	Assimilatory Nitrite to Ammonia
Bacteria				
Habitat (H)	0.14	0.02	0.23	0.07
Precipitation (P)		<0.01		
Temperature (T)		<0.01		
pH				
Organic Carbon (C)	0.12			
Total N	0.05			
H × P	<0.01		0.07	
H × T			0.09	0.23
H × pH	<0.01		0.09	
H × C	0.1			
H × N	0.17	0.49		0.06
P × T			0.02	
C × N				
Adjusted R^2	0.58	0.51	0.5	0.36
Archaea				
Habitat			0.08	
Precipitation				
Temperature				
pH				
Organic Carbon				
Total N			0.05	
H × P			0.21	
H × T	0.09		0.18	
H × pH				
H × C				
H × N				
P × T				
C × N	0.34			
Adjusted R^2	0.43	NS	0.52	NS

NOTES: The models for the individual pathways are based on the residual frequencies of the pathway after correcting for the Total N pathway frequency (see text). Estimates of the fraction of explained variation are only reported for significant variables ($P < 0.05$). Samples were only included when all environmental variables could be obtained for that location ($n = 99$). NA, not assessed; NS, not statistically significant.

N Fixation	Nitrification	Dissimilatory Nitrite to Nitrite	Denitrification	Dissimilatory Nitrate to Ammonia
0.29	0.11	0.06	0.09	
0.02			0.02	
	<0.01			
0.13				
	0.08	0.32		
0.31	0.21	0.03	0.31	
0.05				
	0.05		<0.01	
		<0.01		
0.8	0.45	0.41	0.41	NS
0.09	0.03		0.12	
<0.01				
	0.02		0.03	
			<0.01	
	0.04			
0.09				
0.33	0.13			
0.12			0.06	
0.63	0.22	NS	0.21	NA

Second, the pathway encoding dissimilatory nitrite reduction to ammonium was also broadly distributed across soil bacteria (Fig. 10.4), as noted before (Welsh et al., 2014). However, the dominant soil taxa were restricted to two phyla, the Deltaproteobacteria and Verrucomicrobia (Fig. 10.5*B*). Verrucomicrobia are known to be abundant in soils, but their ecological role remains unclear (Bergmann et al., 2011; Fierer et al.,

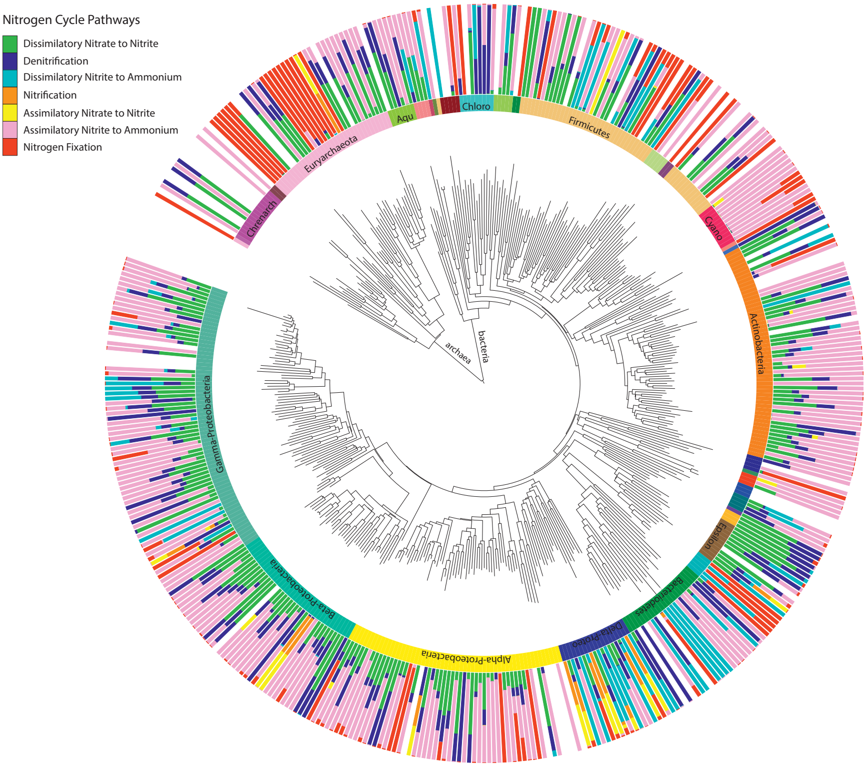


FIGURE 10.4 Phylogenetic distribution of N pathways in the soil metagenomes. A neighbor-joining tree was constructed using 16S rRNA sequences (*Materials and Methods*) and includes all Archaea and Bacteria genera associated with N cycling sequences in the dataset. The outer circle plots the proportion of N cycle reads assigned to each pathway within the genus. The ammonia assimilation pathway is excluded because it was found in all genera represented by at least 10 sequences. The inner circle indicates major classes and phyla. See *Fig. S5* for a high-resolution figure with genus labels.

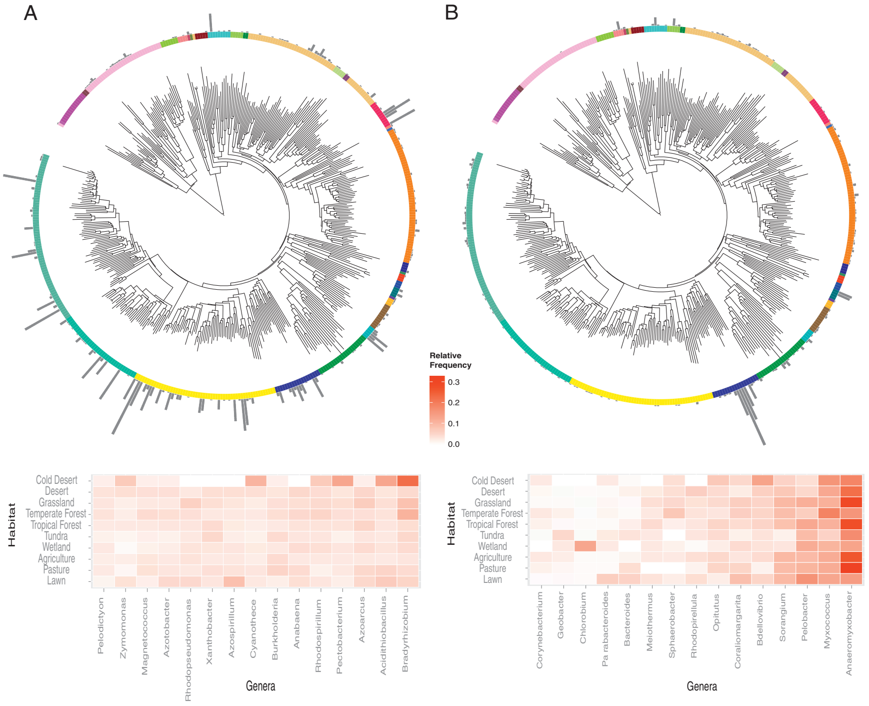


FIGURE 10.5 Phylogenetic distribution of genera encoding specific pathways. The neighbor-joining tree was constructed using 16S rRNA data (*Materials and Methods*). The relative abundance of genera associated with (*A*) N fixation and (*B*) dissimilatory nitrite reduction to ammonium. Within a pathway, the proportion of sequences associated with each genera was calculated for each sample. The proportions were averaged across libraries within each location and then averaged across the 10 habitat types (to provide equal weighting to the habitats). The heat maps give the relative frequency by habitat for the 15 most abundant genera associated with each pathway. See *Fig. S6* for plots of all pathways with genus labels.

2013). The pathway's most abundant genus, *Anaeromyxobacter* (phylum Deltaproteobacteria), is common in agricultural soil and has recently been shown to carry out a previously unrecognized process of nondenitrifying N_2O reduction to N_2 (Sanford et al., 2012). The relative abundances of genera encoding the other six N pathways in the soil samples are reported in *Fig. S6*.

Drivers of Taxonomic Composition by N Pathway

The same environmental variables that explained the overall frequency of the N pathways well explained much less of the variation in the taxonomic composition of the organisms encoding the pathways. For the eight pathways, the models only explained 7–19% of the composition variation of the individual N pathways (*Table S1*). However, as for pathway frequency, habitat type was the best predictor of composition, explaining up to 14% of the compositional variation in the assimilatory nitrite to ammonium pathway. Temperature also explained 11% of the compositional variation for the nitrification pathway. All other predictors, including pH, explained at most 3% of the variation for any pathway.

A closer examination of two pathways confirms weak compositional differences between the habitats. The 15 most abundant genera carrying the N fixation pathway were similarly abundant across all habitats except in cold deserts (Fig. 10.5*A*). The most abundant genera encoding the dissimilatory nitrite reduction to ammonium pathway displayed greater variability among habitats (confirming the model results in *Table S1*), but of these only one genus (*Chlorobium*) seemed specialized on a habitat (wetland) (Fig. 10.5*B*).

DISCUSSION

Here, we used metagenomic data to characterize the biogeographic patterns of microbial N cycling traits in soil. The advantage of this approach is that it allows us to identify the traits—and the organisms harboring them—involved in many key functions at once. Specifically, the analysis provides a comprehensive map of the dominant lineages involved in eight N processes. The approach also allowed us to search all known genes in a pathway, while avoiding primer biases toward particular lineages (Myrold et al., 2014).

The overall structure of microbial N traits—the relative frequency of the eight pathways—seems to be quite consistent across soils. This is not unexpected but had not been previously tested. For instance, the ammonia assimilation pathway was relatively common, and the pathways for N fixation and nitrification were relatively rare, as observed previously in soil and other environments (Varin et al., 2010; Martiny et al., 2013; Quinn et al., 2014; Souza et al., 2015). Less expected, however, was that N pathway frequencies within a soil sample were overwhelmingly positively correlated (Fig. 10.3). This result suggests that soil communities with high numbers of cells able to use one N pathway also generally support higher numbers of cells that can use other N pathways. Greater numbers of metagenomic sequences associated with nutrient cycles have previously been interpreted to be indicative of faster nutrient cycling rates

(Fierer et al., 2012). The positive correlations between pathways within the N cycle would seem to support this hypothesis. We also found a high frequency of Bacteria encoding the dissimilatory nitrite reduction to ammonium pathway, which leads to recycling of N in soils. The balance between DNRA and denitrification, which leads to the loss of N to the atmosphere, is thought to be key to soil N budgets. Our results confirm previous studies suggesting that this pathway may be more common than previously thought (Rutting et al., 2011; Nelson et al., 2015), but the taxa encoding the process in soil environments remain to be carefully characterized (Kraft et al., 2011).

The frequency of N traits further displayed clear biogeographic patterns. At the broadest scale, N trait frequency in Bacteria tended to decrease at higher latitudes, perhaps reflecting a general trend in N limitation in high-latitude ecosystems (Yergeau et al., 2007). Beyond latitude, the frequency of N cycling traits in soil communities depended largely on habitat type as well as soil carbon and N concentrations. N traits were highest in human-dominated habitats, where N inputs tend to be high, and tropical forests, which are generally thought to be less limited by N than temperate ecosystems (Vitousek, 1984). In contrast, N traits were lowest in cold deserts (Antarctic and Arctic), which are highly nutrient-limited (Jonasson et al., 1999; Yergeau et al., 2007). However, given the low sample numbers for some habitat types, it will be important to retest these patterns as more data accumulate.

Contrary to our hypothesis, the taxa responsible for each N pathway did not vary greatly by habitat type. Within a pathway, genera that were dominant in one habitat tended to be dominant in all habitats. More generally, the environmental variables in our analyses were poor predictors of the compositional variation of the N functional groups. One possible reason for this result is that environmental preferences are conserved below the genus level and therefore would not be detected by our analysis. However, this reasoning does not explain why soil pH seems to have little influence on composition, because pH preference seems to be conserved at a broader taxonomic level (Lauber et al., 2009; Martiny et al., 2015). Perhaps N functional groups are less specialized for a particular pH environment than microorganisms with other functional roles, but distinct pH-associated lineages in ammonia-oxidizing Archaea indicate that this is not always the case (Gubry-Rangin et al., 2011). Alternatively, the estimates of soil pH might have been too spatially coarse to detect a pattern.

A well-recognized issue in calculating the frequencies of genes or pathways from metagenomic data is how to normalize for overall genome abundance in the library (Manor and Borenstein, 2015). This normalization step is prone to uncertainties related to variation in mean genome size among communities. To address this issue, we estimated the frequencies of N pathways in two ways: using a set of conserved marker genes as well as the total number of annotated sequences within a domain. The first approach should be sensitive to differences in genome size, whereas the second approach includes more sequence reads and is thus more statistically robust. Because the two approaches led to similar findings, we conclude that the overall patterns in N pathway frequencies are likely not an artifact of normalization.

In sum, this study provides a foundation for future trait-based investigations of soil N cycling but also highlights two major challenges. First, we still know very little about how variability in the frequency and composition of microbial N traits will affect process rates in soil environments (Prosser, 2015). Indeed, a recent review found little correlation between an individual gene's abundance and the process rates that such genes encode. However, assessment of these links using metagenomic datasets is still needed (Rocca et al., 2015). Second, assigning function and taxonomy from short-read sequences is limited by genomic databases where annotations in some cases may be sparse and/or erroneous (Wu et al., 2009; Thomas et al., 2012). The N cycle is an archetype of this problem, because new N processes and lineages continue to be identified (Strous et al., 1999; Könneke et al., 2005; Farnelid et al., 2011; Sanford et al., 2012; van Kessel et al., 2015). Despite these challenges, the application of metagenomic data to a trait-based framework offers a powerful avenue for elucidating the role that microbial communities play in regulating biogeochemical processes (Barberán et al., 2012; Fierer et al., 2014).

MATERIALS AND METHODS

Dataset and Curation

Metagenomic samples (sequencing type "whole genome sequencing" and environmental package "soil," n = 809) in the MG-RAST database (Meyer et al., 2008) were classified into one of 10 habitat types (desert, cold desert, grassland, temperate forest, tropical forest, tundra, wetland, agriculture, pasture, and lawn). Samples that could not be classified into these habitats (e.g., oil spill, mines, and microbial mats) were not considered further.

Global Positioning System coordinates and sample date associated with each metagenome identification were downloaded from MG-RAST via the R package *matR* (R Core Team, 2011; Braithwaite and Keegan, 2013). To minimize the problem of pseudoreplication, we only considered samples from one date per location (the date with the most samples). Based on the statistics provided by MG-RAST, we further removed samples if (*i*) the number of uploaded sequences was equal to the number of post-QC sequences, which seemed to indicate a preprocessing step; (*ii*) the number of identified protein features was <10,000; or (*iii*) the total bacterial reads was <10,000. The remaining metagenomic libraries ($n = 365$) encompassed 118 unique locations. These were downloaded using the MG-RAST API version 3.2 with KEGG database annotations. Each sequence was assigned to the closest related genus in the database using an e-value of $\leq 10^{-5}$.

Data Standardization Across Metagenomic Libraries

Because sequencing effort varied greatly among samples, we standardized the bacterial and archaeal sequences by a suite of conserved, single-copy (i.e., marker) genes to control for possible variation in average genome size among samples (Nayfach and Pollard, 2015) (*Fig. S1*). The Kegg orthology numbers for 30 Bacteria and Archaea marker genes (Nayfach and Pollard, 2015) were matched to MD5 IDs using the nonredundant M5nr database. We then searched for these MD5 IDs in the samples annotated by the MG-RAST server.

The number of marker genes was also highly correlated with the total number of annotated sequences in a sample ($R^2 = 0.86$; *Fig. S1*). Thus, when comparing across Archaea, Bacteria, and Fungi, we standardized the samples by total annotated sequences. Sequencing effort varied greatly among the samples but was not overtly biased toward any particular habitat type (*Fig. S1*).

Identification of N Cycle Pathways

In each metagenomic library, we searched for sequences from eight N pathways, defined previously in Nelson et al. (2015). These pathways included nitrification (number of genes targeted: $n = 2$), N fixation ($n = 20$), denitrification ($n = 20$), dissimilatory nitrate to nitrite reduction ($n = 9$), dissimilatory nitrite to ammonia reduction ($n = 4$), assimilatory nitrate to nitrite reduction ($n = 2$), assimilatory nitrite to ammonia reduction ($n = 2$), and ammonia assimilation ($n = 10$) (Fig. 10.2*A*). If a gene from a pathway was detected, we assumed the presence of the entire pathway.

Environmental Metadata

Environmental data were retrieved from a variety of publically available sources. In all cases, gridded spatial data files were downloaded, and data were extracted using the R packages *raster*, *rdgal*, and *sp* (Hijmans and van Etten, 2012; Bivand et al., 2013). The data included average precipitation (millimeters) and temperature (degrees Celsius) from the month of sampling (Hijmans et al., 2005), soil pH (Batjes, 2000), total organic carbon (kilograms per square meter) (Batjes, 2000), total organic N (grams per square meter) (Group GSDT, 2000), and N deposition (milligrams of N per square meter per year) (Dentener et al., 2006). Approximate data grid resolution for precipitation and temperature was 0.01°, for soil pH and organic carbon was 0.5°, for total organic N was 0.1°, and for N deposition was 4.0°. Environmental metadata were assigned to each sample using the associated latitude and longitude coordinates. Where data were categorized into ranges (soil pH and total organic carbon), the average value from the range was used.

Statistical Analyses

To compare the relative abundance of N pathways across samples, we calculated the frequency of each pathway in a sample for both the Bacteria and Archaea. This frequency is the estimated number of times the pathway was detected per marker gene detected, or [number of pathway reads/number of pathway genes searched]/[number of marker gene reads/30]. Thus, a pathway's frequency of detection was also standardized for the number of genes in the pathway.

To test for differences in the frequency across pathways, we used a one-way analysis of variance, using the aov function in R. To test for correlations between the frequencies of the individual pathways within a sample, we used Spearman's correlation coefficient. To calculate the total N pathway frequency of each sample, we summed the frequency of all eight pathways. We used lm in R to calculate the residuals of each N pathway against a sample's total N pathway frequency.

To tease apart the relative importance of environmental variables on the frequency of N pathways, we used a multiple regression model (lm function in R) including the following variables: habitat type, temperature, precipitation, soil pH, organic carbon, and total N. For this analysis, we averaged data across multiple samples from the same location at just one sampling time, yielding 118 datasets. Based on a priori expectations (Ramette, 2007), we also included the following interaction terms: habitat by temperature, habitat by precipitation, habitat by soil pH, habitat by organic carbon, habitat by total N, precipitation by temperature, and organic carbon by total N. To determine the relative importance of the various significant environmental factors from our model in contributing to variation in the frequency of N pathways across samples, we used a backward selection procedure (Mac Nally, 2002; Ramette, 2007). Starting with the significant terms ($P < 0.01$) from our original model, we removed variables one at a time; the differences in R^2 values between each step were used to calculate the relative importance of the independent variable removed from the model. If there was no change or only a marginal change in R^2 when the term was removed, the term was assigned a relative importance of <0.01. After the initial analysis, N deposition was added to test whether this parameter improved the model.

To analyze the composition within each pathway, we calculated the proportional abundance of the genera in a sample and averaged these proportions across multiple samples from the same location. We then calculated a Bray–Curtis distance matrix for all sample locations. We used a distance-based linear model [DISTLM; PRIMER v6; PERMANOVA++ (Clarke and Warwick, 2001; Anderson et al., 2008)] to test the significance and importance (an estimate of the proportion of R^2 explained) of the predictor variables for each pathway's composition, using a forward selection procedure.

Phylogenetic Visualization

We constructed a phylogenetic tree including a representative species from all genera encoding N sequences using 16S rRNA amplicon data (chosen for their sequence quality and length of ~1,400 bp) from the SILVA database (Quast et al., 2013). We aligned the sequences using SINA (Pruesse et al., 2012) and created a neighbor-joining tree with the default parameters in Geneious v9.0.5. We used the Interactive Tree of Life (iTOL) (Letunic and Bork, 2007) to plot (*i*) the proportion of N pathways (excluding ammonia assimilation) detected within each genus and (*ii*) the relative abundance of genera encoding each individual pathway across the unique sampling locations ($n = 118$). For the N fixation and dissimilatory nitrate reduction pathways, we used the *ggplot2* package (Wickham, 2009) in R to plot heat maps of the relative frequencies of the 15 most abundant genera by habitat.

ACKNOWLEDGMENTS

We thank John Avise, Francisco Ayala, and Brian Bowen for the invitation to participate in this colloquium and Alex Chase for helpful feedback on earlier drafts of the manuscript. This work was supported by a U.S. Department of Education Graduate Assistance in Areas of National Need Fellowship (to M.B.N.) and U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research Grant DE-PS02-09ER09-25.

11

Phenotypes in Phylogeography: Species' Traits, Environmental Variation, and Vertebrate Diversification

KELLY R. ZAMUDIO,*# RAYNA C. BELL,†‡§
AND NICHOLAS A. MASON*

Almost 30 years ago, the field of intraspecific phylogeography laid the foundation for spatially explicit and genealogically informed studies of population divergence. With new methods and markers, the focus in phylogeography shifted to previously unrecognized geographic genetic variation, thus reducing the attention paid to phenotypic variation in those same diverging lineages. Although phenotypic differences among lineages once provided the main data for studies of evolutionary change, the mechanisms shaping phenotypic differentiation and their integration with intraspecific genetic structure have been underexplored in phylogeographic studies. However, phenotypes are targets of selection and play important roles in species performance, recognition, and diversification. Here, we focus on three questions. First, how can phenotypes elucidate mechanisms underlying concordant or idiosyncratic responses of vertebrate species evolving in shared landscapes? Second, what mechanisms underlie the concordance or discordance of phenotypic and phylogeographic differentiation? Third, how can phylogeography contribute to our understanding of functional phenotypic evolution? We

---

*Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853; †Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720; ‡Department of Integrative Biology, University of California, Berkeley, CA 94720; and §Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560. #To whom correspondence should be addressed. Email: kelly.zamudio@cornell.edu.

demonstrate that the integration of phenotypic data extends the reach of phylogeography to explain the origin and maintenance of biodiversity. Finally, we stress the importance of natural history collections as sources of high-quality phenotypic data that span temporal and spatial axes.

Phylogeography, as originally defined, focused on processes governing the spatial distribution of genealogical lineages within species (Avise et al., 1987). One of the strengths of the field at its inception was formalizing conceptual links among heredity (processes at the level of individual pedigrees), divergence at the population level, and phylogenetic relationships among species (Avise et al., 1987). This analytical framework bridged microevolutionary processes acting within populations and macroevolutionary patterns at larger spatial and temporal scales. From the earliest applications, empirical phylogeographic studies described spatial patterns of genetic diversity and inferred underlying mechanisms, thus contributing to the explanatory and predictive power of the field (Buckley, 2009). If most species show phylogeographic structure caused by landscape features that impede gene flow, then the geographic distribution of divergent lineages should coincide among species that coinhabit those landscapes. Further, phylogeographic breaks or contact zones should arise as lineages diverge allopatrically or come into secondary contact after divergence, respectively. This explicit prediction (Avise et al., 1987) resulted in a search for shared geographic patterns in genetic structure among species and the birth of comparative phylogeography (Schneider et al., 1998; Hewitt, 2000). Now, with thousands of taxon-specific phylogeographic studies published and synthesized in comparative studies (Beheregaray, 2008; Camargo et al., 2010), we have learned a tremendous amount about the geography of genetic structure both within and among species.

Phylogeography has progressed rapidly in the last three decades, with new genetic markers (Edwards et al., 2015; Garrick et al., 2015), analytical techniques (Hickerson et al., 2010), and synergies with landscape ecology and population genetics (Chan et al., 2011; Hickerson et al., 2006b; Oaks, 2014). Combined, these advances have revealed previously unrecognized genetic variation and its spatial and environmental correlates; however, phenotypic variation in those same diverging lineages has not received the same attention. Phenotypic variation among populations across a species' range is common and often serves as the initial motivation for phylogeographic studies; however, most studies focus primarily on spatial variation in genetic lineages (Edwards et al., 2015) and their distribution relative to environmental or geographic features of the landscape. Phenotypes are targets of selection and affect the performance of organisms in variable environments; combined, both processes contribute to diversification. Furthermore, different classes of phenotypes vary in how they impact processes such as dispersal, colonization, and persistence, thereby providing a window into the importance of various evolutionary processes in current and historical selective environments. Genetic structure of neutral genes, on the other hand, primarily reflects demographic processes (e.g., drift, expansion, changes in effective population size) that are a consequence of historical biotic and abiotic conditions during a species' evolutionary history. Thus, a new conceptual framework that explicitly integrates quantitative analyses of phenotypic variation within a phylogeographic framework can greatly enhance our knowledge of how genetic and phenotypic divergence arise, how they are linked, and how they respond to changing ecological and evolutionary contexts.

Here we review three research areas that exemplify the benefits of integrating phenotypic and genetic datasets in vertebrate phylogeography. First, we review how species-specific traits and their interactions with the environment predict concordant or idiosyncratic phylogeographic patterns among codistributed species. Second, we examine mechanisms that underlie the spatial concordance or discordance between phenotypic and genetic diversification. Third, we consider how phylogeography contributes to our understanding of functional phenotypic variation. For each topic, we describe case studies to highlight how the integration of phenotypic and genetic evolution has contributed to long-standing questions in evolutionary biology and has advanced our understanding of biodiversity. Finally, we emphasize the importance of natural history and field collections for the successful integration of organismal phenotypes and phylogeographic studies.

SPECIES-SPECIFIC TRAITS AND IDIOSYNCRATIC PHYLOGEOGRAPHIC PATTERNS

Comparative phylogeography seeks to characterize concordant phylogeographic breaks or contact zones, biogeographic "hotspots" for understanding mechanisms shaping genetic structure within and among species with shared distributions (Rissler and Smith, 2010; Moritz et al., 2009). A common assumption of comparative phylogeography is that taxa evolving in particular landscapes respond similarly to the abiotic and biotic elements that cause genetic divergence. We know, however, that species and populations vary in tolerance, plasticity, adaptive potential, and biotic interactions, all of which mediate responses to environmental variation (Bernardo and Spotila, 2006; Ridenhour et al., 2007; Satler et al., 2016) and ultimately dictate the degree of spatial and temporal concordance in genetic structure. The early definition of phylogeographic response categories acknowledged that differences could stem from species-specific traits such as dispersal potential and life history (Avise et al., 1986, 1987). Not surprisingly, species that are exceptions to regional phylogeographic patterns have been identified in most, if not all, phylogeographic hotspots, precluding generalizations and challenging expectations for shared causes of organismal diversification.

Given that species-specific phenotypes can dictate spatial variation in population responses to environmental change, phylogeography would benefit from a more integrative and inclusive framework, one that incorporates predictions based on those phenotypes, an approach that has been termed "trait-based phylogeography" (Paz et al., 2015). A parallel example of trait-based approaches can be found in the emerging field of biodiversity and ecosystem function, which arose at the interface of community and ecosystem science (Loreau, 2010). This new framework breaks from the view of species diversity as an epiphenomenon driven by a combination of abiotic environmental factors (e.g., temperature, rainfall, soil fertility), ecosystem processes that are themselves determined by these abiotic factors (e.g., productivity, biomass and nutrient cycling), and biotic interactions among species within communities (e.g., competition and predation). Instead, this new field considers biodiversity—in particular, the identity and diversity of species—as a driver of ecosystem functioning (Loreau, 2010) and establishes causality between a species' traits and the processes that in turn have functional consequences for ecosystems (Loreau, 2010). Our current view of biodiversity in phylogeography parallels the "old view" in ecology by considering the genetic structure of species as a consequence of abiotic conditions and the evolutionary "function" of lineage births and deaths. In other words, species themselves, and their traits, are typically not considered as functionally causal in the processes that ultimately shape them. When we consider that traits can alter an organism's demography and interactions with the environment, we can no longer ignore the dynamic nature of these interactions and their impact on lineage diversification (He et al., 2013). Thus, this paradigm shift challenges the expectation that temporal and spatial concordance among species should be the expected pattern in comparative phylogeography (Papadopoulou and Knowles, 2015b).

Phenotypes can either promote or constrain population divergence, depending on their function and interaction with the environment. For example, phenotypes that directly affect dispersal or persistence in new environments, such as those related to locomotor efficiency, physiological tolerance, or body size, can influence the frequency of migration and gene flow among subdivided populations. Others, such as recruitment rate, life span, and time to maturity will affect population size and turnover and thus the amount of genetic variation in subdivided populations. Finally, sexually selected phenotypes may not affect demography directly but can affect the distribution of genetic diversity indirectly via assortative mating, species recognition, and inbreeding avoidance. Variation in the distribution of phenotypes with different functions and the concordance of these phenotypes across species provide opportunities to quantify the importance of specific evolutionary processes for species inhabiting similar environments.

We are not implying that researchers have completely ignored species' traits in interpreting phylogeographic patterns. Many studies consider dispersal capacity, environmental tolerance, and other characteristics that contribute to diversification (Wang and Summers, 2010; Bell et al., 2012; Moritz et al., 2012; Smith et al., 2014b; Paz et al., 2015) or apply predictions derived from species-specific traits in a priori hypothesis testing (Brumfield and Capparella, 1996; Brown and Knowles, 2012; He et al., 2013; Massatti and Knowles, 2014). To date, studies adopting a comparative trait-based framework typically have focused on groups of organisms evolving in and adapting to particular habitats. A comparison of four distantly related and allopatric temperate amphibian species demonstrated that population divergences are significantly lower in two desert species that breed in ephemeral habitats than in two species inhabiting mesic forested landscapes (Chan and Zamudio, 2009). The stochastic persistence of breeding ponds across years in arid habitats may select against site fidelity and favor increased dispersal and larger physiological tolerances to inhospitable environments (Chan and Zamudio, 2009). A second study in the tropics confirmed that topographic complexity and especially macrohabitat preferences had strong effects on population divergence, so that species occupying forests and topographically complex regions showed deeper phylogeographic structure (Rodríguez et al., 2015). Lower vagility across complex terrain and reliance on specific breeding habitats may lead to greater phylogeographic divergence in rainforest species. In contrast, species in more open landscapes typically use ephemeral and unpredictable breeding sites suitable for vagile generalists, possibly reducing intraspecific divergence (Rodríguez et al., 2015).

Although trait-based analyses of shared phylogeographic structure yield important correlational evidence for divergence mechanisms (Smith et al., 2014b; Paz et al., 2015), the next step in this predictive framework is to examine species-specific traits that are selected for in particular landscapes and to quantify the extent to which those traits then contribute to diversification. An important advance in this direction is the development of model-based phylogeographic methods that incorporate phenotypic variation. These efforts stem from the realization that lack of concordance in temporal and spatial clustering in codistributed taxa may not mean that taxa are not responding to a common landscape or climatic barrier (Oaks et al., 2013); rather, discrepancies may reflect variation in ecological traits and dispersal capabilities of taxa sampled across the presumed barrier (Papadopoulou and Knowles, Chapter 8, this volume). These efforts refine expectations for spatial concordance and temporally clustered divergences by explicitly including geography and trait-based responses for each species (Massatti and Knowles, 2014). A recent study examined trait-based phylogeographic predictions using flightless beetles that coinhabit the Cycladic Plateau in the central Aegean archipelago. The species differ in body size and associate with different soil types, both traits that affect dispersal capacity and persistence in habitat patches of different sizes (Papadopoulou and Knowles, 2015b). The authors investigated the relative importance of geographic factors and species-specific traits (soil-type preference, body size) on population divergence of 13 codistributed species. They found greater support for phylogeographic concordance when the null expectation of divergence times incorporated geographic and species-specific trait data (Papadopoulou and Knowles, 2015b). Efforts such as these to inform phylogeographic inferences with relevant differences among species have great potential for improving our understanding of how landscapes and species-specific traits interact during diversification (Knowles and Alvarado-Serrano, 2010; He et al., 2013).

These empirical studies demonstrate that integrating species' traits in phylogeographic studies can help explain the concordance (or discordance) of genealogical differentiation for species living in shared environments. This exciting prospect will greatly increase the impact of phylogeography in biodiversity science, and future studies need to consider explicitly how phenotypes can be incorporated into their predictive frameworks. Just as common responses to phylogeographic barriers became a null hypothesis early in the history of the field (Papadopoulou and Knowles, Chapter 8, this volume), an understanding of how organismal traits mediate responses to variable environments and demography will be necessary for a complete picture of the expected mode and rate of phylogeographic diversification (Papadopoulou and Knowles, 2015b). Achieving this understanding will require the development of metrics for quantifying phenotypic divergence and methods that explicitly incorporate those data in phylogeographic predictions and analyses (Knowles and Alvarado-Serrano, 2010). To be comparable, studies of geographic variation in phenotypes should strive to collect analogous data across species' ranges. Doing so may be challenging, because phenotypes of interest will vary among systems, and each field has unique standards for replicable data, but these efforts will greatly expand the database of comparative traits for broader investigations of organismal divergence.

EVOLUTIONARY MECHANISMS LINKING GENETIC AND PHENOTYPIC DIVERSIFICATION

All lineages harbor phenotypic and genetic variation among individuals, and this variation can be geographically partitioned in different ways (Fig. 11.1 and Table 11.1). In populations that remain geographically and genetically isolated, theory predicts that phenotypic differences will become more pronounced over time as the result of both neutral and selective processes (Gould and Johnston, 1972). However, geographic variation in phenotypes does not always coincide with phylogeographic breaks (Endler, 1973). One prominent goal of phylogeography is to infer the biological mechanisms that partition genetic and phenotypic variation among populations. Recent advances in modeling (Thomé and Carstens, Chapter 7, this volume), predictive frameworks for incorporating phenotypes into phylogeography (Papadopoulou and Knowles, Chapter 8, this volume), and DNA sequencing technologies (Edwards et al., 2015)

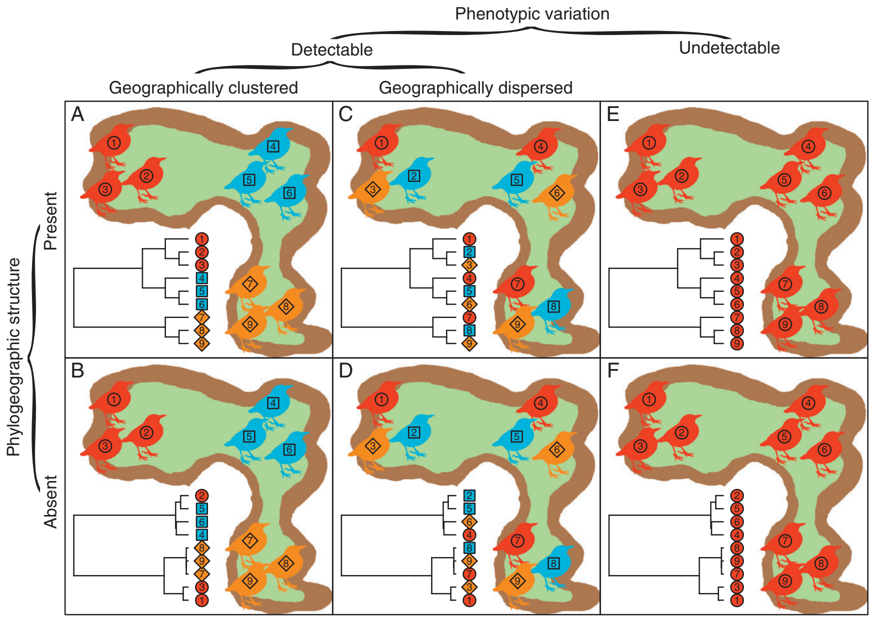


FIGURE 11.1 Potential patterns of phylogeographic structure and phenotypic diversity among populations (*A–F*). Although discrete morphs are depicted here, similar patterns can arise for phenotypes with continuous variation.

TABLE 11.1 Patterns of Phylogeographic Structure and Phenotypic Diversity in Vertebrates

Geographic Variation in Phenotypes	Phylogeographic Structure	Biological Mechanisms or Processes	Empirical Examples
Clustered	Present	Neutral divergence, local adaptation, divergent sexual selection	Neotropical oscines (Winger and Bates, 2015); treefrogs (Warwick et al., 2015); butterflyfishes (Waldrop et al., 2016)
Clustered	Absent	Rapid and recent diversification, phenotypic plasticity, gene flow	Dark-eyed juncos (McCormack et al., 2012); redpoll finches (Mason and Taylor, 2015); perch and roach (Faulks et al., 2015); spadefoot toads (Rice and Pfennig, 2010)
Dispersed	Present	Retention of ancestral polymorphism, convergent local adaptation, phenotypic plasticity, balancing selection	Side-blotched lizards (Corl et al., 2010); desert cichlids (Magalhaes et al., 2015); threespine sticklebacks (Colosimo et al., 2005); trout (Pearse et al., 2014); mollies (Pfenninger et al., 2015); humans (Tishkoff et al., 2007)
Dispersed	Absent	Convergent local adaptation, phenotypic plasticity, balancing selection, gene flow	Red crossbills (Benkman, 1996; Parchman et al., 2006); ground snakes (Cox and Davis Rabosky, 2013)
Undetectable	Present	Stabilizing selection, "cryptic" diversification	Sun skinks (Barley et al., 2015); black salamanders (Reilly and Wake, 2015); plain-backed thrushes (Alström et al., 2016); field voles (Paupério et al., 2012); rainforest skinks (Singhal and Moritz, 2013)
Undetectable	Absent	Stabilizing selection, gene flow	Straw-colored fruit bats (Peel et al., 2013)

NOTE: For each of the patterns illustrated in Fig. 11.1, we include potential mechanisms contributing to the distribution of phenotypic and genetic variation and empirical examples.

have bolstered our ability to identify and quantify the mechanisms that generate this diversity. Here, we illustrate those biological mechanisms that favor spatial concordance (or discordance) of genetic and phenotypic variation within a phylogeographic framework.

Geographically Clustered Phenotypes with Phylogeographic Structure

Phylogeographic structure often coincides with phenotypic variation (Fig. 11.1*A*) (Gould and Johnston, 1972). In fact, many phylogeographic studies are initially motivated by phenotypic differences among lineages separated by putative barriers to gene flow. Phylogeographic structure is most pronounced when populations are separated by nonpermeable barriers to gene flow and/or in taxa with limited capacity for dispersal across physical barriers (Hewitt, 2000). As a corollary, those lineages with deeper phylogeographic structure likely have greater levels of phenotypic divergence, arising from longer periods of independent neutral and selective evolution (Lande, 1980). A recent study compared genetic and phenotypic differentiation between allopatric, ecologically similar sister species from eight genera of Peruvian passerine birds in cloud forests separated by the Maranon Valley, a prominent physical barrier to gene flow among central Andean taxa (Winger and Bates, 2015). Allospecies showed a positive association between genetic divergence and phenotypic differentiation, especially in plumage differentiation, a trait important for social signaling and species recognition. Empirical studies in treefrogs (Warwick et al., 2015) and butterflyfishes (Waldrop et al., 2016) indicate that although the degree of phenotypic differentiation among vertebrates may scale with the strength and duration of genetic isolation, patterns can be highly idiosyncratic among taxa. Additionally, differences in the degree of phenotypic divergence among taxa or categories of traits can represent variation in evolutionary constraints or selective pressures driving the tempo and mode of phenotypic differentiation, which in turn can shape neutral genetic connectivity among populations.

Selective pressures vary temporally and geographically, thereby potentially altering the context-dependent fitness of phenotypic variants within lineages (Wang and Bradburd, 2014). If environmental conditions favor one phenotype, then populations may diverge phenotypically and genetically through local adaptation (Kawecki and Ebert, 2004). The spatial arrangement of suitable habitat in heterogeneous landscapes, such as mosaics or clines, can also promote geographically clustered phenotypic variation (Forester et al., 2016). For example, strawberry poison dart frogs are highly polymorphic, and genetic distances among populations are more strongly associated with phenotypic differences than with geographic distances, suggesting a role for local adaptation related to predation and aposematism (Wang and Summers, 2010). Similarly, genetic isolation imposed by ecological variation also contributes to geographically clustered phenotypic and genealogical diversity in Mediterranean blue tits (Charmantier et al., 2015) and bats (Morales et al., 2016). Thus, a phylogeographic framework can reveal patterns of phenotypic and genetic variation and their regional concordance (or lack thereof), thereby elucidating roles of neutral and selective processes in lineage differentiation.

Geographically Clustered Phenotypes Without Phylogeographic Structure

In many systems, geographically clustered phenotypic variation exists without phylogeographic structure (Fig. 11.1*B*). Phenotypic differences can arise quickly through localized divergent selection, or, alternatively, they may seem to arise quickly because traditional molecular markers may not detect the phylogenetic structure underlying rapid diversification (Edwards and Bensch, 2009). Dark-eyed juncos, for example, exhibit pronounced phenotypic variation that is geographically clustered among subspecies that exhibit little to no phylogeographic structure (McCormack et al., 2012). This pattern likely reflects subtle phylogeographic structure resulting from recent, rapid genetic isolation or adaptive divergence with ongoing gene flow among subspecies. Fortunately, high-throughput sequencing offers increased resolution and the capacity to distinguish between incomplete lineage sorting and ongoing gene flow, thereby improving our ability to infer ongoing biological processes in these cases (Edwards et al., Chapter 9, this volume).

Alternatively, clustered phenotypic variation among populations that lack phylogeographic structure may reflect biological processes, such as phenotypic plasticity (Crispo, 2008; Pfennig et al., 2010). Phenotypic plasticity involves developmental and phenotypic responses to different environmental regimes, thereby generating phenotypic diversity without genetic differentiation. For example, redpoll finches exhibit a gradient in which longer-billed individuals with streaking occur at lower latitudes, shorter-billed individuals with little to no streaking occur at higher latitudes, and many individuals express intermediate phenotypes (Mason and Taylor, 2015). However, genome-scale SNP analyses revealed that redpolls constitute a single gene pool, regardless of their phenotype or geographic origin, so the paucity of genetic differentiation among phenotypes at a continental scale is probably not the result of recent divergence and/or insufficient molecular data (Mason and Taylor, 2015). Furthermore, polygenic patterns of gene expression are strongly correlated with continuous variation in bill shape and plumage patterns, indicating possible roles for plasticity or variation in regulatory elements in generating and maintaining geographically clustered phenotypes (Mason and Taylor, 2015). Recent studies in perch and roach (Faulks et al., 2015) and spadefoot toads (Rice and Pfennig, 2010) similarly highlight the potential role of phenotypic plasticity in generating geographically clustered phenotypic variation without phylogeographic structure.

Distinguishing between recent, rapid bouts of adaptive genetic differentiation and phenotypic plasticity in natural populations can be difficult (Merilä and Hendry, 2014). Definitively demonstrating phenotypic plasticity requires experimental studies of captive populations or long-term longitudinal datasets, which are logistically challenging in many vertebrate, nonmodel species. Nonetheless, combining high-throughput sequencing with experimental studies to disentangle phenotypic plasticity and adaptive genetic divergence within a phylogeographic context is an exciting area of ongoing research.

Geographically Dispersed Phenotypes with Phylogeographic Structure

Many species are polymorphic with discrete or continuous phenotypic variation shared among phylogeographically structured populations (Fig. 11.1C). This pattern can be maintained through various evolutionary processes, including retention of ancestral polymorphism, balancing selection, parallel adaptation to locally variable conditions, and phenotypic plasticity. Multiple populations of side-blotched lizards in the western United States, for example, share polymorphism in throat color associated with different mating strategies that is maintained through negative frequency-dependent selection (Sinervo et al., 2000) and has persisted through multiple bouts of genetic isolation among populations (Corl et al., 2010).

Phenotypic plasticity also can contribute to geographically dispersed variation with phylogeographic structure. If plastic responses are retained among multiple populations that experience genetic isolation, then similar environmental conditions can result in shared polymorphism with phylogeographic structure. Desert cichlids restricted to the Cuatro Ciénegas valley in northern Mexico exhibit strong phylogeographic structure, and populations from each lagoon are genetically distinct from all others. However, both deep-bodied benthic morphs and slender-bodied limnetic morphs occur in each population, suggesting a role for phenotypic plasticity in generating and maintaining shared polymorphism among isolated populations (Magalhaes et al., 2015).

Parallel adaptation, or the independent evolution of similar adaptive phenotypes in different populations, is another potential mechanism for shared phenotypic variation in species with marked phylogeographic structure. Parallel adaptations evolve under similar selective pressures, often acting on standing ancestral genetic variation (Barrett and Schluter, 2008), resulting in similar phenotypes among genetically isolated lineages (Stern, 2013). Prominent examples of parallel evolution include changes in the ectodysplasin signaling pathway in threespine sticklebacks that result in reduced armor plating in multiple, independent freshwater populations (Colosimo et al., 2005). Life history differentiation in trout (Pearse et al., 2014), adaptations to sulfidic habitats in mollies (Pfenninger et al., 2015), and lactase persistence in humans (Tishkoff et al., 2007) underscore the potential for parallel evolution of shared phenotypic variation among geographically and genetically isolated populations across a large diversity of taxa.

Geographically dispersed and shared phenotypic variation with phylogeographic structure (Fig. 11.1C) may arise through similar processes that generate geographically clustered phenotypes without phylogeographic structure (Fig. 11.1*B*), albeit at different temporal, spatial, and phylogenetic scales. Parallel adaptive and plastic responses to similar temporal or spatial variation in environmental conditions within the range of each lineage can generate shared polymorphisms among populations, even if they are separated by prolonged bouts of genetic isolation (Fig. 11.1*C*). Furthermore, rapid adaptation and phenotypic plasticity in response to regional conditions also can induce geographically clustered phenotypic variation without perceivable phylogeographic structure at larger spatial scales, especially if gene flow reduces genetic isolation among populations (Fig. 11.1*B*). Thus, the same adaptive and plastic processes can shape geographic and phylogenetic partitioning of phenotypic and genetic variation over space and time at different scales.

Geographically Dispersed Phenotypes Without Phylogeographic Structure

Phenotypic variation also can be dispersed and shared among populations that do not exhibit phylogeographic structure (Fig. 11.1*D*). Intuitively, this pattern can arise if processes that generate shared phenotypic variation (e.g., adaptations to spatial or temporal environmental heterogeneity, balancing selection, or phenotypic plasticity) occur among populations that readily exchange genes with one another. Polymorphism in red crossbills provides an empirical example: Their lower mandibles curve either left or right to facilitate feeding on conifer cones, and polymorphism is maintained by frequency-dependent selection in multiple populations connected by gene flow and thus weakly differentiated (Benkman, 1996; Parchman et al., 2006). Likewise, highly polymorphic ground snakes include various color morphs that are present in multiple, genetically undifferentiated populations (Cox and Davis Rabosky, 2013). Although discrete forms of polymorphism are perhaps easier to identify, variation in continuous phenotypes—such as body size or limb length—is also common among populations that lack phylogeographic structure. Dispersed polymorphism among populations that lack phylogeographic structure is more likely in vagile than dispersal-limited taxa, and the maintenance of phenotypic diversity will depend on the strength of selection acting on phenotypic variants among populations connected by high levels of gene flow.

Uniform Phenotypes with Phylogeographic Structure

Many lineages exhibit phylogeographic structure with little or no detectable phenotypic variation (Fig. 11.1*E*). In the case of strong stabilizing selection acting on traits that characterize a species' niche, populations will track suitable habitat as it appears and disappears over time (Wiens et al., 2010). As such, ancestral populations can subdivide and accumulate genetic differences without morphological divergence, generating "cryptic" lineages or species (Bickford et al., 2007). Philippine sun skinks, for example, exhibit deep phylogenetic splits concordant with geography but show little to no morphological variation among lineages (Barley et al., 2015). Similar patterns in black salamanders (Reilly and Wake, 2015), plain-backed thrushes (Alström et al., 2016), and field voles (Paupério et al., 2012) indicate that cryptic lineages are prevalent among vertebrates. Although cryptic lineages may not persist if previously isolated populations come into contact and exchange genes freely, speciation theory predicts that reproductive isolation—and postzygotic reproductive isolation in particular—increases with divergence time between lineages, in part through the accrual of Dobzhansky-Muller incompatibilities (Bolnick and Near, 2005). Thus, study systems in which cryptic lineages occur in secondary contact provide an opportunity to address the evolution of postzygotic reproductive isolation between cryptic populations in nature. For example, divergence time and the degree of reproductive isolation are tightly correlated across five contact zones among cryptic lineages of rainforest skinks previously isolated in glacial refugia, indicating that, even in morphologically cryptic lineages, phylogeographic splits of increasing depth represent stages along the speciation continuum (Singhal and Moritz, 2013). Lineages with limited dispersal and prominent evolutionary or developmental constraints are most likely to develop phylogeographic structure with no perceivable phenotypic diversity (Bickford et al., 2007). We expect this pattern where strong biogeographic barriers to gene flow generate genetic divergence, and the selective environments between genetically isolated populations are relatively similar over space and time.

Uniform Phenotypes Without Phylogeographic Structure

Occasionally, vertebrate taxa do not vary perceivably in phenotypes within and among populations that do not also exhibit phylogeographic structure (Fig. 11.1*F*). Phenotypic uniformity among populations will be more likely if gene flow homogenizes populations (Lenormand, 2002). If environmental conditions are temporally and spatially consistent, then phenotypic variation may become fixed at an optimum, such that no phenotypic variation exists across a species' range (Kawecki and Ebert, 2004). Furthermore, if individuals consistently disperse and exchange genes with other populations, phylogeographic structure will not accumulate over time. Straw-colored fruit bats in continental Africa, for example, display no phylogeographic structure and do not vary in phenotype across their expansive distribution (Peel et al., 2013). Vagile taxa that have expanded their range rapidly from a single glacial refugium, such as the blackpoll warbler (Ralston and Kirchman, 2012), may display phenotypic uniformity and little to no phylogeographic structure among populations. A phylogeographic framework can disentangle cases in which continued gene flow homogenizes genetic and phenotypic variation among populations and cases in which historical demographic events, such as recent and/or rapid range expansions, have limited the time for phenotypic differences to accumulate (Marko and Hart, 2011).

As illustrated by the empirical and hypothetical examples discussed above (Fig. 11.1 and Table 11.1), the geographic partitioning of phenotypic and genotypic diversity is highly variable among vertebrates. Geographic partitioning of phenotypic variation across environmental gradients can occur with or without genetic differentiation, and in some cases the mechanisms for concordance (or lack thereof) are difficult to disentangle. Integrative approaches that combine high-throughput sequencing, experimental manipulations, and high-quality phenotypic datasets allow us to differentiate among biological mechanisms underlying phenotype-genotype concordance. Comparative studies adopting this framework will yield further examples of neutral divergence, local adaptation, and phenotypic and developmental plasticity, balancing selection and the prevalence of different evolutionary processes across taxa. Phylogeographic studies adopting this framework will also enhance our understanding of how rates and modes of phenotypic diversification vary among taxa. Finally, clarifying the evolutionary mechanisms underlying patterns of phenotypic and genetic diversity has implications for conserving biodiversity and for making accurate predictions of how species will respond to environmental change.

PHYLOGEOGRAPHY AND FUNCTIONAL VARIATION IN PHENOTYPES

Phylogeographic studies of functional variation in phenotypes have the potential to identify selective regimes that structure variation within and between species and ultimately shape the evolutionary history of functional traits. This approach can build on classic studies of hybrid zones and character evolution at macroevolutionary timescales by focusing on the spatial distribution of functional (or selected) traits within species. For instance, most hybrid zones result from secondary contact between populations or species that were previously allopatric; thus, the selective environment within the hybrid zone may not reflect the selective pressures that initiated divergence between parental lineages. Likewise, methods for quantifying character evolution above the species level typically do not account for within-species variation in phenotype and genotype (but see Revell and Graham Reynolds, 2012) or the range of environmental conditions across a given species' distribution. Thus, a phylogeographic approach that encompasses phenotypic and environmental variation within species and contextualizes the demographic history of functional traits can provide exceptional insights into how organismal diversity evolves.

Identifying the role of extrinsic barriers in shaping the geographic distribution of functional phenotypic variation is an essential first step for investigating local adaptation. Thus, many studies quantify divergence at neutral genetic markers to investigate whether historical barriers coincide with the geographic distribution of variation in phenotypes such as coloration (Wang and Summers, 2010) or physiology (Moritz et al., 2012). Within this historical context, phylogeographers then can investigate whether regional environmental variation and local adaptation contribute to phenotypic divergence (e.g., Ng et al., 2013) and identify instances of parallel phenotypic evolution among phylogeographic lineages that occupy similar selective environments in different geographic areas (Richmond and Reeder, 2002; Hoekstra et al., 2005). Even in study systems with only a single evolutionary origin of the trait of interest, a broader phylogeographic framework can inform the timing and/or direction of phenotypic change between genetically differentiated lineages. For example, conspicuous coloration and toxicity vary across the range of the granular poison dart frog, and prevailing evolutionary theory contends that these traits should evolve in a correlated fashion. By reconstructing phylogeographic relationships across the species' range, Wang (2011) demonstrated that the less conspicuous, more toxic population evolved from a more conspicuous, less toxic ancestor, thus challenging the view that conspicuousness and toxicity are tightly coupled. Finally, quantifying demographic processes, such as patterns of gene flow or changes in population size, and the distribution of phenotypic variation among divergent lineages can provide preliminary insights as to the strength of selection on adaptive phenotypes and whether gene flow between populations introduces adaptive phenotypes to new environments. For example, in the rock pocket mouse, in which melanic pelage evolves repeatedly on dark lava flows, high levels of gene flow between neighboring populations that differ phenotypically indicate that selection for color matching is strong (Hoekstra et al., 2004, 2005). Furthermore, high rates of gene flow between melanic mice populations inhabiting neighboring lava flows suggest that on a finer spatial scale, adaptive phenotypes in these different populations have a common genetic basis (Hoekstra et al., 2005). Thus, characterizing geographic patterns of divergence at neutral genetic markers and in phenotypes contextualizes the demographic history of adaptive traits, and this evolutionary perspective then can inform in-depth investigations to identify the selective environment in which these traits evolved.

Phylogeographic studies of functional traits can identify the processes that shape adaptive variation and estimate the strength of selection acting on phenotypic variation by building links between locally adapted genotypes/phenotypes, population demography, and environmental variation in selective regimes. The challenge of this approach, however, is that it requires identifying and quantifying adaptive phenotypes, obtaining samples that encompass relevant environmental and phenotypic variation across the species' range, and characterizing functional genetic variation underlying adaptive phenotypes as well as neutral variation to estimate population demographic history. Given that the genetic architecture of adaptive phenotypes is unknown in most nonmodel vertebrate taxa, this approach has been applied primarily in systems with extensive genomic resources (e.g., the threespine stickleback; Deagle et al., 2013) or in systems in which the links between a particular adaptive phenotype and the underlying functional genetic variation are well defined. These systems include adaptive shifts in coloration caused by variation in genes encoding the melanin pathway (Hoekstra et al., 2004), physiological adaptation to high-altitude environments caused by variation in hemoglobin subunit genes (Bulgarella et al., 2012), tetrodotoxin resistance caused by variation in skeletal muscle sodium channels (Feldman et al., 2009), and differences in adaptive immunity caused by variations in genes encoding the MHC class II subunits (Savage et al., 2015; Savage and Zamudio, 2016).

One common result of functional-trait studies is the identification of recurrent novel mutations underlying similar phenotypes (Nachman et al., 2003; Feldman et al., 2009; Savage and Zamudio, 2016). For example, mutations in the melanocortin-1 receptor gene (*Mclr*) are highly correlated with adaptive melanism in an Arizona population of rock pocket mice, but melanic populations in New Mexico show no association with variation at *Mclr*, indicating that their dark coloration must result from changes at different genes (Hoekstra and Nachman, 2003; Nachman et al., 2003; Hoekstra et al., 2005). This finding is in contrast to traits associated with threespine sticklebacks in which the same adaptive alleles underlie multiple independent freshwater invasions on a regional scale (Fisher et al., 2012). A second theme is the role of gene flow in promoting adaptive evolution in some contexts (Hoekstra et al., 2005) and impeding adaptation in others (Savage et al., 2015; Savage and Zamudio, 2016). Finally, geographic sampling that captures environmental variation is especially valuable for understanding complex adaptive scenarios such as host–pathogen dynamics. Across its range, the lowland leopard frog exhibits population-level variation in survival after infection by the fungal pathogen *Batrachochytrium dendrobatidis,* which causes chytridiomycosis, a disease implicated in population declines or extinction in hundreds of amphibian species worldwide (Fisher et al., 2012). Variation in immunity loci determines susceptibility to the pathogen such that several MHC alleles are strongly associated with increased survival or susceptibility in both experimental (Savage and Zamudio, 2011) and natural settings (Savage and Zamudio, 2016). These associations are decoupled, however, for populations surrounding a thermal spring, where warm water shields frogs from developing high pathogen loads but also precludes selection from increasing the frequency of MHC survival alleles (Savage and Zamudio, 2016).

In summary, quantifying functional genetic variation within the context of the phylogeographic history of a species and across the range of environments it inhabits can reveal how regional variation in selective regimes and demographic processes drives the evolution of adaptive phenotypes. Just as phylogeography initially formalized conceptual links among heredity, population divergence, and phylogenetic relationships among species, an analytical framework that advocates genealogical and spatially explicit analyses of intraspecific functional genetic and phenotypic variation will bridge microevolutionary processes acting on individual populations and macroevolutionary patterns at larger spatial and temporal scales. These integrative and rigorous approaches have been possible only in select systems to date, but identifying the underlying genetic basis of phenotypic variation within species is becoming increasingly tractable in vertebrates. Clearly, examining functional phenotypic variation in a phylogeographic framework holds great promise for exploring links between genotypic and phenotypic diversity and adaptation across variable environments.

FIELD STUDIES AND NATURAL HISTORY COLLECTIONS: SOURCES OF PHENOTYPIC DATA

Organismal phenotypes, many of which are the target of selection and play important roles in species performance in variable environments, are important components of how we identify and categorize biodiversity. We have argued that, despite the clear benefits of integrating patterns of phenotypic evolution into phylogeographic predictions and inferences, this integration has yet to be fully realized. One reason is that high-quality phenotypic data are difficult to obtain; establishing a phenotypic database with robust sample sizes and fine-scale spatial sampling can be laborious. This challenge is solvable by relying on well-established methods in biodiversity science.

For centuries, naturalists relied solely on phenotypes to document diversity, study the relationship of organisms with their environment, and infer evolutionary change. Unaware of the genetic underpinnings of those phenotypes, early naturalists focused on explaining phenotypic diversity not only among but also within species. They did so with a large number of phenotypes, including behavior, color, morphology, life history, and ecological traits, among others. Fortunately, many of those efforts are archived in publications or are preserved in natural history collections, providing a sample of Earth's biota that typically extends back to the 19th century, and often includes representative coverage of species' distributions (Graham et al., 2004; Holmes et al., 2016). For some kinds of phenotypes, such as certain aspects of morphology preserved in museum specimens, natural history collections are a rich source of phenotypic data. In contrast, for phenotypes that are not easily preserved—such as behavior, ecological associations, or physiological parameters—field studies will be the ultimate source, although the metadata associated with many preserved specimens often contain important information on behavior, habitat preferences, and other ecological associations (Holmes et al., 2016).

Natural history collections also will have a large role in the analysis of more recent evolutionary change in phenotypes (Holmes et al., 2016). For some species, long-term series of collections offer a unique opportunity to infer ancestral phenotypes and how those have changed with documented changes in the environment (Ożgo and Schilthuizen, 2012). Examining these data within a phylogeographic framework provides the evolutionary context to identify rangewide dynamics of phenotypic change and may highlight regional sources of adaptive variation. Furthermore, many of these phenotypic changes can be tied to strong selection imposed by changing environments, thus providing the link between genetic and phenotypic changes under different environmental contexts. These phenotype–genotype associations provide a mechanistic basis for inferring past changes at both recent and longer evolutionary time frames and a predictive framework for understanding how organisms will adapt to future natural and anthropogenic global change. Phenotypes, genetic structure, and environmental characteristics are intimately coupled in the processes of organismal divergence; thus advances in all three fields will enable the integrative study of divergences in natural populations.

CONCLUSIONS

We see great promise in ongoing methodological and conceptual advances that explicitly incorporate trait evolution in phylogeographic predictions and inferences. Our goal here was to highlight the many valuable avenues for future work in this area. The field of phylogeography has changed since its origin, incorporating new techniques, new analyses, and increasingly, different sources of data. Although only a fraction of extant taxa have been surveyed, the field has revealed many common patterns and mechanisms underlying diversification within broadly divergent taxonomic groups. Meanwhile, our ability to quantify genetic and phenotypic variation also has expanded; thus, the field of phylogeography now is poised for another integration, this time by incorporating data on phenotypic variation in diverging lineages, understanding the selective and genetic basis for that variation, and quantifying the role that phenotypes play in diversification. This integration has the potential to unify once again disparate fields in evolutionary biology, and address how interactions among abiotic landscape features and biological features of species shape biodiversity (Greene, 2005). This integrative framework is a powerful tool for understanding the effects of past global change on current biodiversity and for predicting the adaptive potential and resilience of species adapting to novel environments of the future.

ACKNOWLEDGMENTS

We thank J. Avise and F. Ayala for the invitation to participate in the Sackler Symposium *In the Light of Evolution X: Comparative Phylogeography*; and the K.R.Z., Lovette, Searle, and McGuire laboratories, A. Corl, A. Chavez, H. Greene, and two anonymous reviewers for constructive feedback on the manuscript. Our work is funded by National Science Foundation Research Grants DEB-0542848, DEB-1601072, and DEB-1309171; EPA Science to Achieve Results Fellowship F13F21201 (to N.A.M.); and a University of California Chancellor's Postdoctoral Fellowship (to R.C.B.).

12

Geogenetic Patterns in Mouse Lemurs (Genus *Microcebus*) Reveal the Ghosts of Madagascar's Forests Past

ANNE D. YODER,[*,†,§§] C. RYAN CAMPBELL,[*]
MARINA B. BLANCO,[†] MARIO DOS REIS,[‡]
JÖRG U. GANZHORN,[§] STEVEN M. GOODMAN,[‖,#]
KELSIE E. HUNNICUTT,[*] PETER A. LARSEN,[*]
PETER M. KAPPELER,[**] RODIN M. RASOLOARISON,[**,††]
JOSÉ M. RALISON,[††] DAVID L. SWOFFORD,[*] AND
DAVID W. WEISROCK[‡‡]

Phylogeographic analysis can be described as the study of the geological and climatological processes that have produced contemporary geographic distributions of populations and species. Here, we attempt to understand how the dynamic process of landscape change on Madagascar has shaped the distribution of a targeted clade of mouse lemurs (genus *Microcebus*) and, conversely, how phylogenetic and population genetic patterns in these small primates can reciprocally advance our understanding of Madagascar's prehuman environment. The degree to which human activity has impacted the natural plant communities of Madagascar is of critical and enduring interest. Today, the eastern rainforests are separated from the dry deciduous forests of the west by a large expanse of presumed anthropogenic grassland savanna, dominated by the

---

[*] Department of Biology, Duke University, Durham, NC 27708; [†] Duke Lemur Center, Duke University, Durham, NC 27705; [‡] School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom; [§] Tierökologie und Naturschutz, Universität Hamburg, 20146 Hamburg, Germany; [‖] Field Museum of Natural History, Chicago, IL 60605; [#] Association Vahatra, BP 3972, Antananarivo 101, Madagascar; [**] Behavioral Ecology and Sociobiology Unit, German Primate Centre, 37077 Goettingen, Germany; [††] Département de Biologie Animale, Université d'Antananarivo, BP 906, Antananarivo 101, Madagascar; and [‡‡] Department of Biology, University of Kentucky, Lexington, KY 40506. [§§] To whom correspondence should be addressed. Email: anne.yoder@duke.edu.

Family Poaceae, that blankets most of the Central Highlands. Although there is firm consensus that anthropogenic activities have transformed the original vegetation through agricultural and pastoral practices, the degree to which closed-canopy forest extended from the east to the west remains debated. Phylogenetic and population genetic patterns in a five-species clade of mouse lemurs suggest that longitudinal dispersal across the island was readily achieved throughout the Pleistocene, apparently ending at ~55 ka. By examining patterns of both inter- and intraspecific genetic diversity in mouse lemur species found in the eastern, western, and Central Highland zones, we conclude that the natural environment of the Central Highlands would have been mosaic, consisting of a matrix of wooded savanna that formed a transitional zone between the extremes of humid eastern and dry western forest types.

Madagascar is one of the most enigmatic landmasses on Earth and has long been identified as a unique biodiversity hotspot (Myers et al., 2000). Current estimates hold that close to 100% of the island's land mammals and native amphibians, 92% of its reptiles, and >90% of its plants are found nowhere else on Earth (Vences et al., 2009; Callmander et al., 2011). Best estimates indicate that the island harbors nearly 5% of Earth's species-level biodiversity even though it comprises only a little more than 0.01% of the planet's land-surface area. The mechanisms by which so much biodiversity arose on such a relatively small and remote corner of the Earth beg for phylogeographic and biogeographic investigation (Wilmé et al., 2006; Samonds et al., 2013).

Presently, Madagascar exhibits a remarkable east–west trend in precipitation, which is primarily the result of eastern trade winds coming off the Indian Ocean and the orographic effect of the island's eastern mountain range (Gautier and Goodman, 2003). An evergreen humid forest biome covers portions of the eastern lowlands and extends ~100 km inland along the north–south aligned eastern chain of mountains. At elevations above 800 m, and extending well into the island's interior, the humid forest transitions to the Central Highlands, which are dominated by moist montane forest. At higher elevations (generally above 1,900 m), the montane forest habitat gives way to an Ericaceae thicket. Along the western half of the island, below 800 m elevation and to the west of the Central Highlands, the montane forests shift to dry deciduous forest dominated by drought-adapted trees and shrubs. All these habitats have been extensively degraded and fragmented by human activities over the past few hundred years (Green and Sussman, 1990; Harper et al., 2007), and in most areas, particularly in the Central Highlands, little of the former natural vegetation remains. Although there is uniform consensus that human activities dramatically transformed the landscape at about 1,000 y BP (Dewar, 2014;

Goodman and Jungers, 2014; McConnell and Kull, 2014; Vorontsova et al., 2016), the prehuman condition remains debated. Interpretation of the prehuman vegetation via current phytogeographic classification of the island is constrained by the extant floristic affinities of plants (Perrier de la Bâthie, 1921; Humbert, 1955), with little data on natural formations from the Central Highlands remaining. Thus, the present classification separating the eastern and western biomes as extreme endpoints ignores the possibility that the Central Highlands might naturally have formed a vast zone of transitional vegetation.

The eastern humid forest, western dry deciduous forest, and the Central Highlands zones are now covered in part with what have been referred to as "secondary grasslands" or savanna composed primarily of Poaceae dominated by the subfamily Gramineae and subject to frequent burning. Today, the Central Highlands forms a stark habitat separation between the mesic zones of the east and the arid habitats of the west and south (Fig. 12.1). It has been the subject of ongoing debate whether this grassland formation is entirely the result of human-mediated transformation or whether it is better viewed as a landscape that has been modified from the original native wooded savanna interspersed with areas of forest (Bond et al., 2008; Goodman and Jungers, 2014). Our definition of "wooded savanna" is structurally very similar to the Miombo woodlands of southern Africa, which has irregular canopy cover and contains herbaceous plants, including grasses. The classic view of Madagascar's natural vegetation (hereafter referred to as the "forest hypothesis") holds that Madagascar was blanketed by closed-canopy forest across its entirety before the arrival of humans and that a 90% reduction in forest cover occurred primarily through human agency. Supporting this hypothesis, a recent stable isotope analysis from calcium carbonate cave deposits from Anjohibe Cave in the northwest found evidence for a rapid and complete transformation from a flora dominated by C_3 plants to a C_4 grassland system over the course of a single century, coincident with early human habitation of this portion of the island (Burns et al., 2016). This view, however, leaves unanswered the degree to which intermediate habitats already might have been subject to Quaternary climatic cycles, including the impacts of natural fires, before the arrival of humans as recently as 2,000–4,000 y BP (Dewar and Wright, 1993; Dewar et al., 2013).

Alternatively, a global analysis of the species richness, levels of endemism, and phylogenetic placement of Malagasy grasses indicates that naturally occurring grass formations have long been a feature of Madagascar's prehuman vegetation, existing extensively throughout the landscape from the Neogene (Vorontsova et al., 2016). An extension of these findings suggests that the grassland savanna that presently characterizes the Central Highlands is thus a remnant of the former naturally occurring

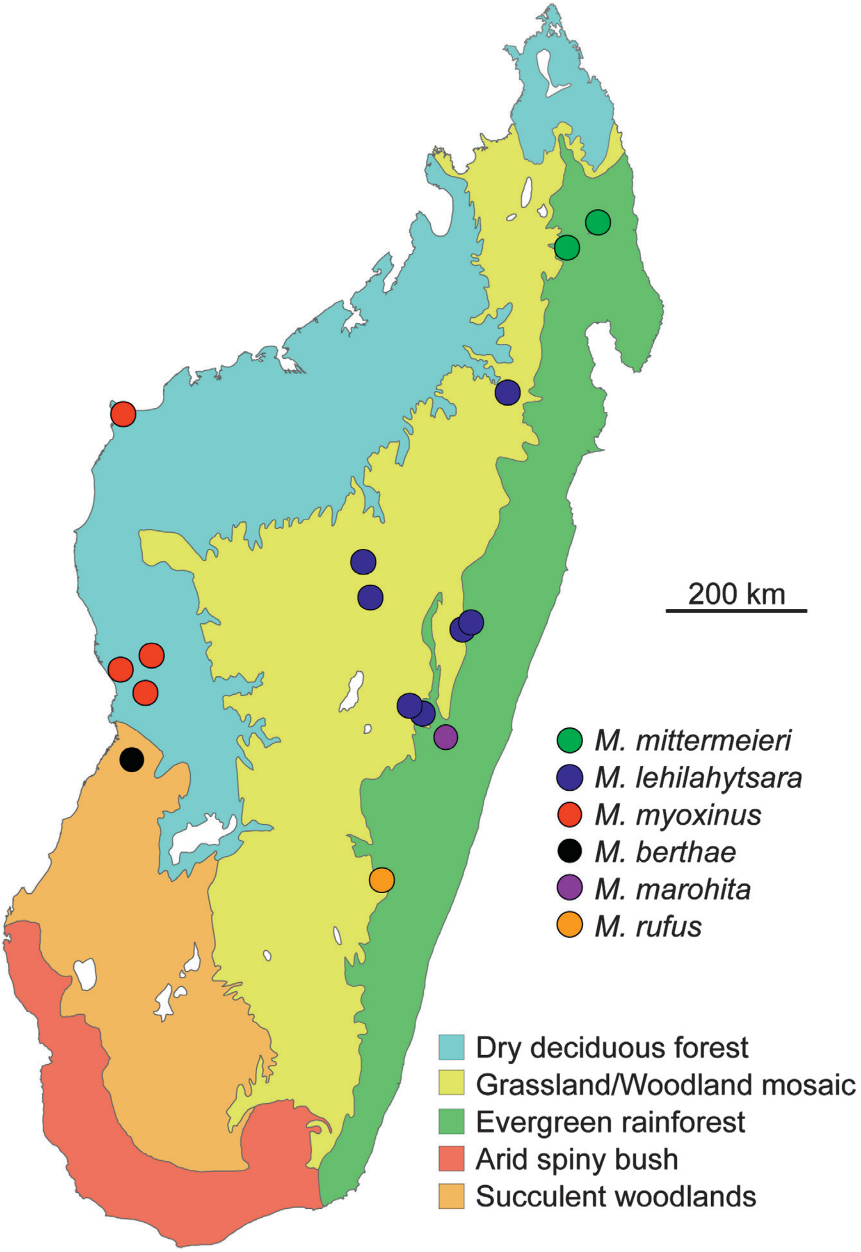


FIGURE 12.1 A map of Madagascar with the major biome types recognized in modern times and sampling locations used in this study for all 30 samples across six species within the genus *Microcebus*. Sampling counts at a given location varied between one and three individuals.

graminae vegetation, a hypothesis hereafter referred to as the "grassland hypothesis." An intermediate and more nuanced view of prehuman vegetation in Madagascar, hereafter referred to as the "mosaic hypothesis," holds that the Central Highlands were characterized by a mixture of closed-canopy forest and wooded savanna, with some regions being open and others closed with dense tree growth, and with marshlands presumably being partially dominated by Poaceae (Goodman and Jungers, 2014). The most detailed available palenological record in support of the mosaic hypothesis comes from a 40-m core of lake sediment collected from Lake Tritrivakely near Antsirabe in the center of the island (Gasse and Van Campo, 2001). The core spans ~150,000 years and shows that six cycles of warming and cooling occurred during this period. Coincident with these cycles was a progressive shift from heathland, to grassland, to grassland–forest mosaic, with a return to ericaceous heath in the coldest periods (Dewar, 2014). This view of Madagascar's prehuman vegetation has been developed over years of study focusing on the analysis of pollen spectra and charcoal influx from lake sediments across the Central Highlands (Burney, 1987, 1997; Gasse et al., 1994; Burney et al., 1997) and holds that this mosaic environment would have become fully established in the Quaternary, with grass formations being a far more recent feature of the Malagasy landscape than is proposed in the grassland hypothesis.

THE CENTRAL HIGHLANDS RUBICON

Our study aims to examine the relationship of the western and eastern zones and the extent to which the hypothesized Central Highland habitat matrix between these zones provided conduits for the dispersal of forest-dwelling mammals. There are a number of questions to be tested: Is the contemporary sharp habitat divide between the wet east and dry west biomes natural, or is it the remnant of a transitional east–west cline? If the latter, what would have been the ecological characteristics of the intervening Central Highlands? What effects have ecological separations and/or corridors had on the distribution and evolution of Madagascar's biota? Most importantly, when and how might forest-dwelling terrestrial vertebrates have dispersed between eastern and western habitats?

To address these questions, we examine phylogenetic and geogenetic (Bradburd et al., 2016) patterns in mouse lemurs (genus *Microcebus*) to determine both the timing and directionality of dispersal events between eastern and western Madagascar. We also investigate patterns of genetic structure within two broadly distributed species, one from western Madagascar and one from the Central Highlands, to determine if there are differential signatures of continuous versus discontinuous habitat in the two species. Such differential patterns will have potential consequences for the interpretation of the prehuman landscape and its impact on mouse lemur evolution. If the forest hypothesis applies, and the grassland savannas of the Central Highlands are an entirely recent phenomenon, we would expect to see extensive and recent patterns of connectivity among populations that presently are distributed in eastern and western biomes. Alternatively, if the grassland hypothesis holds, we should observe a strong and ancient separation between the species distributed in the eastern and western biomes. If, however, the mosaic hypothesis applies, we would expect to see a fairly rapid and relatively recent (i.e., Quaternary-age) diversification of lineages. Moreover, in the mosaic scenario, it is likely that the pulsing cycles of temperature and humidity characteristic of the Quaternary would have left genetic signatures of population isolation and reconnection among contemporary populations (Wilmé et al., 2006).

WHY MOUSE LEMURS?

Mouse lemurs are the world's smallest living primates, ranging in average adult body weight from 30 g in *Microcebus berthae* to ~80 g in *Microcebus marohita* (Rasoloarison et al., 2013). They are found in virtually every forested habitat in Madagascar and are known to have the highest reproductive rate, and thus the "fastest" life history, of all primates (Dewar and Richard, 2007). Martin (1972) reported that in natural populations mouse lemurs may give birth for the first time at the age of 12 mo. This finding has been confirmed in other natural populations (Eberle and Kappeler, 2004) and also in the captive colony of *Microcebus murinus* housed at the Duke Lemur Center (Zehr et al., 2014). Field studies of various species within the genus have further verified that they are highly seasonal breeders, with female estrous showing a periodicity close to 365 days (Blanco, 2011). Thus, mouse lemurs are likely to track environmental variation with fine temporal resolution, with the added benefit of relatively precise generation times that can readily be translated parametrically in coalescent-based modeling methods. Some of these advantages for phylogeographic study in Madagascar have been noted by others (Schneider et al., 2010). Notably, some species of mouse lemur can occur in high densities in degraded forests and along forest ecotones and can thrive in areas with invasive fruiting trees. Thus, of all of the endemic Malagasy primates, they are perhaps most likely to be robust to human-mediated landscape modification (Gerard et al., 2015), although this ecological flexibility appears to be true for some, but not all, species (Schäffler et al., 2015). Despite their diminutive size, they have been observed to use terrestrial means for dispersal between trees.

Ongoing work from the past two decades has shown conclusively that although the morphological differences among species of mouse lemur can be subtle, their genetic and ecological differences are consistent with an evolutionary radiation that presently contains 24 described species (Schmid and Kappeler, 1994; Thalmann and Rakotoarison, 1994; Ganzhorn and Schmid, 1998; Zimmermann et al., 1998; Rasoloarison et al., 2000; Yoder et al., 2000; Lahann et al., 2006; Louis et al., 2006; Craul et al., 2007; Heckman et al., 2007; Weisrock et al., 2010; Rakotondranary and Ganzhorn, 2011; Rasoloarison et al., 2013; Hotaling et al., 2016), with the basal diversification occurring as long ago as 9–10 Ma (Yang and Yoder, 2003; Yoder and Yang, 2004). Several studies have shown strong support for three deep lineages, one that contains *M. murinus* plus *Microcebus griseorufus*, another deeply diverged lineage represented by *Microcebus ravelobensis*, *Microcebus danfossi*, and *Microcebus bongolavensis*, and a third lineage that comprises all other mouse lemur species, including strong support for a distal subclade containing *M. berthae*, *Microcebus rufus*, and *Microcebus myoxinus* (Yoder et al., 2000; Heckman et al., 2007; Weisrock et al., 2010, 2012). This latter subclade is especially intriguing, given that *M. rufus*, an eastern humid-forest animal, is markedly divergent both ecologically and geographically from *M. berthae* and *M. myoxinus*, both of which occur in western dry deciduous forests. The observation that this relatively recently diverged clade contains both eastern and western representatives has prompted speculation about the existence of former forested corridors between eastern and the western Madagascar (Yoder et al., 2000).

Our study aims to examine the fit of mouse lemur phylogeography to hypotheses regarding the natural plant community composition of Madagascar's Central Highlands. To test the fit to the three paleoenvironmental hypotheses described above, we focus on phylogenetic and geogenetic patterns in five species of mouse lemur: two with eastern distributions (*Microcebus mittermeieri* and *M. rufus*), two with western distributions (*M. myoxinus* and *M. berthae*), and one that has been proposed to be limited to the Central Highlands (*M. lehilahytsara*) (Radespiel et al., 2012). Phylogenetic analysis of two mtDNA loci, cytochrome *b* (*cytb*) and cytochrome oxidase II (*cox2*), was conducted to place the five targeted species within the larger evolutionary framework of the mouse lemur radiation. A genome-wide restriction site-associated DNA sequencing (RADseq) approach was used to assess genetic diversity among and within the five targeted species and to test the fit of these data to both spatial and historical predictions associated with the forest, grassland, and mosaic hypotheses. The genome-wide SNP data were analyzed with coalescent methods to estimate the species tree structure and its congruence with the mtDNA gene tree and to estimate divergence times for the five targeted species.

RESULTS AND DISCUSSION

mtDNA Tree

Maximum likelihood trees for concatenated *cytb* and *cox2* loci largely agree with previous analyses (Weisrock et al., 2010; Hotaling et al., 2016). Although the mtDNA analysis sampled only 17 of the 24 currently described mouse lemur species, the fundamental pattern found in previous studies is repeated here: Described species are resolved as reciprocally monophyletic, and the phylogeny shows a basal split between a clade formed by *M. murinus* plus *M. griseorufus* and all other species. Within the much more taxonomically diverse sister lineage, two species (*M. marohita* and *M. ravelobensis*) show long branches radiating from their shared ancestral node, suggesting that these species have long been diverged from other species within the sister lineage. The very short internal branches that separate the remaining 11 species within this lineage suggest that they experienced rapid radiation, making phylogenetic resolution highly problematic. The mtDNA gene tree illustrates the depth and diversity of the mouse lemur radiation through time and across the geographic expanse of Madagascar (Fig. 12.2). The clade containing the eastern *M. rufus* plus western *M. myoxinus* and *M. berthae* continues to receive robust support with the Central Highlands species *M. lehilahytsara* as its sister lineage, also in agreement with previous studies (Weisrock et al., 2010; Hotaling et al., 2016).

Species Tree Estimation and Divergence Times

The species tree was generated from double digest restriction site-associated DNA (ddRAD) libraries from 30 individuals across the five target species and a single outgroup, *M. marohita.* Paired-end 150-bp sequencing on two Illumina NextSeq runs yielded 557.3 million reads. After several quality-filtering steps (*Methods*) the dataset was trimmed to a total of 340.5 million reads across 29 samples for analysis with the software pyRAD (Eaton, 2014). pyRAD yielded 124,916 total loci after the removal of putative paralogs and implementation of filtering criteria. The 29 samples averaged 57,845 ± 13,901 loci. Among the four taxa with two or more sequenced samples, the average number of loci was 60,076 ± 10,387.

The species tree estimated with BPP (Yang, 2015) resulted in a posterior distribution with perfect (100%) statistical support for all nodes (Fig. 12.3). The tree is largely congruent with the mtDNA gene tree for the five targeted species (shown by a dashed box in Fig. 12.2), with a critical difference: *M. rnfus*, endemic to southeastern rainforest, is found to be the sister species to *M. berthae* from the western regions of Madagascar, south of the Tsiribihina River (node B in Fig. 12.3). *M. myoxinus*, also from

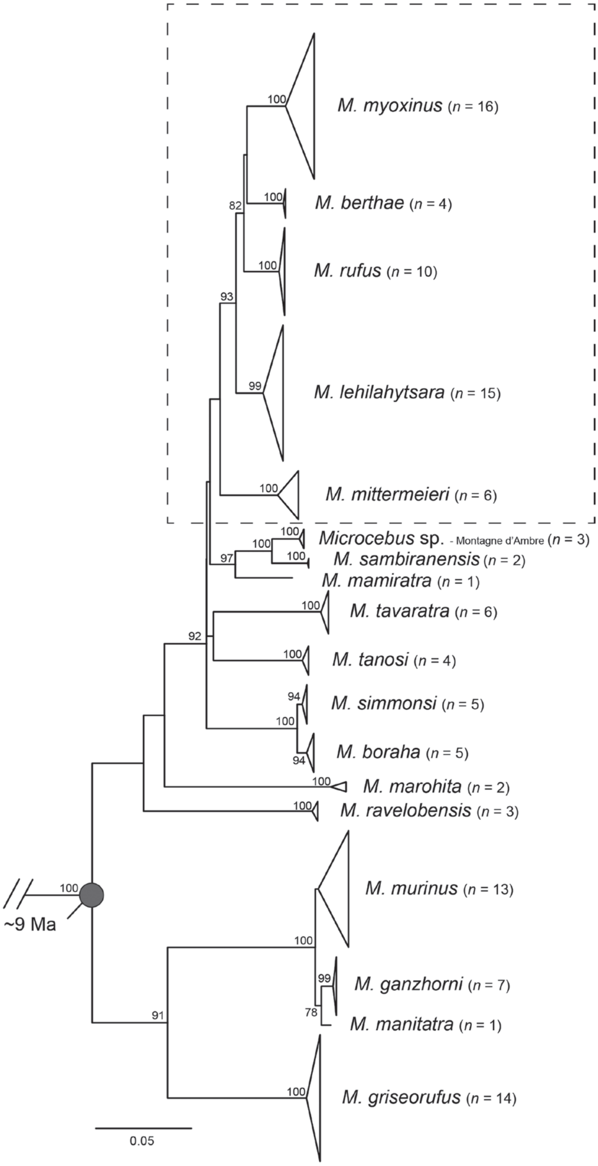


FIGURE 12.2 Maximum likelihood tree for concatenated mtDNA data (*cytb* and *cox2*) from 117 *Microcebus* sequences plus four outgroup sequences (*Cheirogaleus* and *Mirza*; not shown in figure). Bootstrap support is shown for internal nodes (100 replicates). The age of the basal node was previously estimated using phylogenetic methods by Yoder and Yang (2004). The dashed box highlights the five species targeted for genomewide SNP analysis.

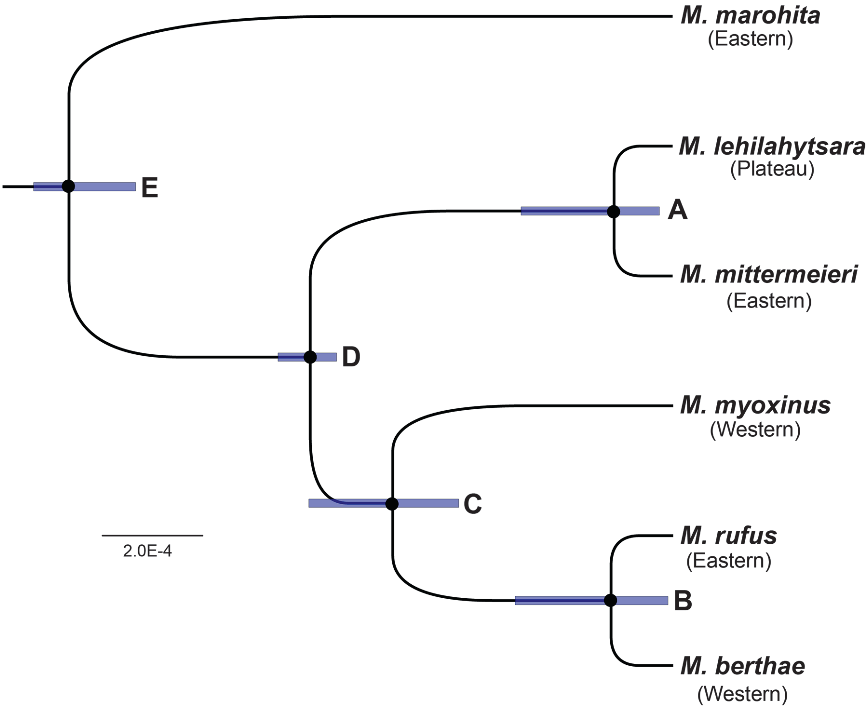


FIGURE 12.3 Maximum clade credibility species tree from a BPP analysis of nuclear RAD-seq data. Node labels correspond to those in Table 12.1. All nodes in the tree have 100% statistical support. Blue bars are the 95% highest posterior densities of node heights. Species are identified by their contemporary geographic range (eastern versus western); plateau, Central Highlands plateau.

western Madagascar, is placed as the sister lineage to the *M. berthae* plus *M. rufus* clade (node C in Fig. 12.3). The species trees inferred with SVD-Quartets and SNAPP (*SI Methods*[1]) are identical to the BPP tree, with both also yielding maximum support for all nodes. Using BPP, posterior branch lengths (τ) and nucleotide diversity ($\theta = 4N\mu$) were converted to geological times of divergence and effective population sizes by using priors on the

---

[1] Supporting information for this chapter, which includes additional *Methods* text, *Figures S1* through *S3*, and *Dataset SI*, is available online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1601081113/-/DCSupplemental. Reference numbers in the *SI Methods* text refer to the original PNAS article.

TABLE 12.1 Ancestral Effective Population Size (N_e) and Estimated Divergence Times

Node	N (× 10^3)	95% CI	Divergence Time, ka	95% CI
E	308	211, 450	539	346, 822
D	19.7	11.2, 31.2	324	207, 495
A	82.7	39.1, 13.3	52.5	10.6, 159
C	13.8	2.13, 34.7	250	147, 403
B	65	18.1, 11.9	55.4	4.56, 153

NOTE: Node labels correspond to those identified in Fig. 12.3. Results were derived using BPP (Yang, 2015).

per-generation mutation rate and the generation time. This analysis indicates that the ancestral node for the targeted species has a posterior mean age of ~540 ka (Table 12.1). The divergence times between *M. lehilahytsara* and *M. mittermeieri* (node A in Fig. 12.3) and between *M. berthae* plus *M. rufus* (node B in Fig. 12.3) are much more recent, with both divergences occurring ~55 ka [although note the considerable uncertainty represented by the 95% credible intervals (CIs)].

The age of the basal radiation of these five species is placed well within the Quaternary when climatic changes associated with glacial and interglacial periods would have radically and episodically desiccated much of the Central Highlands (Burney, 1986, 1987, 1988, 1997; Gasse et al., 1994). Thus, the results of our study would appear to agree with other work in which Quaternary climatic and vegetation changes have been invoked to explain patterns of diversity and speciation in different groups of lemurs (Wilmé et al., 2006) and in mouse lemurs in particular (Olivieri et al., 2008; Schneider et al., 2010).

Geogenetic Analysis

One of the intriguing results of this study was the discovery of a population of *M. lehilahytsara* from a tiny (<2 km^2) forest patch known as "Ankafobe," in close proximity to a series of fragmented forest parcels within the Ambohitantely protected area (Fig. 12.4). Although only three individuals from this locality were sampled, the level of genetic diversity among these animals at the mtDNA locus is greater than 1% and exceeds that of any other single locality from which this species has been sampled. Although geographic sampling is far too limited for definitive conclusions, such a pattern would be consistent with the hypothesis that remnant forest patches such as Ankafobe and Ambohitantely, although clearly impacted

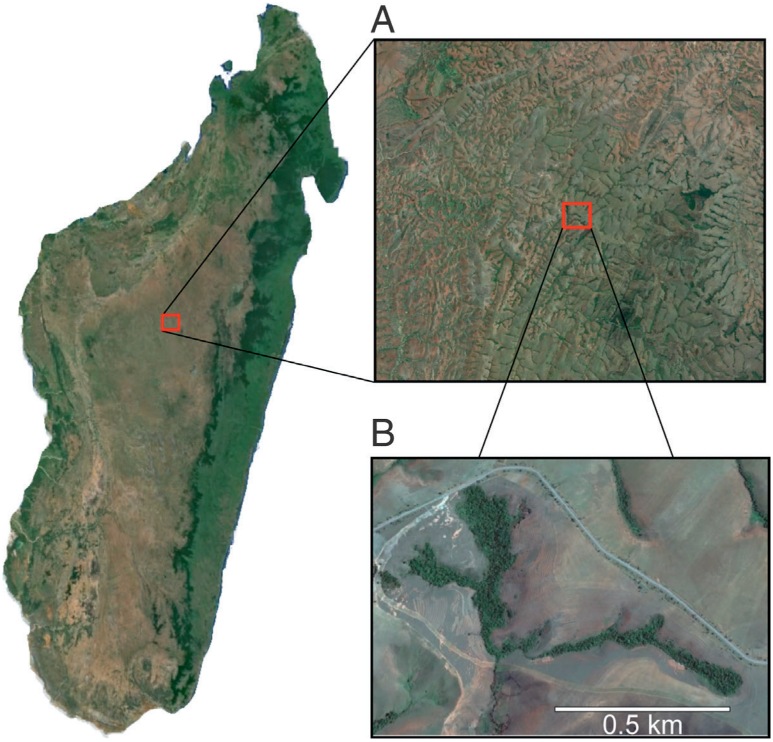


FIGURE 12.4 Google Earth view of the Ankafobe sampling locality for *M. lehilahytsara*. *Inset A* illustrates the depauperate environment that currently is comprised of anthropogenic grasslands. Areas of green are largely comprised of rice cultivation but suggest the potential for naturally occurring forests along these stream basins. *Inset B* illustrates the extremely isolated position of the Ankafobe Reserve.

by recent degradation caused by anthropogenic agency, may have experienced alternating episodes of connection, isolation, and reconnection with similar relict forests across the Central Highlands. Taking this hypothesis to its logical extreme, patterns of genetic diversity within and among these relictual habitats thus might share a signature of genetic diversification driven by long-term isolation but punctuated by periods of gene flow. These fragments thus might be described as "museums" of genetic diversity within an otherwise desolated grassland savanna.

Only unlinked SNPs with no missing data were used for geogenetic analysis using the software package SpaceMix (Bradburd et al., 2016). There were 1,583 SNPs without any missing data across the four taxa (*M. lehilahytsara*, *M. rufus*, *M. myoxinus*, and *M. mittermeieri*). However, when individuals from the same sampling location were consolidated and analyzed together, this number increased to 7,303. We used this latter dataset for all final SpaceMix analyses. As a potential means for testing the punctuated gene flow hypothesis, we compared patterns of genetic diversity within *M. lehilahytsara* to those within *M. myoxinus.* The two species have inferred ranges that are equivalent in geographic area (as measured in square kilometers), although in the case of *M. myoxinus*, there is the null expectation that ancestral habitat would have been largely continuous before anthropogenic fragmentation. The ancestral habitat for *M. lehilahytsara*, on the other hand, is unknown and dependent on the inferred conditions consequent to the forest, grassland, and mosaic hypotheses. By comparing genetic structure in a geographic range among contiguous forests (*M. myoxinus*) equivalent in area (as measured by square kilometers) with a geographic range among mosaic forests within the Central Highland species (*M. lehilahytsara*), we see that the two species show significantly different patterns of genetic distance relative to geographic distance (Fig. 12.5). The *t*-tests yielded statistically significant differences in genetic distance in paired long-distance dyads (average $P = 9.80 \times 10^{-4}$ across the 1,000 randomly sampled *t*-tests), paired short-distance dyads (average $P = 8.59 \times 10^{-3}$), and when considering long and short dyads together (average $P = 1.34 \times 10^{-4}$) (*Figs. S1–S3*). At both short and long geographic distances, the intraspecific genetic distance within *M. myoxinus* was significantly lower than that of the closely related *M. lehilahytsara.* Given that *M. myoxinus* and *M. lehilahytsara* presumably inhabited their respective habitats for millennia before the arrival of humans, the starkly higher genetic distances relative to geographic distances in *M. lehilahytsara* suggests that this species has long lived in a mosaic environment.

The geogenetic analysis yields robust support for both the phylogenetic and the genetic distance analyses (Fig. 12.6). SpaceMix represents the geogenetic locations of the samples (the colored haloes) relative to their true geographic location (the solid dots), with arrows pointing in the direction of hypothesized genetic history and admixture. Each pair of sampling and geogenetic locations represents a population, summarized by allele count, and the halo corresponds to a 95% confidence interval of the spatial location of the SNP profile. The size of the halo decreases with the number of individuals in the population.

There are notable patterns regarding ancestral connections between the east and the west. The first, and probably most obvious, pattern is

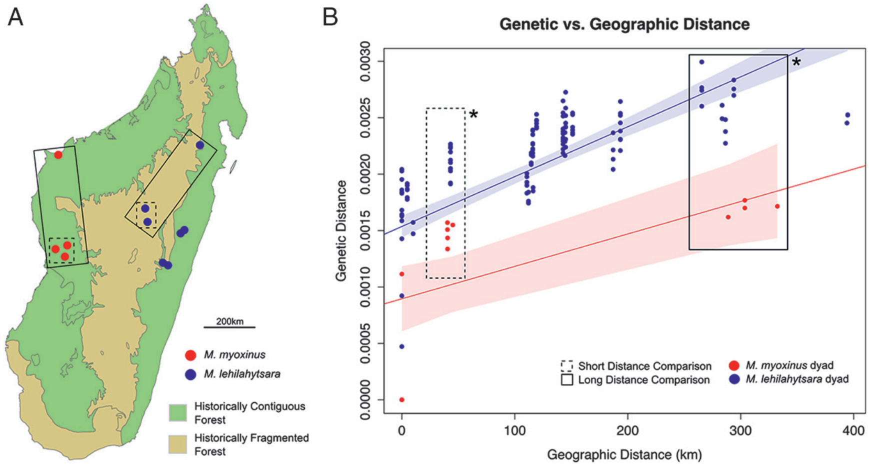


FIGURE 12.5 (*A*) A map of Madagascar with sampling locations for two focal species, *M. myoxinus* and *M. lehilahytsara*, hypothesized historical forest type labeled by color, and boxes marking the sampling locations used for short and long geographic distance *t*-tests of genetic distance. (*B*) Plot of genetic distance versus geographic distance in within-species dyads, colored as on the sampling location map. The dashed and solid boxes on both the distance plot and sampling map illustrate the short and long geographic distance comparisons of genetic distance being made. The *t*-tests within each box are significantly significant at the $P < 0.01$ level (*Figs. S1–S3*). The line of best fit and shaded 95% confidence interval, plotted with R, are matched to the species by color.

the common geogenetic space shared by *M. rufus* (from the east) and *M. berthae* (from the west), consistent with the species tree in which they are shown to be sister lineages. Although the actual geographic coordinates of *M. rufus* fall well into the southeastern forests of Madagascar, the genomic signature recovered from the ddRAD data place this species squarely in the center of the island, forming a genetic "bridge" between *M. lehilahytsara* and *M. myoxinus.* This result is in contradiction to the expectation that western neighbors *M. myoxinus* and *M. berthae* should share a more recent ancestry, given their proximity relative to the geographically distant *M. rufus.* Instead, there is a strong separation of *M. myoxinus* and *M. berthae*, consistent with the biogeographic separation of these species by the Tsiribihina River, suggesting that the river is a significant barrier

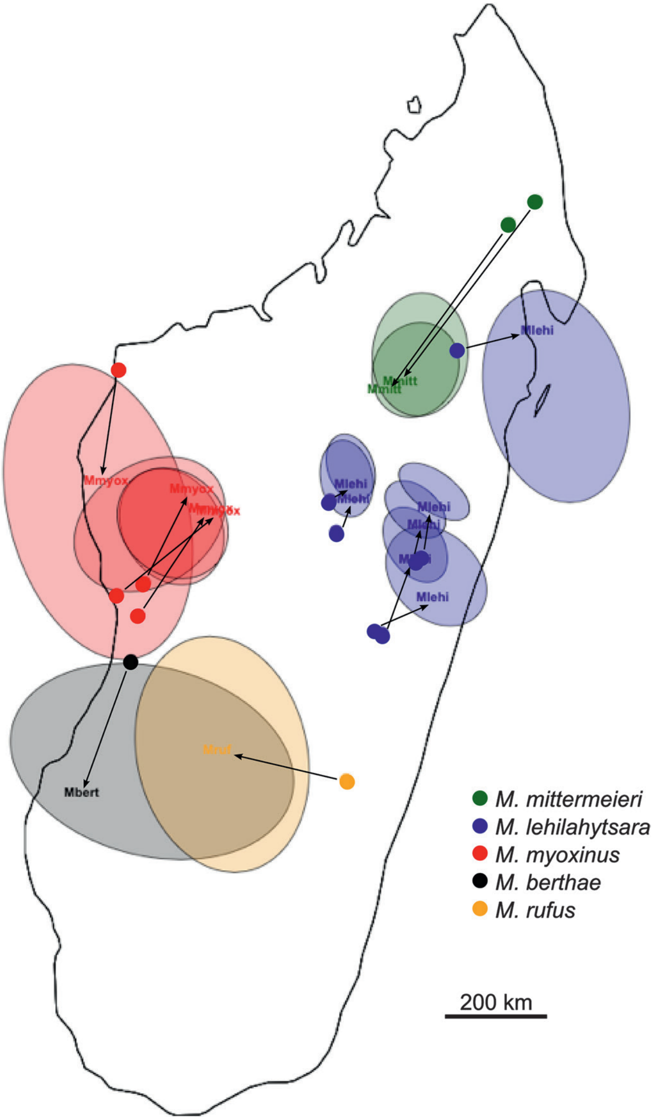


FIGURE 12.6 SpaceMix-inferred geogenetic locations of samples based on prior of true sampling site and population consolidated SNP data. Abbreviated names indicate the geogenetic location, and ellipses represent 95% confidence interval of geogenetic location. The true sampling site is labeled with a colored dot, and arrows indicate direction of geographic pull via admixture.

to gene flow. This biogeographic barrier is further emphasized given that the *M. myoxinus* dots are pulled northward and *M. berthae* dots pushed southward, opposite of what would be expected in the presence of gene flow between these two species.

There is also a strong geogenetic signal regarding the inference of ancestral habitat type in the Central Highlands. Although the full set of *M. myoxinus* samples (four populations) are all pulled to a central location, and all four geogenetic confidence intervals overlap, the opposite is true for *M. lehilahytsara*. In that species, measured over a similar geographic range, the populations do not share any common geogenetic space. Even when the northern *M. lehilahytsara* population is excluded, the six southernmost populations occupy two distinct geogenetic spaces. This finding is in robust agreement with the results of the genetic vs. geographic distance plots, indicating that a substantial intraspecific genetic structure is built up in *M. lehilahytsara* relative to *M. myoxinus*. Moreover, the shift of *M. mittermeieri* from the northeastern corner of the island toward the south and into the geographic range of *M. lehilahytsara* is a further indication that this putative plateau specialist displays remarkable levels of genetic diversity, despite its fragmented distribution.

CONCLUSIONS

The results of this study have specific implications regarding the timing and geographic patterns of divergence events among the five species of *Microcebus* targeted for phylogeographic analysis, as well as more general implications for the inferred habitat of the Central Highlands during the period of their diversification. Species-tree analysis of a robust genomewide SNP dataset, using three independent analytic toolkits (BPP, SNAPP, and SVDQuartets) shows unequivocal support for a rapid radiation of a large ancestral population (node E in Table 12.1 and Fig. 12.3) into five geographically separated species. This diversification is estimated to have commenced at ~540 ka, with the two most terminal divergence events occurring at ~55 ka (see nodes A and B in Fig. 12.3). The hypothesized timing of these events is coincident with paleoclimatological patterns in the African and Indian Ocean regions that suggest tremendous climatological variation (Cannon et al., 2009; Scholz et al., 2011; Blome et al., 2012). Even so, the inferred ages are more recent than might have been predicted, given previous analyses of divergence times that used phylogenetic methods (Yang and Yoder, 2003; Yoder and Yang, 2004).

What could not have been predicted is the inferred paleogeographic positioning of these five species, closely related in geological time but broadly dispersed geographically. The most surprising pattern is the close phylogenetic relationship between *M. rufus* and *M. berthae* (node B in

Fig. 12.3). *M. rufus* is native to the humid forests of the southeast, whereas *M. berthae* occupies a restricted range within the dry forests of western Madagascar. Thus, these species are presently separated by the Rubicon of the Central Highlands, yielding the preconception that longitudinal dispersal should have been impossible. It is here that a consideration of the phylogenetic and geogenetic patterns among the targeted species allows for discrimination among the forest, grassland, and mosaic hypotheses. If the forest hypothesis is applicable, we would not expect to see a divergence that is tens of thousands of years old. Rather, the forest hypothesis would predict ease of geographic dispersal until historic times. Conversely, although the grassland hypothesis would predict the separation of eastern and western species, it should do so uniformly across all species and with the expectation of divergence ages on the order of millions of years old rather than the tens of thousands of years found in this study.

The totality of evidence instead points to the mosaic hypothesis, with *M. lehilahytsara* as something of a telltale species, retaining the signal of a mixed environment of forests and grasslands, perhaps experiencing cyclic episodes of isolation and reconnection. The timescaled phylogeny indicates that although longitudinal dispersal would have been relatively uncomplicated throughout the Pleistocene, ease of dispersal was apparently arrested around 53–55 ka, a time of climate change in the African region (Scholz et al., 2011). In all likelihood, Holocene aridification and rewarming would have promoted natural fragmentation into a mosaic of grasslands and interconnecting woodlands. In fact, the results of this study are congruent, both temporally and geogenetically, with those of Quéméré et al. (2012), who observed similar genetic symptoms of long-term habitat fragmentation in the golden-crowned sifaka (*Propithecus tattersalli*) in northern Madagascar. We therefore conclude that although the longitudinal dispersal of mouse lemurs across Madagascar was achieved with relative ease throughout the Pleistocene, habitat fragmentation via natural processes was established in the Holocene, yielding the contemporary distribution of Madagascar's mouse lemurs.

METHODS

Sample Collection

Samples analyzed in this study consisted of 318 samples from previous studies and 22 samples from previously unsampled populations of *M. lehilahytsara* and *M. mittermeieri* collected by M.B.B. (*Dataset S1*). Mitochondrial DNA sequencing was conducted to compare newly collected samples with existing data. The targeted species analysis consisted of ddRAD sequencing of 30 wild-caught mouse lemurs from 15 study sites and are largely from five closely related species (*M. berthae*, *M. myoxinus*, *M. rufus*, *M. lehilahytsara*, and *M. mittermeieri*) and one outgroup (*M. marohita*). Nine of the ddRAD-sequenced samples were newly collected; the remaining 21 samples were analyzed, in part, in previous work (Yoder et al., 2000; Weisrock et al., 2010).

mtDNA Sequencing

Sequence data were collected from 18 newly collected samples and four previously unsequenced samples for both the *cytb* and *cox2* genes. Approximately 1,140 bp of *cytb* were amplified using the following primers: TGA-YTA-ATG-AYA-TGA-AAA-AYC-ATC-GTT-G and TCT-CCA-TTT-CTG-GTT-TAC-AAG-ACC-A. Approximately 684 bp of *cox2* were amplified using primers L7553 and H8320 (Adkins and Honeycutt, 1994). Cycle sequencing reactions were performed using BigDye Terminator V1.1 following standard conditions, and all samples were sequenced using an Applied Biosystems 3730xl at the Duke Sequencing Core resource. Sequences have been submitted to the National Center for Biotechnology Information (NCBI) under Accession Nos. KX070700–KX070743.

mtDNA Analyses

The full dataset consists of *cytb* and *cox2* sequences for 332 individuals from 26 described species and three undescribed putative species within Cheirogaleidae. These sequences include 318 individuals from previous studies and new data from 14 individuals sequenced for this study. Analyses were conducted using PAUP* version 4.0a149 except where otherwise indicated. Models chosen by corrected Akaike information criteria (AICc) (Burnham and Anderson, 2002) for the three codon positions were SYM+I+G, HKY+I+G, and TrN+I+G, respectively. All selected models include an invariable-sites category with variable sites drawn from a γ distribution (+I+G). Bootstrapping was performed using RAxMLversion 8.2.4 (Stamatakis, 2014) using the closest available model (GTR+I+G, with partitioning by codon position). Bootstrap results were transferred to the optimal PAUP* trees using the SumTrees command available in DendroPy version 4.1.0 (Sukumaran and Holder, 2010). (Additional details are given in *SI Methods*.)

RAD Genotyping

ddRAD libraries were generated from whole genomic DNA and whole genome amplified DNA following the protocol of Peterson et al. (2012) and Blair et al. (2015). The double digest was completed with the enzymes SphI and MluCI, and IDT primers were used to barcode uniquely the first paired end of all 30 samples. We used Stacks (Catchen et al., 2013) to demultiplex reads by unique barcode. Given the overall poor quality of the second paired-end read, we dropped those reads from the analysis and used only the first read. We used the program pyRAD (Eaton, 2014) to cluster, align, and call SNPs within the curated ddRAD dataset. All sequence reads used for analysis are available at the NCBI Short Read Archive (study Accession No. SRP073945). (Additional details are given in *SI Methods*.)

RAD Data Analyses

For the assessment of genetic distances, the R package APE (Paradis et al., 2004) was used, taking as input the concatenated, aligned ddRAD loci. The R package SpaceMix (Bradburd et al., 2016) was run to determine the samples' geogenetic positions with 95% confidence intervals. SpaceMix was run both for all 29 individuals, yielding 1,583 SNPs, and for all 15 sample sites separately, in which we collapsed the locality data to a single entry in the SpaceMix input matrix, yielding 7,303 SNPs. All SpaceMix analyses were run without any missing data.

Species Tree and Divergence Time Analyses

The program BPP (Yang, 2015), which implements the multispecies coalescent, was used to estimate the species tree topology, branch lengths (τ), and nucleotide diversity ($\theta = 4N\mu$) using whole ddRAD loci (as opposed to SNP data). The posterior of τ and θ can be converted to geological times of divergence and effective population sizes by using priors on the per-generation mutation rate and the generation time (Angelis and dos Reis, 2015). Bayesian analysis using the multispecies coalescent is computationally expensive, so here we used a small dataset to estimate the tree topology and a large dataset to estimate τ and θ values more precisely by fixing the topology to that obtained with the small dataset. Tree topology is identical to that derived using SNAPP and SVDQuartets (*SI Methods*).

Small Dataset Analysis

The ddRAD loci (82 loci; 11,624 bp) with data for all 30 individuals were analyzed with BPP to obtain the tree topology (BPP's A01 analysis). The species tree prior was set to uniform rooted trees. The prior for τ is Gamma (2, 250) and for θ is Gamma (2, 1,000). The posterior species tree obtained was identical to that estimated with SNAPP.

Large Dataset Analysis

The RAD-seq fragments (80,662 fragments; 11,247,917 bp) were analyzed for a subset of one individual from each species from our focal ingroup. The use of many sites leads to narrower (more precise) posterior estimates of τ and θ, whereas the use of few individuals reduces the amount of time needed for Bayesian Markov chain Monte Carlo computation. The priors on τ and θ were as for the small dataset. To convert τ and θ values to divergence times (t) and population sizes (N), we sampled values for the mutation rate (μ) and generation time (g) from priors and used these values to calculate posterior estimates of t and N [see Angelis and dos Reis (2015) for details]. The prior on μ was Gamma (27.80, 31.96), which roughly has a 95% prior credible interval (CI) between 0.5 and 1.2×10^{-8} substitutions per site per generation. This result matches NextGen estimates of the per-generation mutation rate in the laboratory mouse (Uchimura et al., 2015) and human (Scally and Durbin, 2012). The prior on g was Gamma (100, 26.6), which roughly has a 95% prior CI between 3 and 4.5 years.

Estimation of Average Generation Time for Genus *Microcebus*

We estimated an average generation time for all species of mouse lemurs using a combination of data from both wild and captive populations, yielding an informed estimate of 3.0–4.5 years. Field observations indicate that for eastern humid-forest species, age at first reproduction in wild populations is 10–12 months, with females undergoing estrus and reproduction annually thereafter. Two litters per year can be common for some species and wild populations but is rare in other species and localities (Lahann et al., 2006; Blanco et al., 2015). Captive data for *M. murinus* at the Duke Lemur Center show that two litters in a single year has occurred only once in 101 litters. Extrapolating from figure 2A in Zohdy et al. (2014), we calculated a survival probability of 0.879–0.0714 * (age in years) that, when extrapolated to 12 years of age, yields 17% for 10 years, 9% for 11 years, 2% for 12 years, and 0% above 12 years. These figures are consistent with observations from animals bred in captivity. Survival probabilities are considerably lower for wild populations, however. In the dry forests of western Madagascar, individuals appear to live only 2–3 years on average (Hämälainen et al., 2014). Based on capture/recaptures at more resource-rich habitats, such as the eastern humid forests, indications are that lifespans are likely longer, with 4 years being a conservative average lifespan. Information from both captive and wild populations indicates that reproductive senescence does not show major effects until 5 years of age in females; thus, it is likely that the period of reproductive fitness exceeds lifespan in natural populations.

Comparison of Genetic and Geographic Distance

The genetic distance (RAD Data Analyses) was compared between populations of similar geographic distances in *M. myoxinus* and *M. lehilahytsara*. Individuals were assessed at every possible within-species dyad, and the resulting genetic and geographic distances were plotted to inspect visually for isolation by distance (IBD). Given the relatively low sample count of *M. myoxinus* compared with *M. lehilahytsara*, the latter was randomly sampled to conduct a paired, one-sided *t*-test. Paired short- and long-distance comparisons were tested individually and together for a total of three separate *t*-tests, each sampling 1,000 times from the measured *M. lehilahytsara* dyads (*Figs. S1–S3*).

ACKNOWLEDGMENTS

We thank the Organizers of the Symposium, Francisco Ayala and John Avise, and the invited speakers for an inspiring synthesis of ideas and methods. This paper is very much a reflection of their contributions to the field of phylogeography. We thank the Malagasy authorities for permission to conduct this research and Chris Birkinshaw for suggesting the importance of Ankafobe as an ecological isolate to M.B.B. This study was funded by a grant from the Duke Tropical Conservation Initiative (to A.D.Y.) and by National Science Foundation Grant DEB-1354610 (to D.W.W. and A.D.Y.). This is Duke Lemur Center Publication No. 1322.

All sequence reads used for analysis are available at the National Center for Biotechnology Information (NCBI) Short Read Archive (Study Accession No. SRP073945). Sequences have been submitted to NCBI (Accession Nos. KX070700–KX070743).

13

Bison Phylogeography Constrains Dispersal and Viability of the Ice-Free Corridor in Western Canada

PETER D. HEINTZMAN,[*] DUANE FROESE[†§§§]
JOHN W. IVES,[‡] ANDRÉ E. R. SOARES,[*]
GRANT D. ZAZULA,[§] BRANDON LETTS,[‖]
THOMAS D. ANDREWS,[#] JONATHAN C. DRIVER,[**]
ELIZABETH HALL,[§] P. GREGORY HARE,[††]
CHRISTOPHER N. JASS,[‡‡] GLEN MACKAY,[#]
JOHN R. SOUTHON,[§§] MATHIAS STILLER,[‖‖]
ROBIN WOYWITKA,[†] MARC A. SUCHARD,[##***†††] AND
BETH SHAPIRO[*‡‡‡§§§]

---

[*] Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95064; [†] Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, AB T6G 2E3, Canada; [‡] Institute of Prairie Archaeology, Department of Anthropology, University of Alberta, Edmonton, AB T6G 2H4, Canada; [§] Yukon Palaeontology Program, Government of Yukon, Whitehorse, YT Y1A 2C6, Canada; [‖] Department of Biology, The Pennsylvania State University, University Park, PA 16802; [#] Prince of Wales Northern Heritage Centre, Government of the Northwest Territories, Yellowknife, NT X1A 2L9, Canada; [**] Department of Archaeology, Simon Fraser University, Burnaby, BC V5A 1S6, Canada; [††] Yukon Archaeology Program, Government of Yukon, Whitehorse, YT Y1A 2C6, Canada; [‡‡] Royal Alberta Museum, Edmonton, AB T5N 0M6, Canada; [§§] Keck-CCAMS Group, Earth System Science Department, University of California, Irvine, CA 92697; [‖‖] Department of Translational Skin Cancer Research, German Consortium for Translational Cancer Research, D-45141 Essen, Germany; [##] Department of Biomathematics, David Geffen School of Medicine at University of California, Los Angeles, CA 90095; [***] Department of Human Genetics, David Geffen School of Medicine at University of California, Los Angeles, CA 90095; [†††] Department of Biostatistics, University of California, Los Angeles Fielding School of Public Health, Los Angeles, CA 90095; and [‡‡‡] University of California Santa Cruz Genomics Institute, University of California, Santa Cruz, CA 95064. [§§§] To whom correspondence may be addressed. Email: duane.froese@ualberta.ca or bashapir@ucsc.edu.

The Ice-Free Corridor has been invoked as a route for Pleistocene human and animal dispersals between eastern Beringia and more southerly areas of North America. Despite the significance of the corridor, there are limited data for when and how this corridor was used. Hypothetical uses of the corridor include the first expansion of humans from Beringia into the Americas, northward postglacial expansions of fluted point technologies into Beringia, and continued use of the corridor as a contact route between the north and south. Here, we use radiocarbon dates and ancient mitochondrial DNA from late Pleistocene bison fossils to determine the chronology for when the corridor was open and viable for biotic dispersals. The corridor was closed after ~23,000 until 13,400 calendar years ago (cal y BP), after which we find the first evidence, to our knowledge, that bison used this route to disperse from the south, and by 13,000 years from the north. Our chronology supports a habitable and traversable corridor by at least 13,000 cal y BP, just before the first appearance of Clovis technology in interior North America, and indicates that the corridor would not have been available for significantly earlier southward human dispersal. Following the opening of the corridor, multiple dispersals of human groups between Beringia and interior North America may have continued throughout the latest Pleistocene and early Holocene. Our results highlight the utility of phylogeographic analyses to test hypotheses about paleoecological history and the viability of dispersal routes over time.

A central question in New World biogeography and archaeology has been the role of an "Ice-Free Corridor" along the eastern slopes of the Rocky Mountains in facilitating biotic exchange between Beringia—unglaciated Alaska and Yukon—and southern interior parts of the Americas (Johnston, 1933; Mandryk, 1996; Mandryk et al., 2001; Ives et al., 2013). Of central importance is the potential role for the corridor between the Laurentide and Cordilleran ice sheets in defining the pattern and timing of the initial entry of humans into the New World (Fig. 13.1). Early geological models from western Canada suggested that the ice sheets did not coalesce during the Last Glacial Maximum (LGM) (Stalker, 1977) and implied that the corridor was a viable route for dispersal throughout the late Pleistocene (~125,000–11,500 years ago). This view gained support from numerous, misleadingly "old" radiocarbon ages from bulk deposits in the corridor region (MacDonald et al., 1987). The suggestion of an open and viable corridor throughout the LGM gave rise to its prominence as the most likely pathway taken by the first people to colonize the Americas from Beringia (Fladmark, 1986; Mandryk, 1996; Mandryk et al., 2001). Eventually, this notion of a corridor became tightly linked with the "Clovis First" archaeological model, in which widespread fluted projectile point technology, often found in association with remains

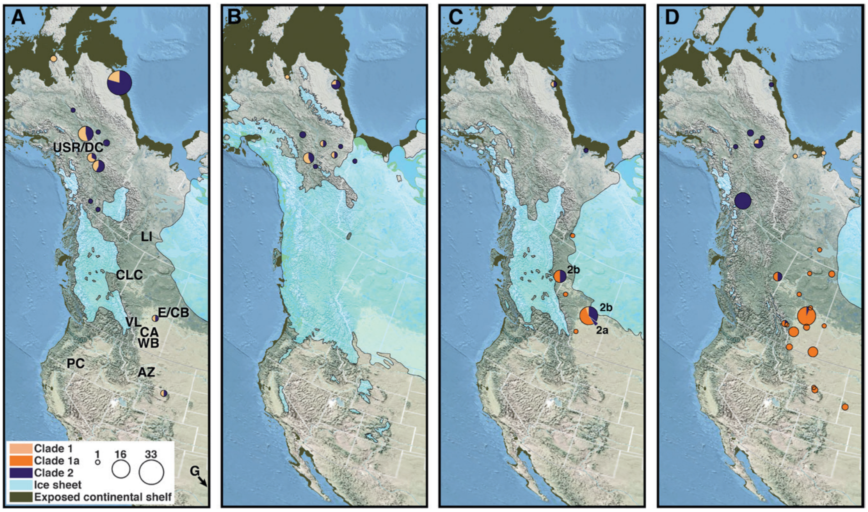


FIGURE 13.1 The geographic distribution and clade assignment of 192 radiocarbon-dated or present-day North American bison. Bison distribution (*A*) >23,000 cal y BP, before the period of glacial coalescence; (*B*) 23,000–13,500 cal y BP, during the period of glacial coalescence; (*C*) 13,500–12,000 cal y BP, as the corridor opens; and (*D*) <12,000 cal y BP. Bison mitochondrial sequences fall into two main genetic clades (Shapiro et al., 2004): clade 1a arises in the south during the interval depicted in *B*. Ice margins are from Dyke (2004): (*A*) pre-LGM margin at 23,000 cal y BP estimated following discussion of probable marine isotope stage 3 margin in Dyke (2004); (*B*) 17,900 cal y BP; (*C*) 13,000 cal y BP; (*D*) 10,700 cal y BP. Sea levels are from Hu et al. (2010): (*A*) –55 m; (*B*) –100 m; (*C*) –65 m; and (*D*) –40 m. Specific localities: Anzick (AZ), Calgary (CA), Charlie Lake Cave (CLC), Dry Creek (DC), Edmonton/Clover Bar (E/CB), Gault (G), Liard River (LI), Paisley Caves (PC), Upward Sun River (USR), Vermilion Lakes (VL), and Wally's Beach (WB). SOURCES: Map modified from ESRI ArcGIS Online, U.S. National Park Service, and U.S. Geological Survey, east view cartographic.

of Pleistocene megafauna, was believed to reflect initial human dispersal into the Americas (Goebel et al., 2008).

More recent geological interpretations indicate that the Laurentide Ice Sheet coalesced with the Cordilleran Ice Sheet along the eastern foothills of the Rocky Mountains by ~21,000 calendar years before present (cal y BP, with present defined as CE 1950) (Jackson et al., 1997). Ice sheet coalescence would have formed an impenetrable barrier to terrestrial dispersals between Beringia to the north and interior North America to the south that lasted at least until glaciers retreated near the end of the Pleistocene. As the ice sheets receded, a corridor formed that opened a route between Beringia and the continental interior, with its most likely route running through Yukon Territory along the rapidly wasting Cordilleran Ice Sheet or perhaps along the Mackenzie Valley (Ives et al., 2013). A gap in the regional vertebrate radiocarbon date record in Alberta (Burns, 2010) and genetic evidence of a strong barrier to gene flow within large mammal populations to the north and south of the ice sheets during this time interval (Shapiro et al., 2004) both support the geological interpretation of coalescence of the Laurentide and Cordilleran ice sheets.

As geological and paleoenvironmental evidence mounted that the corridor was not available during the LGM, a hypothesized Pacific coastal route emerged as the more likely route to account for the earliest people in southerly regions of the Americas (Mandryk et al., 2001; Amick, 2016). This hypothesis was coupled with the suggestion that, even if an early deglacial corridor existed, the landscape was probably ecologically unproductive and unable to support large mammal or human populations, perhaps for millennia following the retreat of glaciers (Mandryk et al., 2001). The Clovis First model lost favor as sites south of the LGM ice sheets, such as Monte Verde (Chile), the Gault locality (Texas), Manis (Washington), and Paisley (Oregon), provided archaeological evidence that humans were present in the Americas at least a millennium before the appearance of Clovis technology (Goebel et al., 2013). At the same time, genetic analyses suggested that New World founding populations began dispersing from Beringia ~15,000–20,000 cal y BP (Tamm et al., 2007; Achilli et al., 2013; Raghavan et al., 2015).

Despite its importance in North American paleoecology and archaeology, interpretations of the corridor are based on data from a sparsely studied, vast region that was impacted significantly by glacial processes. Questions remain about the precise timing of the postglacial opening of the corridor (Amick, 2016), its biological carrying capacity following deglaciation (Mandryk et al., 2001), and when, if at all, it could have served as a viable route for the movement of people and animals between Beringia and the rest of the Americas. Available archaeological data demonstrate that human populations were present near the southern end of the corridor just before deglaciation (Ives et al., 2013; Waters et al., 2015). Unfortunately, the limited and often perplexing archaeological evidence from the corridor region itself makes it difficult to assess directly when people, technologies, or both traversed the corridor after deglaciation (Ives et al., 2013). More recently, archaeological interest in the corridor has focused on its potential as a route for northward postglacial dispersal, most notably to explain the appearance of fluted projectile point technology in Alaska (Goebel et al., 2013).

BISON PHYLOGEOGRAPHY TO INFER THE CHRONOLOGY OF THE ICE-FREE CORRIDOR

Advances in paleogenetic approaches and DNA sequencing technologies have enabled the recovery of large datasets of populations sampled over time (Shapiro and Hofreiter, 2014). Although DNA preservation and the chronological range of radiocarbon dating have limited ancient DNA studies largely to the last ~50,000 years, this temporal interval spans two events that were undeniably important in shaping the present-day distribution of biodiversity: the transition into and back out of the LGM, which reached its peak ~26,000–19,000 cal y BP (Clark et al., 2009), and the global expansion of human populations (Koch and Barnosky, 2006). Although some genetic signal of both of these events is retained in living populations, paleogenetic data provide temporal snapshots of phylogeographic structure that may not be otherwise detectable (Ramakrishnan and Hadly, 2009). For example, paleogenomic data from ancient Europeans have revealed present-day Sardinians to be more closely related to the ancestral human population in Europe than to any other present-day population, which probably reflects a genetic replacement associated with the expansion of agriculture (Lazaridis et al., 2014; Skoglund et al., 2014). Paleogenetic data have also been used to test hypotheses about the influence of humans on the phylogeography of other species, for example, via domestication (Larson et al., 2007a; Kistler et al., 2014).

Bison (*Bison* sp.) are an ideal taxon for assessing when the corridor was open and available for biotic dispersals. Although bison taxonomy is complex, with up to several dozen species named based largely on skull and horn morphology (Guthrie, 1990; Wilson, 1996), the first bison to enter North America from Asia during the Pleistocene are generally referred to as the steppe bison, *Bison priscus.* Bison are one of the most abundant large mammals recovered from faunas within the western interior of North America, although previous studies have provided only limited radiocarbon and mitochondrial data from the corridor region. Bison are one of the first species for which ancient DNA data were used to reconstruct changing patterns of population structure spanning tens of thousands of years at continental scales (Shapiro et al., 2004). In addition, unlike many of their Pleistocene counterparts such as mammoths (*Mammuthus*), horses (*Equus*), and lions (*Panthera*), bison survived the extinction event at the end of the Pleistocene and persist in North America. Bison genomes, however, contain the signature of a genetic bottleneck at the end of the Pleistocene that can be used to test hypotheses about how ice age environmental changes affected their distribution and abundance (Shapiro et al., 2004).

Previous work using radiocarbon and ancient DNA data demonstrated that bison mitochondrial genetic diversity (but not necessarily their morphological diversity) (Wilson, 1996) is phylogeographically partitioned across North America and has been since the LGM (Shapiro et al., 2004). Although bison survived the interval of glacial coalescence both north and south of the continental ice sheets, population bottlenecks and barriers to gene flow affected their mitochondrial diversity. By the time the glaciers began to retreat, bison populations that had been isolated to the south of the continental ice sheets were mitochondrially distinct from their contemporary northern counterparts in Beringia. A mitochondrial clade of southern bison, including the two present-day bison subspecies in North America, the plains bison (*B. bison bison*) and wood bison (*B. b. athabascae*), shares a common ancestor dating to the period of glacial coalescence ~15,000–22,000 cal y BP (Shapiro et al., 2004). Thus, the identification of bison from this southern clade within and north of the corridor region can be interpreted as reflecting northward dispersal. Likewise, the appearance of bison from a Beringian mitochondrial clade further south in interior North America can be interpreted as southward dispersal.

Here, we report new radiocarbon dates and mitochondrial haplotype data from fossil bison recovered from the corridor region and adjacent areas in Beringia (Fig. 13.1, Table 13.1, and *Dataset S1*[1]) and use these data to assess when the corridor opened and was available as a dispersal route. In addition to resolving the chronology for when the corridor was open, we use a Bayesian phylogeographic approach to infer the timing and directionality of dispersals for ancient bison between the north and south. The specific evolutionary history of bison mitochondria makes this an ideal genetic marker to assess the timing and viability of the corridor and thereby provides insights relevant not only to bison, but also to other species, including humans, that may have used the corridor to disperse between Beringia and interior North America during the late Pleistocene.

RESULTS AND DISCUSSION

A Refined Chronology for the Opening of the Postglacial Corridor

To infer the chronology of the corridor linking Beringia and interior North America, we generated radiocarbon dates from 78 North American bison fossils, 49 of which were recovered from the corridor region (Table 13.1 and *Dataset S1*). Sites included Clover Bar and Charlie Lake Cave in western Canada (Fig. 13.1), both of which previously yielded dates in the time frame of interest (11,500–13,500 cal y BP) (Burns, 2010). The Charlie Lake Cave fossils were also associated with archaeological materi-

---

[1] Supporting information for this chapter, which includes *Datasets S1* and *S2*, *Fig. S1*, and *Table S1*, is available online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1601077113/-/DCSupplemental.

als (Driver et al., 1996). We generated mitochondrial haplotypes from 45 of these dated fossils, including from 18 of the 22 from Clover Bar and Charlie Lake Cave that fall within the time frame of interest. We then used these and previously published haplotypes to estimate a mitochondrial genealogy for a total of 192 late Pleistocene, Holocene, and present-day North American bison, including 37 from the corridor region and within the time frame of interest (Fig. 13.1, *Fig. S1*, Table 13.1, and *Dataset S1*). To facilitate discussion, we divide the mitochondrial genealogy into two major clades, clade 1 and clade 2, which are based on the most deeply diverging lineages within the mitochondrial tree. We also highlight several well-supported subclades within these two major clades.

Our recovered genealogy is similar to previously published mitochondrial genealogies for bison (Shapiro et al., 2004; Zazula et al., 2009), in which the most striking feature of the tree is the clustering of all present-day bison into clade 1a, with a maternal common ancestor that postdates the LGM (Fig. 13.2*A*). The increased density of new postglacial bison included in our analysis refines the pattern of extinction of clade 2, within which the latest surviving lineages tend to cluster together both phylogenetically and geographically (Fig. 13.2*A*). Clade 2a comprises a late-surviving population of bison that appears to be geographically restricted to southern Yukon and interior Alaska and includes a fossil that dates to as recently as ~325–490 cal y BP. Clade 2b is geographically isolated to the region of the postglacial corridor and includes a skull recovered from Banff National Park that dates to ~305–430 cal y BP (*Dataset S1*). This increased phylogeographic clustering during postglacial times, compared with before the LGM, probably reflects a trend toward physical isolation of bison populations as open habitats were largely replaced by spruce forest and increasing paludification across northwestern Canada (Dyke, 2005).

The pattern of bison dispersal into the postglacial corridor provides increased resolution of both the timing of establishment of an ecosystem that can support grazing herbivores and the capacity for the animals to traverse this region. Previously published dates for a horse from Vauxhall, Alberta (Burns, 2010), and a bison from Tsiigehtchic, Yukon (Zazula et al., 2009), indicate a postglacial corridor route had begun to form at both the northern and southern ends by at least 13,500 cal y BP. Bison from clade 1a were present in the southern corridor near present-day Calgary by 13,430–12,875 cal y BP, and bison from clade 2b were in northeastern British Columbia by 13,175–13,045 cal y BP. The postglacial corridor was fully open for dispersals by 13,000 cal y BP, when multiple overlapping radiocarbon dates belonging to bison from all three late-surviving clades (Fig. 13.2; clades 1a, 2a, and 2b) are found at the Clover Bar site (Fig. 13.1 and Table 13.1). The fact that bison from all three late-surviving clades are present during this same interval is crucial; although we can-

TABLE 13.1 Bison Specimens Recovered from the Corridor Region During the Time Frame of Corridor Opening (13,500–11,500 cal y BP)

Locality	^{14}C Age (y BP)	^{14}C Accession Number	Calibrated Date (1σ)	Clade	Reference
Tsiigehtchic, NWT	11,830 ± 45	OxA 18549	13,715–13,595	2	Zazula et al., 2009
Gallelli Pit, Calgary, AB	11,300 ± 290	RL-757	13,430–12,875	1a	Shapiro et al., 2004
Chetwynd, BC	11,240 ± 70	OxA 11274	13,175–13,045	2b	Shapiro et al., 2004
CloverBar, AB	11,255 ± 45	UCIAMS 117399	13,145–13,070	1a	This study
Tuktoyaktuk, NWT	11,185 ± 30	UCIAMS 81883	13,090–13,040	2	This study
CloverBar, AB	11,140 ± 25	UCIAMS 125537	13,065–13,005	2b	This study
CloverBar, AB	11,110 ± 25	UCIAMS 125527	13,060–12,975	1a	This study
CloverBar, AB	11,115 ± 25	UCIAMS 125531	13,060–12,985	2a	This study
CloverBar, AB	11,105 ± 25	UCIAMS 125533	13,060–12,965	1	This study
CloverBar, AB	11,100 ± 25	UCIAMS 125526	13,055–12,945	1a	This study
CloverBar, AB	11,100 ± 30	UCIAMS 125532	13,055–12,935	2b	This study
CloverBar, AB	11,085 ± 35	UCIAMS 125541	13,040–12,900	1a	This study
CloverBar, AB	11,080 ± 35	UCIAMS 117391	13,030–12,890	1a	This study
CloverBar, AB	11,080 ± 35	UCIAMS 117392	13,030–12,890	1a	This study
CloverBar, AB	11,080 ± 25	UCIAMS 125528	13,025–12,905	2b	This study
CloverBar, AB	11,075 ± 30	UCIAMS 117388	13,020–12,885	1a	This study
CloverBar, AB	11,050 ± 25	UCIAMS 125544	12,980–12,855	2b	This study
CloverBar, AB	11,040 ± 30	UCIAMS 117390	12,970–12,845	2b	This study
CloverBar, AB	11,030 ± 25	UCIAMS 125529	12,955–12,830	1a	This study
CloverBar, AB	11,010 ± 25	UCIAMS 125540	12,920–12,795	2b	This study
TwinBridges, AB	10,530 ± 30	UCIAMS 117384	12,555–12,430	1a	This study
Williston Lake, BC	10,460 ± 65	OxA 11272	12,540–12,190	2b	Shapiro et al., 2004
Charlie Lake Cave, BC	10,505 ± 45	OxA 12085	12,540–12,420	1a	Shapiro et al., 2004

Tumbler Ridge, BC	10,475 ± 40	UCIAMS 142224	12,535–12,395	2b	This study
Athabasca, AB	10,450 ± 55	OxA 11584	12,530–12,175	1a	Shapiro et al., 2004
Charlie Lake Cave, BC	10,440 ± 40	UCIAMS 142221	12,520–12,185	N/A	This study
Charlie Lake Cave, BC	10,435 ± 25	UCIAMS 11346	12,515–12,190	1a	Waters and Stafford, 2013; this study
Athabasca, AB	10,425 ± 50	OxA 11592	12,420–12,160	1a	Shapiro et al., 2004
Charlie Lake Cave, BC	10,430 ± 30	UCIAMS 11347	12,420–12,170	1a	Waters and Stafford, 2013; this study
Smoky River, AB	10,410 ± 30	UCIAMS 117382	12,400–12,170	1a	This study
Liard River, NWT	10,340 ± 90	Beta 255289	12,390–12,035	1a	This study
Henkel Ranch, AB	10,365 ± 40	UCIAMS 117401	12,380–12,135	1a	This study
Charlie Lake Cave, BC	10,340 ± 40	OxA 12084	12,375–12,065	1a	Shapiro et al., 2004
Charlie Lake Cave, BC	10,290 ± 40	UCIAMS 142220	12,145–11,980	N/A	This study
Charlie Lake Cave, BC	10,285 ± 40	UCIAMS 142219	12,140–11,975	2b	Shapiro et al., 2004; this study
Charlie Lake Cave, BC	10,260 ± 40	UCIAMS 142222	12,105–11,845	N/A	This study
Charlie Lake Cave, BC	10,230 ± 55	OxA 10580	12,060–11,825	2b	Shapiro et al., 2004
Charlie Lake Cave, BC	10,170 ± 40	UCIAMS 142218	11,955–11,765	1a	This study
Charlie Lake Cave, BC	10,060 ± 35	UCIAMS 142217	11,715–11,420	2b	This study
Twin Bridges, AB	10,060 ± 30	UCIAMS 117400	11,715–11,420	2b	This study
Charlie Lake Cave, BC	9,980 ± 40	UCIAMS 142223	11,600–11,320	N/A	This study

NOTES: Clade numbers correspond to Fig. 13.2*A*. AB, Alberta; BC, British Columbia; NWT, Northwest Territories; YT, Yukon Territory.

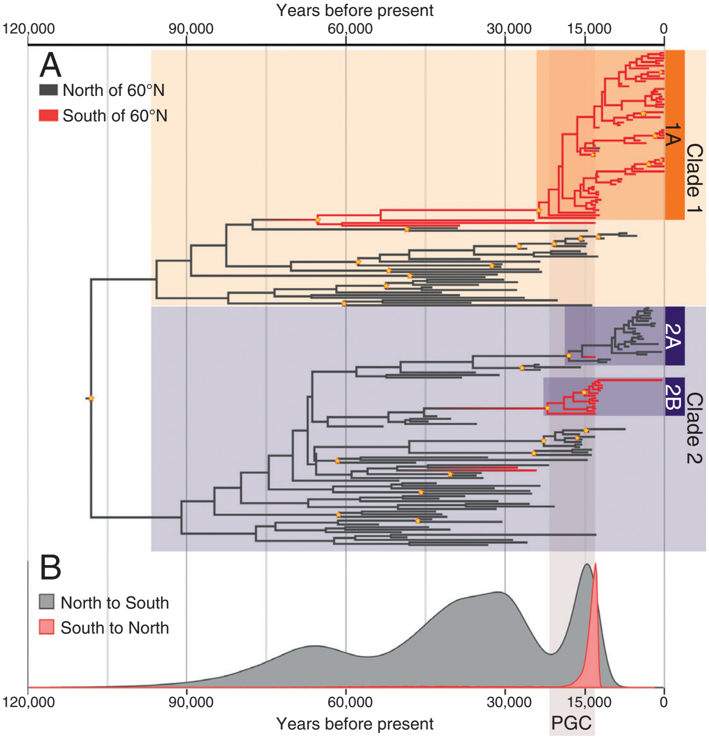


FIGURE 13.2 (*A*) A time-calibrated maximum clade credibility mitochondrial genealogy resulting from a Bayesian phylogeographic analysis of ~600 bp of the mitochondrial control region amplified from 192 bison. Clades 1 and 2 are denoted by orange and purple, as in Fig. 13.1, with the three late-surviving clades (1a, 2a, and 2b) highlighted. Gray and red branches represent haplotypes sampled from localities north and south of 60° N, respectively. This latitude is assumed to be the location of a final barrier to a traversable corridor (Gowan, 2013). Yellow stars denote nodes supported by posterior probabilities >90%, as inferred from the Bayesian analysis. (*B*) Density plots representing the timing of movements between north and south of 60° N. The plots have been rescaled so that the highest peaks are of equal height. The period of glacial coalescence (PGC) is overlain in gray between 23,000 and 13,500 cal y BP.

not determine with confidence the direction of dispersal of bison from clade 2b, our analyses indicate a northern origin for bison from clade 2a (Fig. 13.2*B*). Bison from clades 1a and 2b were present in stratigraphic subzones IIb–d at Charlie Lake Cave for ~1,000 years before the end of the Pleistocene, from ~12,500 to 11,500 cal y BP (Driver et al., 1996) (Fig.

13.1 and Table 13.1). These corridor bison were part of a diverse megafaunal community that included American lion (*Panthera leo atrox*), horse, western camel (*Camelops hesternus*), caribou (*Rangifer tarandus*), tundra muskoxen (*Ovibos moschatus*), helmeted musk oxen (*Bootherium bombifrons*), and mammoth (Jass et al., 2011; Ives et al., 2013), further supporting the notion that the area was productive habitat at the end of the Pleistocene. Based on present understanding of the biogeography of many of these mammal taxa, there is a general hypothesis that, apart from the clade 2a bison, most of the taxa that recolonized the corridor region came from the south, including horses, lions, camels, and muskoxen (Burns, 2010).

Data presented here and elsewhere demonstrate the critical role that continental ice sheets and the postglacial corridor played in biogeographic patterning among species and populations. For many mammalian taxa, paleontological and genetic data reveal distinct separation of populations north and south of the ice. Intriguingly, despite our conclusion that bison disperse into the postglacial corridor from both the north and south, we find only limited evidence of dispersal beyond the region of the corridor. In fact, the only evidence of dispersal completely through the corridor is the occurrence of a bison from clade 1a at Liard River, Northwest Territories, at 12,390–12,035 cal y BP, which corresponds to a northward dispersal (Fig. 13.2*B* and Table 13.1). Similarly to bison, several other taxa with Eurasian origins, such as caribou and American lion, probably traversed the corridor region before the LGM and established populations in the southern interior of North America (Burns, 2010). However, the responses of these species to deglaciation and environmental change at the end of the Pleistocene varied. For example, although bison from the south dispersed as far north as the Liard River, there is no mitochondrial evidence of further northward expansion into Alaska and Yukon during the Holocene. In contrast, mitochondrial data suggest that all present-day wolves (*Canis lupus*) in the North American subarctic and arctic are descended from a population that was south of the continental ice sheets during the LGM and that dispersed northward during the postglacial period (Leonard et al., 2007).

These results have two key implications for the role of the postglacial corridor as a pathway for biotic exchange between Beringia and interior North America. First, the opening of the postglacial corridor may have favored south to north, rather than north to south dispersal. Detailed biome reconstructions indicate that southern and central portions of the deglaciating corridor in this time range featured potentially more productive grasslands, open spruce woodlands, and boreal parkland, whereas northern portions of the corridor were marked by alpine, herb, and shrub tundras (Dyke, 2005). Southward dispersal may also have been limited for biological reasons, for example, if southern bison were better adapted than northern bison to the expanding grasslands within the corridor region (Guthrie, 1990). Second, the interval of time during which it is feasible to transect the corridor may have been limited. For bison, the barrier to further dispersal may have been the relatively quick replacement of grasslands at the northern end of the corridor by increasingly closed spruce forests, which are difficult for grazing herbivores such as bison to transect (MacDonald and McLeod, 1996; Wilson, 1996).

Consequences of the Postglacial Corridor Chronology for North American Human Prehistory

The expansion of bison into the corridor region provides proxy evidence for when this route was viable for human populations and, in doing so, allows further refinement of New World human settlement scenarios. Human genetic and archaeological evidence indicate that eastern Beringia and parts of the Americas well south of the ice sheets were populated by 14,000 cal y BP, suggesting that migration out of Beringia probably began more than 15,000 cal y BP ago (Gruhn and Bryan, 2011; Reich et al., 2012; Waters and Stafford, 2013; Raghavan et al., 2015). Our chronology for the opening of the postglacial corridor indicates that a fully habitable corridor connected Beringia and interior North America by ~13,000 cal y BP. This timing precludes the postglacial corridor as a southward route for initial human dispersal into the Americas, the corollary being that the first indigenous peoples leaving Beringia probably took a coastal route or potentially moved through western North America before glacial coalescence (Madsen, 2004; Erlandson, 2013).

We find that a bison belonging to the northern clade (2a) reached the Edmonton area by 13,000 cal y BP. It is therefore possible that established northern human populations also reached the central corridor by this time. Evidence from the archaeological record supports this hypothesis. For example, Alaskan archaeological sites including Swan Point, Mead, Broken Mammoth, Tuluaq, and Dry Creek, which were occupied from ~14,000 to 11,500 cal y BP, feature a variety of projectile technologies, sometimes associated with microblade industries (Goebel and Buvit, 2011). Similar microblade technologies are present at Vermilion Lakes (Banff National Park) and Charlie Lake Cave by ~11,500 cal y BP (Fedje et al., 1995; Driver et al., 1996; Tackney et al., 2015. In addition, human genetic data from Upward Sun River, Alaska, show founding New World mitochondrial haplotypes B2 and C1b in Alaska at ~11,500 cal y BP. Small, isolated groups of people may therefore have continued to disperse from Beringia to interior North America well after the corridor region opened (Tamm et al., 2007; Achilli et al., 2013; Ives, 2015; Tackney et al., 2015).

Our bison data also suggest that biotic conditions favored northward rather than southward movements through the corridor, paralleling archaeological data involving fluted point technology (Goebel et al., 2013). Bison clade 1a, which originates south of the ice sheets during the period of coalescence, predominates in our corridor sample, with one instance occurring as far north as the Liard River by ~12,200 cal y BP. The oldest recognizable Clovis complex sites in North America are estimated to range from 13,000 to 12,600 cal y BP (Goebel et al., 2008; Waters and Stafford, 2013), which slightly postdates our chronology for the opening of the southern end of the corridor. The Anzick child burial in Montana that dates to ~12,600 cal y BP (Rasmussen et al., 2014), and slightly earlier evidence for human hunting of western camels and horses at Wally's Beach, 420 km north of the Anzick site (Rasmussen et al., 2014; Waters et al., 2015), document the presence of people at the southern end of the corridor. Fluted point density maps indicate that this technology diminishes in frequency from the south to the north in the corridor region, consistent with the northward spread of this technology (Ives et al., 2013; Ives, 2015; Amick, 2016). By 12,500 cal y BP, fluted points are present at Charlie Lake Cave in British Columbia (along with clades 1a and 2b bison) and in sites in Alaska (Driver et al., 1996; Goebel et al., 2013).

CONCLUSIONS

Over the last several decades, phylogeographic research has evolved from being largely qualitative in nature to benefitting from a diversity of statistically rigorous approaches to detect evolutionary structure over both geographic space and, with the addition of paleogenetic technologies, over time. Here, we demonstrate the power of this approach to address a longstanding question in New World biogeography, specifically whether and when an ecologically viable corridor linked Beringia and interior North America during the late Pleistocene.

Until now, much of the research concerning the corridor region has focused on whether or not the corridor was the initial pathway for early human inhabitants of the New World. Our data from fossil bison help shift scientific attention to a different first: the LGM coalescence of the ice sheets marked the first time in the Quaternary that the corridor did not exist as a biogeographic pathway. Once the corridor opened, bison populations that had been isolated for millennia, in Beringia and in interior North America, entered the deglaciated region from both its northern and southern ends. Human populations in eastern Beringia and interior North America had similar possibilities to reengage. In the human case, this involved the added cultural complexity that would accompany societal interactions in an era where exotic raw materials and technological ideas began to circulate widely in the Americas. Increasing genetic and archaeological evidence continue to support the idea that the corridor was a route for dispersal in both directions, with consequences for biological and technological diversity on both sides of the receding ice sheets.

MATERIALS AND METHODS

Radiocarbon Dating

Collagen was extracted from bone and tooth samples using ultrafiltration methods outlined in Beaumont W et al. (2010) and radiocarbon dated using accelerator mass spectrometry (AMS) at the Keck Laboratory, University of California (UC), Irvine (UCIAMS); the Center for AMS (CAMS) at Lawrence Livermore National Laboratory; or Beta Analytics (Table 13.1 and *Dataset S1*). AMS dates were calibrated using the IntCa113 calibration curve (Reimer et al., 2013) in OxCal v4.2 (https://c14.arch.ox.ac.uk/oxcal/OxCal.html) and are reported with 1 standard deviation (SD). For samples that were redated at UCIAMS (SFU 1848, SFU 1849, SFU 3429, and SFU 15004; *Dataset S1*), we used only the new dates for the calibration. All dates reported in the text are in calendar years before present, unless otherwise noted.

DNA Extraction, PCR, Cloning, Library Preparation, Target Enrichment, and Sequencing

To facilitate comparison with previously published data from bison, we isolated ~600 bp of the hypervariable portion of the mitochondrial control region (CR) from 45 Canadian bison bone and tooth samples, dated to the late Pleistocene and Holocene (Table 13.1 and *Dataset S1*). We performed DNA extraction, library preparation, and PCR setup in dedicated ancient DNA facilities at the Pennsylvania State University (PSU) and UC Santa Cruz (UCSC) that were physically isolated from modern molecular biology research. Depending on the sample, we extracted DNA from 100 to 250 mg bone powder using one of two methods that are highly optimized for the recovery of ancient DNA molecules (Rohland et al., 2010; Dabney et al., 2013), performing one negative extraction control for every five to eight processed samples.

To generate CR sequence data, we used a mixture of four approaches: (*i*) direct Sanger sequencing of PCR products, (*ii*) Sanger sequencing of cloned PCR products, (*iii*) Illumina amplicon sequencing of PCR products, and (*iv*) mitochondrial target enrichment followed by Illumina sequencing. For approaches (*i*)–(*iii*), we amplified the target CR fragment either in a single PCR amplification or as a series of overlapping fragments, depending on the preservation of the sample (primer combinations are provided in *Table S1*) (Shapiro et al., 2004). We performed PCR in 25-μL reactions with the following components: 20 μg rabbit serum albumin, 0.25 mM dNTPs, 1× High Fidelity buffer, 1 U Platinum Taq High Fidelity (Life Technologies), 2.4 mM $MgSO_4$, 0.4 μM of each primer, and 1 μL DNA extract, with the following cycling conditions: 12 min at 94°C, 30 s at 94°C, 45 s at variable annealing temperature (*Table S1*), 45 s at 68°C, and 1 min at 68°C, with the middle three steps repeated for 50 cycles. We cleaned PCR products using either Millipore μ96 plates or Sera-Mag SpeedBeads (ThermoScientific) in 18% (wt/vol) PEG-8000, the latter of which followed the bead-based reaction cleanup protocol of Rohland and Reich (2012).

For approaches (*i*)–(*iii*) above, we then used one or more of these approaches to assess the accuracy of the resulting PCR amplicon sequences: (*i*) bidirectional direct sequencing on AB3730xl genetic analyzers at the PSU Genomics Core Facility or the UC Berkeley DNA Sequencing Facility, using BigDye v3.1 chemistry (Life Technologies); (*ii*) cloning PCR products using the TOPO-TA cloning kit (Life Technologies) according to the manufacturer’s instructions, followed by selection and PCR amplification of six to eight colonies following Shapiro et al. (2004); and/or (*iii*) pooled sequencing of barcoded PCR products using the Illumina MiSeq platform, in which PCR products derived from the same sample were pooled in equimolar ratios and turned into Illumina DNA libraries using Meyer and Kircher (2010) with modifications from Heintzman et al. (2015). These indexed libraries were then pooled in equimolar ratios and sequenced on the Illumina MiSeq platform using v2 150-bp paired-end chemistry, following the manufacturer’s instructions.

For the fourth data generation approach described above, we constructed Illumina DNA libraries as above directly from the DNA extract. We then enriched these libraries for the whole bison mitochondrial genome using biotinylated RNA baits (MYbaits v2; MYcroarray), following the manufacturer’s instructions. These enriched libraries were then sequenced on the MiSeq as described above, but using v3 75-bp paired-end chemistry.

Consensus Sequence Construction

For data generation approaches (*i*) and (*ii*), we assessed sequence quality manually and called consensus sequences using Lasergene v9 (DNASTAR) or Geneious v6.1.6 (Biomatters). For approaches (*iii*) and (*iv*), we binned the short-read data by index, and then removed adapters and merged paired-end reads using SeqPrep (https://github.com/jstjohn/SeqPrep). For approach (*iii*), we removed primer sequences from merged and remaining unmerged reads using in-house scripts and mapped each read to both *Bison bison* (GenBank: NC_012346) and *B. priscus* (AY748705)

CR reference sequences using BWA v0.6.1 (Li and Durbin, 2009), resulting in two datasets per sample. We called consensus sequences using Geneious with the base agreement threshold set to 75% and the minimum coverage set to 50×. For approach (*iv*), we aligned merged reads to the reference *B. bison* mitochondrial genome sequence using an iterative assembler MIA (Green RE et al., 2008). We called bases that had >3 times coverage and >67% agreement. For analysis, we then trimmed the resulting consensus sequence to the ~600-bp target. We combined data from all four approaches to create robust consensus sequences for each sample. If we identified conflicts between sequences resulting from the different approaches, we either performed additional analyses or coded conflicting sites as ambiguous. Unsequenced regions are considered missing data.

Phylogeographic Analyses

We aligned the 45 new CR sequences to a dataset of 147 previously published, radiocarbon-dated or present-day, North American bison CR sequences (*Dataset S1*), using Se-Al (Sequence Alignment Editor; v2.0a11). We performed a Bayesian phylogeographic analysis using BEAST v1.8.3 (Drummond et al., 2012). We assumed a generalized time-reversible evolutionary model, with gamma distributed rate variation and a proportion of invariable sites (GTR+G+I), a strict molecular clock with a rate calibrated using the median calendar ages of each radiocarbon-dated specimen, and the flexible skygrid model of the coalescent process (Gill et al., 2013). To infer the timing and directionality of movement between Beringia and interior North America, we adopted the discrete phylogeographic model described in Edwards et al. (2011), assigning each sample to either the North or South population based on whether the sample originated to the north or south of 60° N, which archaeological data and simulation studies indicate was the final barrier to a corridor (Gowan, 2013). To simplify interpretation of the results, the two present-day bison that were sampled at locations just north of this cutoff (at 60° and 61.4°) were classified as southern.

To learn about ancestral movements between north and south, we estimated the posterior distribution of the time of north-to-south and south-to-north movements using Markov jumps (Minin and Suchard, 2008). This technique exploits dynamic programming and tricks from numerical analysis to efficiently compute the expected number and timing of specific transitions within a continuous-time Markov chain (CTMC) that conditions only on the directly observed end states at the tips of phylogenetic tree on which the CTMC acts. The resulting posterior distribution of transition times naturally incorporates uncertainty in the tree and estimated CTMC rates. We ran two MCMC chains for 50 million states each, sampling the posterior states of all model parameters and trees every 5,000 states. We discarded the first 10% of samples from each run as burn-in, combined the remainder using Log-Combiner v1.8.3, and computed posterior means, posterior SDs, and 95% highest posterior density intervals in Tracer v1.6. We summarized the combined set of posterior trees and identified the maximum clade credibility (MCC) tree using TreeAnnotator v1.8.3, which we visualized using Figtree v1.4.2.

Data Availability

The input BEAST file is available as *Dataset S2*. Novel control region mitochondrial DNA sequences have been deposited in GenBank, with Accession Numbers KU705765-KU705809. All fossil specimens used in this study are curated in the repositories listed in *Dataset S1*.

ACKNOWLEDGMENTS

We thank James Cahill and Andrew Fields for technical assistance, Mike Waters for providing samples, and two anonymous reviewers for comments that improved the manuscript. P.D.H., A.E.R.S., and B.S. were funded by National Science Foundation (NSF)-Applied Research Center Grants 1203990 and 1417036 and Gordon and Betty Moore Foundation Grant GBMF3804. D.F. was funded by the Canada Research Chairs program and a grant from the Natural Science and Engineering Research Council of Canada. J.W.I. was funded by the Landrex Distinguished Professorship. M.A.S. was funded by NSF-Division of Mathematical Sciences Grant 1264153 and National Institutes of Health Grant R01 HG006139.

14

Evolutionary Lessons from California Plant Phylogeography

VICTORIA L. SORK,*†# PAUL F. GUGGER,‡
JIN-MING CHEN,§ AND SILKE WERTH‖

Phylogeography documents the spatial distribution of genetic lineages that result from demographic processes, such as population expansion, population contraction, and gene movement, shaped by climate fluctuations and the physical landscape. Because most phylogeographic studies have used neutral markers, the role of selection may have been undervalued. In this paper, we contend that plants provide a useful evolutionary lesson about the impact of selection on spatial patterns of neutral genetic variation, when the environment affects which individuals can colonize new sites, and on adaptive genetic variation, when environmental heterogeneity creates divergence at specific loci underlying local adaptation. Specifically, we discuss five characteristics found in plants that intensify the impact of selection: sessile growth form, high reproductive output, leptokurtic dispersal, isolation by environment, and the potential to evolve longevity. Collectively, these traits exacerbate the impact of environment on movement between populations and local selection pressures—both of which influence phylogeographic structure.

---

*Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095-7239; †Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095-1496; ‡Appalachian Laboratory, University of Maryland Center for Environmental Science, Frostburg, MD 21532; §Key Laboratory of Aquatic Botany and Watershed Ecology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, China; and ‖Institute of Plant Sciences, University of Graz, 8010 Graz, Austria. #To whom correspondence should be addressed. Email: vlsork@ucla.edu.

We illustrate how these unique traits shape these processes with case studies of the California endemic oak, *Quercus lobata*, and the western North American lichen, *Ramalina menziesii*. Obviously, the lessons we learn from plant traits are not unique to plants, but they highlight the need for future animal, plant, and microbe studies to incorporate its impact. Modern tools that generate genomewide sequence data are now allowing us to decipher how evolutionary processes affect the spatial distribution of different kinds of genes and also to better model future spatial distribution of species in response to climate change.

Phylogeographic analyses document the impact of geographic boundaries, geological changes, and climatic fluctuations on evolutionary processes that drive population divergence, speciation, and the formation of communities (Comes and Kadereit, 1998; Avise, 2000; Hewitt, 2001; Hickerson et al., 2010). Early and highly cited studies of phylogeography have emphasized the impact of recent glaciation in Europe and North America and detail population expansion and migration northward (Avise et al., 1987; Hewitt, 1996, 2000). Comparison of phylogeographic data across multiple species often reveals common patterns due to similar changes in space and time associated with specific landscapes (e.g., Hewitt, 1996, 1999; Soltis et al., 1997; Brunsfeld et al., 2001; Ruegg and Smith, 2002; Calsbeek et al., 2003; Carstens et al., 2005; Rissler et al., 2006; Shafer et al., 2010; Ralston and Kirchman, 2012). In regions that were glaciated during the Pleistocene, the location of refugia and regional topography can shape the postglacial migration routes in similar ways across species (Avise et al., 1987; Comes and Kadereit, 1998; Hewitt, 1999; Petit et al., 2003; Shafer et al., 2010; Sproul et al., 2015). The emphasis on formerly glaciated regions where many species expanded from common refugia highlighted the congruence in patterns of historical migration and demography. However, the focus on commonality has overlooked the fact that, ultimately, phylogeography is the result of the interaction between organisms and their environment. Because species respond individualistically to climate fluctuations across the physical landscape, co-occurring species often show discordant phylogeographic patterns (e.g., Soltis et al., 2006). In fact, not only do species respond individualistically, but so too can populations within species (Hu et al., 2009; Gugger et al., 2010). One way to understand these responses is to incorporate life history traits into interpretations of congruence or discordance among species in their phylogeographic patterns (Papadopoulou and Knowles, Chapter 8, this volume).

We propose that plant species have characteristics that amplify their sensitivity to the environment and have the potential to affect phylogeography differently than observed for most animal species. Our argument is motivated by a classic article by Bradshaw (1972), who proposes that important "evolutionary consequences of being a plant" are the particularly strong impacts of natural selection on plant populations. For instance, simply the fact that plants photosynthesize means that they depend on their local environment for light, water, and nutrients to survive, which creates a strong opportunity for selection and local adaptation. In addition to their dependence on the environment, they possess certain traits that enhance the influence of selection. Sessile growth form makes plants very different from animals. Once plants are established, they must survive and reproduce in that environment whereas most animals have some ability to move if the environment is not suitable. Likewise, plants have additional traits that Bradshaw argues enhance the impact of natural selection, including high reproductive output, leptokurtic dispersal, strong influence of environment on gene flow, and potential for longevity. Obviously, these traits are not exclusive to plants. For example, algae photosynthesize, corals and many marine invertebrates are sessile, and fungi exhibit high reproductive output with leptokurtic dispersal. Nonetheless, our goal here is to explore how key characteristics found in plants illustrate an evolutionary lesson about the role of natural selection in shaping migration patterns and demographic history over space and time. We are not necessarily arguing that selection is stronger in plants than animals although Bradshaw (1972) argues strongly for this case. However, we do propose that specific traits found in plants and some other taxa allow selection to shape phylogeography. In the first part of the chapter, we discuss the phylogeographic impact of each of these traits. In the second part, we present two case studies that illustrate how these characteristics shape phylogeography.

PHYLOGEOGRAPHIC CONSEQUENCES OF BEING A PLANT

Sessility

A plant is stationary once established. The obvious consequence of sessility is that it is easy to document the history of movement because the genotypes of the plants provide a clear geographic record of their distribution. More importantly, this attribute means that, once a seed is dispersed, germinates, and becomes a seedling, the plant must survive and reproduce in that same location. In contrast to animals that can move in the course of a day from one environment to another to optimize environmental conditions or can move seasonally for better foraging or mating conditions, a plant must tolerate its location and eventually reproduce, or die. Thus, spatially varying selection pressures are often conspicuous across plant populations, creating local adaptation as well as strong geographic patterns of adaptive genetic variation that will be associated with environmental gradients, sometimes even over short distances (Antonovics, 1971; Wright et al., 2006).

An early demonstration of the importance of natural selection in plants is the work of Clausen et al. (1947), which showed through a series of common garden plots that plants grew best when planted near their native environments and that plants from local environments grew better in those sites than plants derived from sites from different environments. The use of common gardens and reciprocal transplants has a long tradition in plant evolutionary biology to document genetic differences (Linhart and Grant, 1996; Kawecki and Ebert, 2004). In forestry, the use of common gardens, also known as provenance studies, has provided a foundation for forest management practices (Langlet, 1971; Mátyás, 1996). Through these studies, we know that phenotypic differences among natural populations have a genetic basis. Due to the emergence of next-generation sequencing, it is also possible to identify adaptive variation in natural populations by identifying outlier loci that significantly correlate with environmental gradients (Hancock and Rienzo, 2008; Coop et al., 2010; Sork et al., 2013). These loci under selection will produce a geographic pattern of adaptive genetic variation that might differ from that demonstrated by neutral markers. Thus, the sessile nature of plants may, at a minimum, better document the spatial signature of adaptive versus neutral processes, which provides a lesson on the importance of this trait for phylogeographic studies.

Most Plant Species Exhibit a Leptokurtic Pattern of Gene Flow

An obvious difference between plants and animals is their dispersal biology. Gene flow can occur through pollen and seeds in plants (Levin, 1981), which means that the geographic patterns of genetic variation are created through two separate and often asymmetric processes (Oddou-Muratorio et al., 2001; Hamilton and Miller, 2002; Sork and Smouse, 2006; Sork et al., 2015). Plants have chloroplasts, which are haploid and nonrecombining (like mitochondrial DNA), that prove useful for documentation of migration and demographic history. Due to their uniparental inheritance, the organellar genomes can help to separate the impact of the two processes on phylogeography, which has been demonstrated in several tree species (e.g., Latta and Mitton, 1997; Liepelt et al., 2002; Burban and Petit, 2003; Gugger et al., 2010). A unique aspect of plant dispersal is that we can see whether genes are spread through pollen or seed dispersal and colonization. This separation also allows us to determine whether environmental factors, such as phenological differences that could affect differential fertilization success between populations or biotic dispersal vectors, could create individualistic responses to climatic fluctuations and physical landscapes.

A further notable characteristic of plant dispersal biology is that gene flow is often leptokurtic, with the majority of pollen and seeds dispersing locally near where the genotype is likely to be most well adapted (Bradshaw, 1972). At one extreme, plants with inbred mating systems and limited seed dispersal may have highly restricted gene flow and a high degree of genetic structure that will be enhanced by selection (Loveless and Hamrick, 1984). The phylogeography of these species might be similar to animals with restricted dispersal and small population sizes. However, at the other extreme, outcrossing species with leptokurtic dispersal kernels will exhibit both local and long-distance dispersal (Sork and Smouse, 2006; Hardy, 2009). In this regard, plants might show different phylogeographic patterns than animals. In plants with long-distance dispersal, selection will favor individuals who can survive in the new environment, which will then create divergence from the source population for loci under selection. However, for neutral genetic markers, long-distance dispersal will maintain high connectivity among populations (Hardy, 2009). For example, in the Mediterranean regions of Europe, the tree and shrub species with the most genetically divergent populations were those with low seed dispersal abilities. In contrast, the genetically most diverse populations were located at intermediate latitudes, where divergent lineages from separate refugia immigrated and established (Petit et al., 2003). If animals move long distances, we do not expect strong population divergence at adaptive genetic loci because they are likely to select suitable habitats. Of course, if animals disperse and do not maintain ongoing gene exchange, population divergence will emerge. Such a scenario is illustrated in a study of 22 species of Californian amphibians and reptiles and 75 phylogeographic lineages across those species (Rissler et al., 2006), which identified congruent lineage breaks for multiple species across the Central Valley; San Francisco Bay; and the Sierra Nevada, Tehachapi, and Trinity ranges of California. Such results suggest that dispersal and divergence across species are similarly affected by the landscape. In contrast, the lesson from plants is that they often have a distribution of dispersal distances and that the interaction between dispersal and selection for local adaptation can promote divergence at smaller scales for adaptive genetic variation and less divergence among neutral markers.

High Reproductive Output

Most plant species have high fecundity and much higher offspring-to-parent ratios than animal species. This high reproductive output is reminiscent of the high volume of male gametes produced by many organisms relative to the number fertilized; in many plant species, the high reproductive output results in thousands of seeds that can survive if they land in the right environment, but few do—resulting in strong selection. For trees, high fecundity across years reduces the impact of any single reproductive episode, which buffers year-to-year variation in seed production (Petit and Hampe, 2006). A benefit of high reproductive output is that it mitigates the uncertainty of seed survival. When combined with the leptokurtic dispersal kernel for pollen and seeds, it enhances the potential for the evolution of local adaptation within the vast majority of locally dispersed propagules and the opportunity for long-distance dispersal events to transport the seeds somewhere that an individual can survive. This life history trait would enhance the ability of plants or other sessile organisms to have large species distributions. For example, it may be a contributing factor for wide ranges of invasive weedy plant species. The lesson we learn from this trait is that high reproductive output increases the potential for wide dispersal and for selection on propagule after dispersal that enhances their ability to colonize and establish in new sites.

Environmental Influence on Gene Flow

The combination of sessility and local adaptation can create population divergence in heterogeneous landscapes due to the joint effects of isolation by distance (IBD) and isolation by environment (IBE) (Wang and Bradburd, 2014) or isolation by adaptation (IBA) (Andrew et al., 2012). An implicit assumption of phylogeography is that gene flow decreases with distance and is modified by the physical landscape. Because most plant species exhibit leptokurtic dispersal that results in spatial autocorrelation of genotypes (Smouse and Peakall, 1999; Vekemans and Hardy, 2004), isolation by distance (Wright, 1943; Slatkin, 1987) will almost always be involved in plant phylogeography. With IBE, the environmental differences will also shape genetic distances through a variety of processes. Such processes could include isolation caused by differences in flowering time created by environment differences (e.g., Stanton et al., 1997; Yamagishi et al., 2005) or nonrandom gene flow created by dissimilar pollinator communities across environments (e.g., Vickery, 1992; Bradshaw and Schemske, 2003). IBA increases divergence at neutral loci when migrants from one locally adapted population are unable to successfully immigrate into a population in a different environment. For example, Andrew et al. (2012) report that environmental variables such as soil nitrogen and vegetative cover explained more of the variation in genetic differences than IBD or landscape resistance alone. Thus, the dependence of plants on the local environment that they remain in throughout their life span means that the environmental heterogeneity is highly likely to shape plant phylogeography.

In case studies reported below, we highlight how species with plant characteristics are very likely to illustrate the influence of the environment on population divergence. In the literature, we also find other examples of environmental effects on the phylogeography of plants. For example, in a study of the Sonoran desert succulent *Euphorbia lomelii* in Baja California, temperature variables affecting phenological synchrony of flowering significantly affected geographic patterns of genetic variation after controlling for phylogeographic history (Dyer et al., 2010). In another example, we learn how microhabitat preferences may create different phylogeographic outcomes in codistributed and closely related montane sedges from the Rocky Mountains (Massatti and Knowles, 2014). In that study, through the use of next-generation sequence data and species distribution models, Massatti and Knowles (2014) demonstrate that the ecology of the species interacts with glaciations to produce fundamental differences in the past distributions. Thus, the strong impact of the environment on patterns of geographic differentiation found in plants highlights the role of the environment on phylogeography and the opportunity for selection that needs more attention. Of course, the phylogeography of animal species is also likely to show evidence of IBE (Sexton et al., 2014), but the mechanisms for those patterns are likely to differ.

Longevity, or Plants Can Be Trees

One of the most unique differences between plants and animals is that plants can be extremely long-lived with indeterminate growth (Bradshaw, 1972). The local environment experienced by a tree throughout its life span will influence both a plant's size and its reproduction. Even a long-lived turtle grows to adult size that is more or less constant and can move to a new environment to suit its preference. In contrast, for long-lived perennial plants, selection acts strongly on the seedling stage, and it persists to act throughout their long life spans. Longevity helps maintain effective population size and genetic variation of populations (Petit and Hampe, 2006), which means that reproductive episodes throughout a life span will produce genetically diverse offspring that may experience different selective pressures in the face of environmental variability. Thus, longevity can help prevent extinction of local populations (Petit and Hampe, 2006). In addition, trees have great potential for long-distance gene flow and large effective population sizes (Petit and Hampe, 2006). The bristlecone pine (*Pinus longaeva*) of the White Mountains in California and Nevada provides a dramatic example of the impact of longevity and large effective population size on phylogeography (Schierenbeck, 2014). The bristlecone pine can live over 5,000 years and once was widespread at lower elevations throughout the Great Basin during the Last Glacial Maximum. However, its current range is restricted to isolated mountaintops at the western edge of its former distribution (Lee et al., 2002). Despite the limited and fragmented distribution of bristlecone pine, its level of genetic diversity is comparable with that of other pines. Thus, we see many tree species with widespread distributions whose longevity may have prevented population contractions from resulting in complete extinction before environmental conditions would allow expansion.

CASE STUDIES OF PLANT PHYLOGEOGRAPHY

Evolutionary Lessons from a California Endemic Oak, *Quercus lobata*

Study Region

California is a species-rich biogeographic region that simultaneously offers a multifaceted history of geological change, complex geographical structure, and historic climatic fluctuations (Calsbeek et al., 2003), providing a convenient laboratory for phylogeographic study. The majority of California is within the species-rich California Floristic Province (Raven and Axelrod, 1978; Baldwin, 2014), which contains more than 4,700 native plant species, almost half of them endemic (Lancaster and Kay, 2013; Baldwin, 2014), and is one of the world's biodiversity hot spots (Myers et al., 2000; Calsbeek et al., 2003; Mittermeier et al., 2005; Loarie et al., 2008). Many factors could have contributed to this biogeographic pattern. First, this region is a confluence of different floras that allowed a high accumulation of species (Raven and Axelrod, 1978; Swenson and Howard, 2005; Baldwin, 2014). Second, long-term environmental stability of some areas has been inferred to account for substantial paleoendemism (Stebbins and Major, 1965; Raven and Axelrod, 1978). Third, topographic complexity, environmental heterogeneity, and isolation may have created high speciation rates (Baldwin, 2014). Fourth, the emergence of the Mediterranean climate enhanced the diversification of lineages since the Middle Miocene when the transition toward summer dry periods began (Baldwin, 2014). However, based on an analysis of the molecular phylogenies of 16 angiosperm clades, Lancaster and Kay (2013) argue that low extinction rates, not high speciation rates, account for the high biodiversity of the California Floristic Province. Not only would the lack of glaciation foster persistence, but also the topography itself may enhance survival because plants could respond to climatic fluctuations by shifting their altitudinal distributions along steep topographic gradients with a moderating effect of montane areas on precipitation patterns. In addition, the combination of topographical gradients with only moderate temperature oscillations would allow populations to expand and contract regionally and avoid extinction (Grivet et al., 2006; Gugger et al., 2013). Thus, California provides an opportunity to examine phylogeographic patterns that can be distinct from the postglacial population expansion dynamics reflected by species in Europe (Petit et al., 2003), eastern North America (McLachlan et al., 2005; Soltis et al., 2006), and northwestern North America (Soltis et al., 1997; Brunsfeld et al., 2001; Carstens et al., 2005).

The Case

Q. lobata Née (valley oak) offers an excellent system to illustrate some of the phylogeographic lessons to be learned from plants, particularly in a region that has experienced limited glaciation. Valley oak is one of the four major endemic tree oaks of California. Its distribution along the eastern foothills of the Coast Ranges and western foothills of the Sierra Nevada surrounding the Central Valley covers a large portion of California. This distribution has remained relatively stable within California through global climate cycles (Gugger et al., 2013), yet valley oak experiences diverse environments and geography across its distribution. We have analyzed the evolutionary history using several kinds of evidence (Stebbins and Major, 1965): (*i*) nuclear microsatellite markers to assess the overall genetic structure of the species using STRUCTURE (Falush et al., 2003a) (Fig. 14.1*A*) and environmental associations with constrained ordinations; (*ii*) chloroplast genetic markers, using BAPS (Corander et al., 2008) to assess patterns of colonization by seed dispersal over time (Fig. 14.1*B*); (*iii*) species distribution modeling to assess the potential shifts in distribution during glacial and interglacial periods (Fig. 14.2). In addition, we have evidence from next-generation sequencing data to investigate the role of natural selection on divergence (Sork et al., 2016) (Fig. 14.3).

Nuclear DNA phylogeography supports a role for high genetic connectivity among populations, most likely through regionally widespread pollen dispersal (Fig. 14.1*A*). In one study, a clear east–west split was observed (Gugger et al., 2013), and, in others, there was a north–south subdivision (Sork et al., 2010; Ashley et al., 2015; Gugger et al., 2016). A key insight from this work is that patterns of genetic variation on the landscape are not just a product of geographic features but are also a product of climatic influences on migration patterns and local selection. Multivariate analyses suggest that climate explains at least as much genetic variation as spatial/geographic forces do (Sork et al., 2010; Gugger et al., 2013), consistent with a growing body of work in other oak species (Riordan et al., 2016). In valley oak, a signature of climate impacts from both the present and the Last Glacial Maximum implies a stability of these popu-

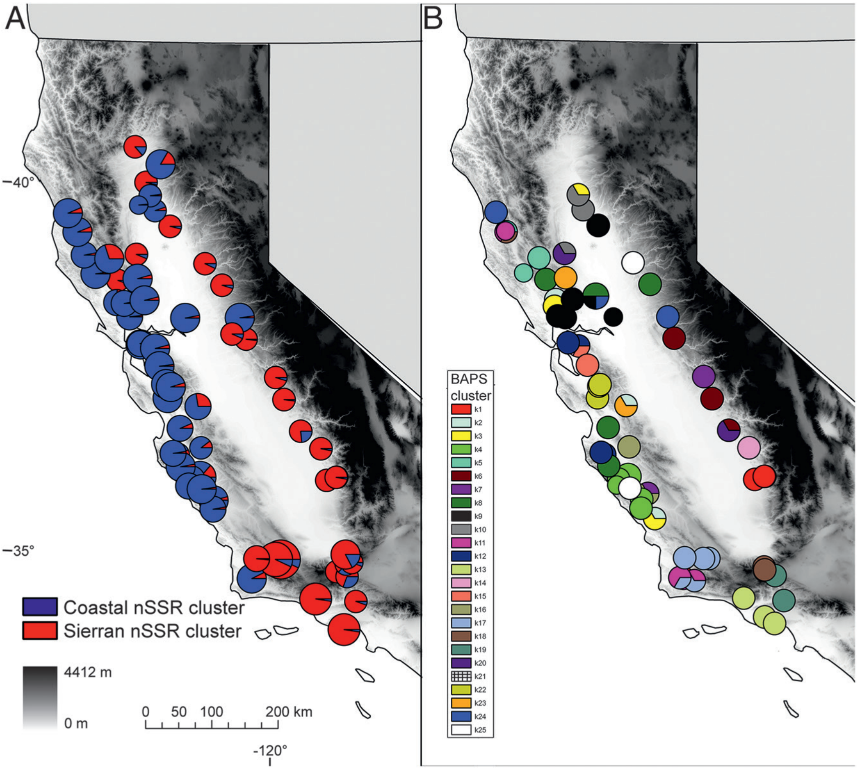


FIGURE 14.1 Maps of genetic structure of *Q. lobata* inferred from the nuclear and chloroplast microsatellite genotypes. (*A*) STRUCTURE analysis of nuclear microsatellites. (*B*) Genetic clusters using BAPS. Pie charts represent the proportion of genetic assignment to each inferred genetic cluster. Background gradient (gray scale) represents elevation (figures modified from Gugger et al., 2013).

lations in this topographically complex region that has lasting impacts. Furthermore, the east–west split is associated with both geographic separation and niche differences detected through species distribution models (Gugger et al., 2013), a finding corroborated in other oak species in the region (Ortego et al., 2015a,b). The strong associations of overall genetic structure with climate are most likely due to isolation by environment (Wang and Bradburd, 2014).

The phylogeographic analyses of chloroplast DNA variation support the inference that valley oak's distribution was stable and that migration was local (Fig. 14.1*A*). We observed many more haplotypes than observed

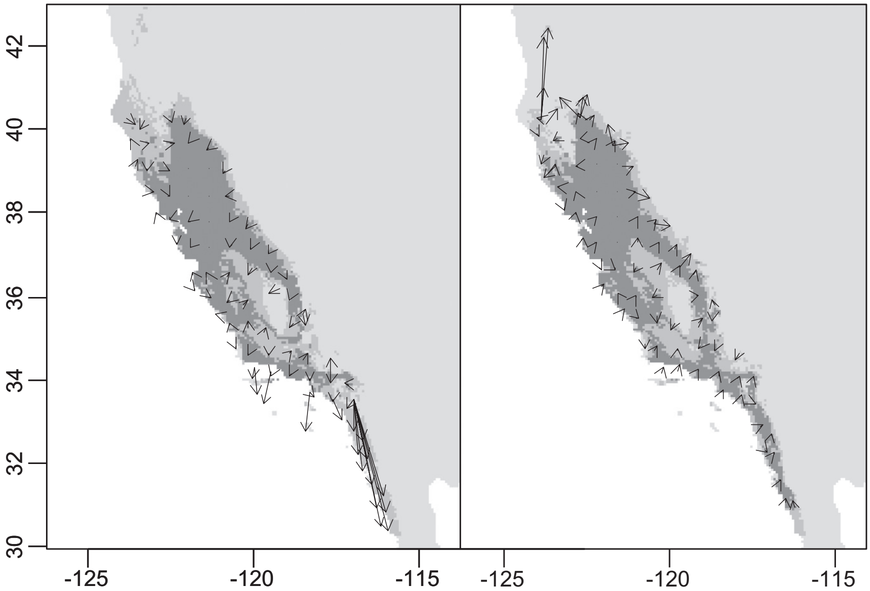


FIGURE 14.2 Vectors represent the local migration direction of *Q. lobata* during recent interglacial and glacial periods predicted by species distribution models. (*A*) Predicted movement from the last interglacial period (~120 ka) to the Last Glacial Maximum (21 ka). (*B*) Movement from the Last Glacial Maximum to the present. Dark gray background represents areas where *Q. lobata* was inferred to be present in both time periods, medium gray indicates presence in one time period, and background light gray is presence in no time period (modified from Gugger et al., 2013).

in Europe, and those haplotypes were only locally distributed rather than dispersed widely across the species range as observed in regions with large impacts of glaciation (Grivet et al., 2008; Gugger et al., 2013). Species distribution models based on climate data indicate that valley oak habitat has been largely stable throughout its current distribution since at least the last interglacial period over 100 ka (Gugger et al., 2013). The limited distribution changes that did occur were likely local elevational shifts, which are suggested by the short distances of vectors between interglacial, glacial, and contemporary periods (Fig. 14.2). This finding is in stark contrast to similar niche models in an eastern North American white oak, *Quercus alba*, which expanded and contracted dramatically in response to climate change (Gugger et al., 2013). Habitat stability and the

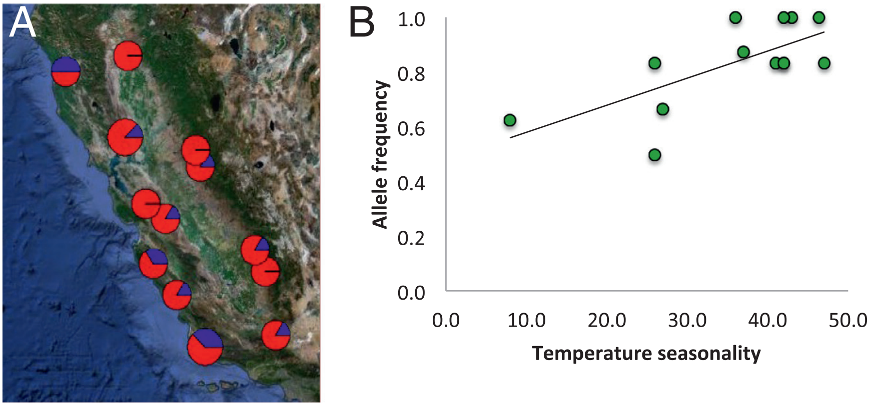


FIGURE 14.3 Association of an outlier SNP from a growth gene with geography and climate gradient detected through environmental analysis. (*A*) Map of allele frequency of a gene putatively involved in local growth adaptation. Pie charts represent proportion of each allele in each sample site. (*B*) Regression of temperature seasonality with allele frequency of the same growth gene (data taken from Sork et al., 2016).

restricted movement of valley oak in a diverse climate landscape may have promoted local adaptation.

Phylogeographic studies use neutral genetic markers to study migration and demographic history, but a growing body of research investigates geographic patterns of adaptive genetic variation. In valley oak, landscape genomic analyses to test for natural selection identify a number of candidate genes for local adaptation to climate (Gugger et al., 2016; Sork et al., 2016). What is most striking is that allele frequencies in these genes do not typically align with neutral genetic structure and instead follow other climate gradients (Fig. 14.3). This pattern demonstrates that adaptive alleles may readily disperse on the landscape despite population structure, such as shown for introgressed genes from invasive salamanders (Fitzpatrick et al., 2010). Another interesting initial finding is that significant evidence of natural selection along climate gradients could be detected with relatively small samples sizes, suggestive of strong local adaptation to environments (Gugger et al., 2016). Sork et al. (2016) found that several functional genes have allele frequencies that correlate with climate gradients in a pattern different from neutral genetic structure. In a landscape genomic study of balsam fir, a gene associated with circadian rhythm showed evidence of local adaptation to temperature (Fitzpatrick and Keller, 2015). A step forward for phylogeography will be to incorporate adaptive genetic variation into a historical demographic context to infer when adaptations arose in response to particular environmental factors (Gavin et al., 2014). Nonetheless, these studies demonstrate how dispersal patterns and environmental forces come together to generate local adaptation and current patterns of phylogeography.

Returning to the five characteristics of plants, valley oak illustrates several lessons about features common to plants. The sessile nature of valley oaks leaves them vulnerable to the local environment around them. As a result, the climate environment has particularly strong influence on genetic variation at the landscape level, leading to isolation by environment and local adaptation. Having two modes of dispersal leads to a dual pattern of local genetic structure with effective regional gene flow. These modes, along with their leptokurtic distribution, permit adaptive genetic variation that enhances fitness to spread relatively freely (presumably through pollen dispersal) despite regional or local genetic structure. In the case of this tree, longevity has led to a lasting signature of past climate on present patterns of genetic variation. Long generation times may hamper the ability of trees to respond to future climate change. On the other hand, high reproductive output, effective gene flow, and large effective population sizes may enhance adaptive responses.

Evolutionary Lessons from California's State Lichen, *Ramalina menziesii*

Study Region

The widely distributed lace lichen *R. menziesii* Taylor (Fig. 14.4) covers a dramatic latitudinal range centered in California but extends south to Baja California of Mexico and north along the western coast of North America to Alaska. The lichen is found in six ecoregions, from fog desert in south ranges to the coniferous forest of the Pacific Northwest and Alaska. This unusual species distribution provides an opportunity to study phylogeographic processes across glaciated and unglaciated regions, to determine whether diversity is higher in the ecoregions within the California Floristic Province than other ecoregions (Fig. 14.5*A*). Within each ecoregion, the lichen will be found on a subset of phorophyte species (i.e., the plant on which the lichen grows). Through an analysis of the four low-copy nuclear genes (Fig. 14.5*B*), we found evidence that the center of genetic diversity and probably the origin of this lichen is the California Floristic Province (Sork and Werth, 2014).

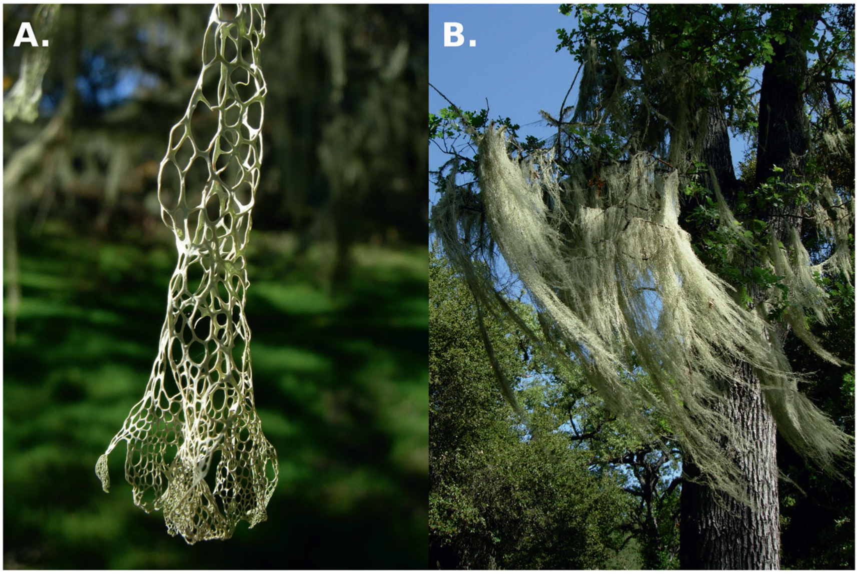


FIGURE 14.4 Photographs of lace lichen (*R. menziesii*). (*A*) Closeup photography of lace lichen thallus (photograph taken by S.W.). (*B*) Deciduous valley oak (*Q. lobata*) heavily colonized by lace lichen.

The Case

R. menziesii, which shares many features with plants, offers a unique opportunity to examine the impact of those features. One key difference among most vascular plants and lichens is that lichens do not tap a root system into a water reservoir, but depend on ambient precipitation and fog to satisfy their demands for water. Nonetheless, like plants, their needs for humidity, nutrients, and light for photosynthesis and their sessile lifestyle make them very dependent on the local environment. Lichens are created by the symbiotic relationship between a mycobiont and a photobiont. In this lichen, the haploid lichen-forming ascomycete *R. menziesii* associates with the haploid photosynthetic green alga *Trebouxia decolorans* (predominantly) (Werth and Sork, 2008, 2010, 2014; Sork and Werth, 2014) and lives for 15–20 years or longer in its lace-like growth form (Fig. 14.4). This symbiosis also provides an opportunity to use comparative phylogeography to understand the extent to which environmental factors have

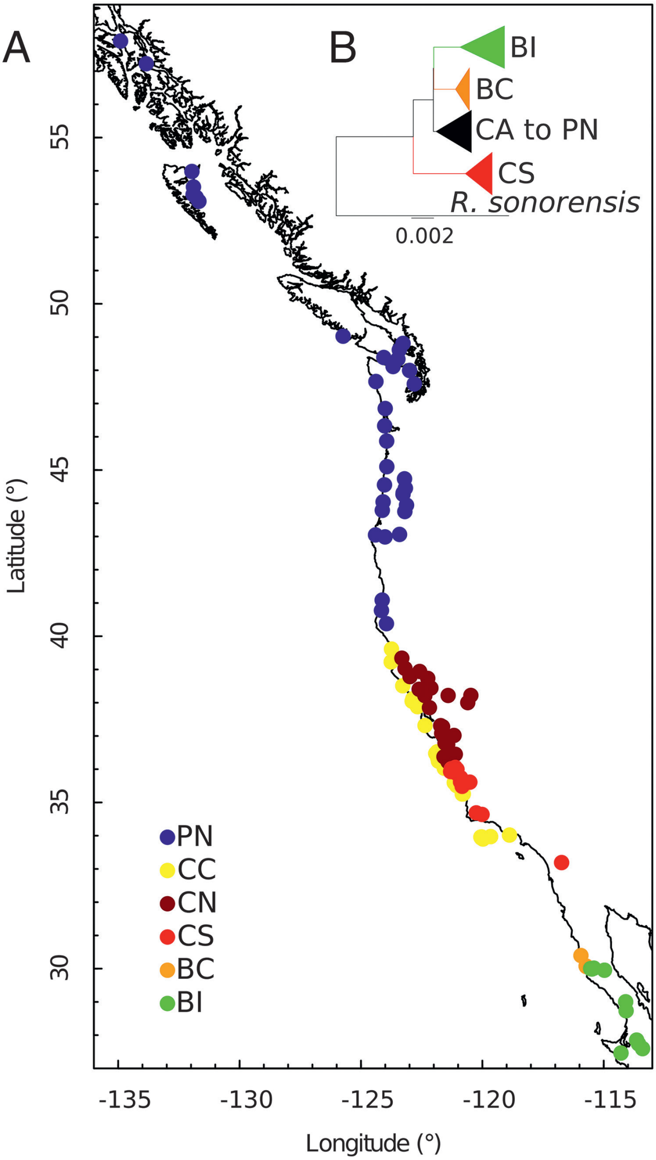


FIGURE 14.5 Spatial distribution of sampling sites and clades of the lichen-forming fungus *R. menziesii* along western North America. (*A*) Sampling sites, color-coded by ecoregion. (*B*) Bayesian Markov Chain Monte Carlo tree generated in BEAST on the basis of four nuclear genes, with ecoregion-specific clades color-coded as indicated in the figure. AK, Alaska; BCC, Baja California coastal; BCI, Baja California interior; CA, California; CAI, California interior (based on data reported in Sork and Werth, 2014).

contributed to the evolutionary history of a species with similar features to many plant species, including the absence of longevity.

To understand the phylogeographic history of the lichen, it is most illustrative to investigate the fungal genome because it is the lichenized fungus that creates the intricate thallus structure of the lichen *R. menziesii.* Below, we discuss how dispersal and high reproductive output shape the complex history of the fungal genome and its green-algal photobiont, and we compare the phylogeographic history of the two symbiotic taxa that have experienced the same environment during evolution because of their association. The dispersal of both the lichen mycobiont and photobiont are important to the phylogeography of *R. menziesii.* The fungus reproduces mainly sexually with high reproductive output and is able to disperse broadly (Werth and Sork, 2008). To form a new lichen thallus, a germinating fungal ascospore has to associate with a compatible local algal strain through a process termed "relichenization" that involves horizontal transmission of the photobiont. The lace lichen has high dispersal capability combined with high reproductive output. In fact, it has been hypothesized for taxa with microscopic dispersal stages, such as lichens that disperse with spores, that dispersal does not limit their geographic distributions: In these cases, everything is everywhere, and "the environment selects" (Baas Becking, 1934). In the case of the lace lichen, spores must land where *T. decolorans* (or in a small region, an alga similar to *Trebouxia jamesii*) is present or lichen will not establish (Werth and Sork, 2014a). The mycobiont must be sufficiently specialized to be compatible with the photobiont's physiology but not so specialized that it cannot relichenize with a range of green algal genotypes. For this particular lichen, high dispersal capability and high reproductive output of the microscopic spores are essential, given that the photobiont is not codispersed in these propagules.

An interesting feature of the lace lichen is that it can persist in a broad range of habitats and ecoregions, most likely due to local adaptation of the photosynthetic green alga, rather than the mycobiont. For example, in a local-scale study, the green algal populations (but not the mycobiont) showed genetic differentiation across their oak phorophytes (Werth and Sork, 2008, 2010). In our species-wide genetic analysis of *T. decolorans*, we found that the abundance of strains differs among phorophyte species and that green-algal populations are differentiated according to phorophyte (Werth and Sork, 2014). Through a multivariate analysis of the genetic association of *T. decolorans* across the landscape with climate and geographic location, we found a significant influence of climate, as well as a significant effect of the species of the phorophyte where the green-alga genotype was found. These findings highlight the potential role of the environment as a selecting force shaping the phylogeographic structure of this symbiotic association by acting on the photobiont. An advantage of this system could be that the fungus associates with locally adapted algal partners (Usher et al., 2007), which may allow the lichen to have a broad distribution range.

The lace lichen study system also offers a noteworthy opportunity to study comparative phylogeography. First, we compared the genealogical trees from a subset of samples that had mycobiont and photobiont genotypes. We discovered some similarities in their clustering with certain ecoregions, but overall the phylogenetic trees were largely incongruent among symbionts, suggesting little coevolution between fungal and algal genotypes (Chen et al., 2016). Next, we compared recent migration patterns for different types of genetic markers to see whether the symbionts were showing similar patterns of movements, which one might expect given that they coexist within the same individual and may respond to climatic fluctuations in a similar way (Chen et al., 2016). The analysis revealed that the mycobiont and photobiont of lace lichen showed a general tendency to move south, with different patterns of movement across ecoregions for each of the markers (Fig. 14.6). The slightly similar clustering of haplotypes in the genealogical trees probably reflects the barriers imposed by differences across ecoregions. However, the overall lack of concordance between the phylogeographic patterns of these two symbiotic taxa might be explained better by dissimilar migration (Chen et al., 2016). In short, the differences in the dispersal abilities of the two taxa, and possibly the differences in their microhabitat preferences, illustrate that plant life history traits can disrupt cophylogeography of mutualistic co-occurring taxa. Overall, our comparative analyses illustrate that the contribution of both migration and local adaptation is shaping the phylogeography of the involved taxa.

The lace lichen, which shares many traits found in plants, also illustrates the evolutionary lesson that selection matters in phylogeography. Their sessile nature and their physiology make them sensitive to the local environment around them with a particularly large opportunity for selection to create locally adapted genotypes. The impact of selection is reflected in fungal and algal clades restricted to regions with unique climatic conditions (Sork and Werth, 2014; Werth and Sork, 2014) and in photobiont strains differentiated among phorophyte species (Werth and Sork, 2010). Lace lichen's widespread dispersal capability, along with its high reproductive output, and its ability to associate with locally adapted algal strains have allowed it to occupy an exceptionally wide geographic range across multiple climate zones, encompassing tropical/subtropical dry-to-humid temperate conditions. Climatic signatures are apparent in its phylogeographic pattern, with several fungal and algal clades being restricted to regions with unique climatic conditions. A unique feature of this study system is the opportunity to conduct a comparative phylo-

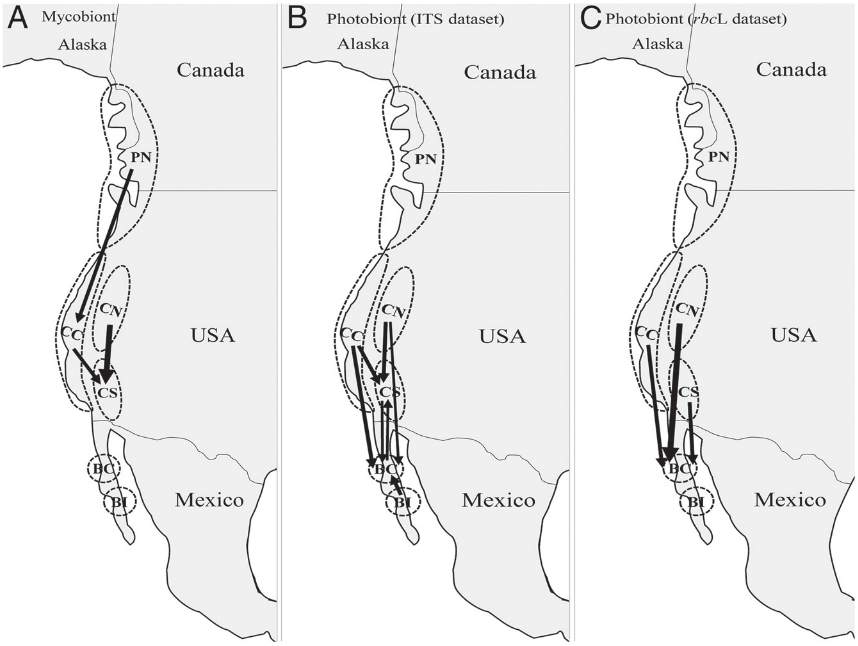


FIGURE 14.6 Comparison of recent migration patterns of *R. menziesii* and *T. decolorans* across ecoregions as inferred from coalescent analysis for three kinds of markers: (*A*) fungal nuclear genes, (*B*) *T. decolorans* nuclear gene, and (*C*) *T. decolorans* chloroplast gene (maps modified from Chen et al., 2016).

geographic analysis of the lace lichen for two organisms experiencing the exact same environments and physical barriers to their migration. The fact that they show incongruent phylogeographic structure within ecoregions and lack of comigration among ecoregions is a dramatic illustration of how selection by habitat, generation time, and dispersal abilities can create different histories and different evolutionary trajectories, even for taxa sharing the same environment.

CONCLUSIONS

The overarching evolutionary lesson from plants and these two case studies that occupy regions with high environmental heterogeneity and high species diversity is the contribution of natural selection in the response of species to climate fluctuations and the presence of major physical barriers. The geographic location of early phylogeographic studies in glaciated regions where populations have expanded and contracted from unglaciated refugia has taught us about the impact of neutral demographic patterns on species distributions. Common physical barriers have often led to concordance among multiple species that highlights shared impacts of neutral demographic processes. Increasingly, however, we are finding that co-occurring species are responding individualistically (Prates et al., Chapter 3, this volume; Papadopoulou and Knowles, Chapter 8, this volume). As studies have been expanded to regions where populations have persisted over longer time frames than the last glaciation and to regions with high environmental heterogeneity promoting ecological impacts of gene movement, we observe the effect of nonneutral processes. Selection pressures across diverse environments will enhance population divergence and often lead to discordance among species in the same region. Selection against migrants from different environments will result in divergent genomewide genetic signatures across populations. In addition, the movement of pollen, seeds, or individuals across a landscape allows the introduction of variants at certain loci that might enhance local adaptation of the populations. The phylogeography of adaptive variation may differ from that of neutral markers. Plants may have traits that can exacerbate or simply highlight the impact of natural selection, but the phylogeographic history of all organisms will be affected by both neutral and adaptive processes. The increased availability of datasets with large numbers of loci derived from next-generation sequencing will create the next phase of phylogeography that allows us to examine the history of movement of not just species but the processes that led to the shifts in species distributions. Future studies that use the new genomic tools that include neutral and adaptive genetic markers, incorporate information about adaptive and demographic responses, and include the contribution of life history traits will more accurately model species-specific shifts in species distributions due to climate change.

ACKNOWLEDGMENTS

We thank the organizers of this symposium, Francisco Ayala and John Avise, its participants, and two anonymous referees. V.L.S. thanks Ryan Harrigan, Brad Shaffer, and Steve Weller for valuable discussion and Krista Beckley for help with the figures. We thank our collaborators on valley oak phylogeography: Frank Davis, Delphine Grivet, Maki Ikegami, Robert Westfall, and Jianli Zhao. We acknowledge the following funding sources: the National Geographic Society (V.L.S. and S.W.), the Swiss National Science Foundation (S.W.), European Commission Marie Curie Fellowship “Lichenomics” 302589 (to S.W.), the University of California, Los Angeles (P.F.G. and V.L.S.), and the Chinese Academy of Sciences (J.-M.C.).

15

Human Phylogeography and Diversity

ALEXANDER H. HARCOURT

Homo sapiens phylogeography begins with the species' origin nearly 200 kya in Africa. First signs of the species outside Africa (in Arabia) are from 125 kya. Earliest dates elsewhere are now 100 kya in China, 45 kya in Australia and southern Europe (maybe even 60 kya in Australia), 32 kya in northeast Siberia, and maybe 20 kya in the Americas. Humans reached arctic regions and oceanic islands last—arctic North America about 5 kya, mid- and eastern Pacific Islands about 2–1 kya, and New Zealand about 700 years ago. Initial routes along coasts seem the most likely given abundant and easily harvested shellfish there as indicated by huge ancient oyster shell middens on all continents. Nevertheless, the effect of geographic barriers—mountains and oceans—is clear. The phylogeographic pattern of diasporas from several single origins—northeast Africa to Eurasia, southeast Eurasia to Australia, and northeast Siberia to the Americas—allows the equivalent of a repeat experiment on the relation between geography and phylogenetic and cultural diversity. On all continents, cultural diversity is high in productive low latitudes, presumably because such regions can support populations of sustainable size in a small area, therefore allowing a high density of cultures. Of course, other factors operate. South America has an unusually low density of cultures in its tropical latitudes. A likely factor is the phylogeographic movement of peoples from the Old World bringing novel, and hence lethal, diseases

---

Department of Anthropology, Ecology Graduate Group, University of California, Davis, CA 95616. Email: ahharcourt@ucdavis.edu.

to the New World, a foretaste, perhaps, of present-day global transport of tropical diseases.

Human phylogeography has been for decades and is still studied under the rubric of physical anthropology or biological anthropology. Among other interests, these fields investigate the global spread of hominids, hominins, and humans. In doing so, they use all methods and sources of information available. The two fields, and hence the discipline, of human phylogeography also investigate the many correlates of our species' distribution across the varied environments of the world—physiology, genetics, behavior, anatomy, commensals, and also a special form of behavior, namely culture. Here, within the topic of human phylogeography, I address the geographic spread of the human species across the world and the influence of the environment on regional cultural diversity.

In both the professional and popular literature, the word "human" sometimes refers to all *Homo* species or even the non-*Homo* genus, *Australopithecus*. I use the word to mean only *Homo sapiens*.

HUMAN SPECIES' GLOBAL DIASPORA

African Origins

The earliest signs so far of humans, *H. sapiens*, are cranial fragments in southwest Ethiopia with a suggested date of 195 kya (McDougall et al., 2005). (All ages are in calendar years ago and not isotope years ago.) However, the range of possible dates of the fragments is, in fact, 195–105 kya. The next oldest remains are from around 165 kya, also in Ethiopia (Clark et al., 2003). Preceding *H. sapiens*, several other *Homo* species and their ancestors, *Australopithecus* species, were African too (Klein, 2009, chap. 5).

This summary hides considerable debate about nearly all aspects of human origins, even whether our species had a single origin (Klein, 2009, chap. 5). Not only that but also, new finds continually change our understanding, especially in a continent as little investigated as Africa. Whether, then, the human species arose in Ethiopia or arrived there from elsewhere in Africa is still an open question (Reed and Tishkoff, 2006; Henn et al., 2011).

Be that as it may, by 100 kya, humans lived over much of Africa (Klein, 2009, chap. 5). Indeed, humans had left Africa by then. Stone tools along with the bones of humans who made them that date to about 125 kya have been found in southeast Arabia (Armitage et al., 2011). Until recently, both archeological and anatomical evidence indicated that humans got no farther until about 60 kya (Mellars, 2006; Reyes-Centeno et al., 2014) or maybe a little over 70 kya (Oppenheimer, 2012). It seemed that the aridity of an approaching ice age might have prevented farther expansion or indeed led to the disappearance of humans from the Arabian Peninsula.

However, increasingly older dates east of Arabia indicate the need for rethinking of at least the dating of our diaspora out of Africa. For instance, a date of about 45 kya for humans' arrival in Australia (O'Connell and Allen, 2004) might need to be extended to 60 kya (Hiscock, 2013). Other dates elsewhere are perhaps too recently published to be independently tested (for instance, the date of 67 kya from a single metatarsal of a potential human in the Philippines) (Mijares et al., 2010). Another in southern India at 74 kya is already disputed. The date comes from volcanic ash around microlithic artifacts (Petraglia et al., 2007). However, no bones are associated with the artifacts, and humans might not even have made them (Mellars et al., 2013). Given the nature of the tools, the next likeliest toolmaker is Neanderthal, but the nearest Neanderthal remains are over 2,000 km away.

If some more distant dates might be wrong, dates almost always get older. In 2015, excavations in southern China revealed "unequivocal" human teeth dated to 120–80 kya (Liu W et al., 2015). A previous similarly aged find for that region is disputed (Oppenheimer, 2012), but if the recent find and claim is substantiated, and especially if the older end of the range of dates is substantiated, it becomes possible that, in fact, the first exodus from Africa continued on east.

The last interglacial period was at its height 125 kya. The climate and hence distribution of vegetation were much the same as now. In other words, the environment was favorable by comparison with the Old World aridity of a glacial period. The Sahara might have been in some regions wooded savannah, and hence good habitat then for migrating humans (Osborne et al., 2008; Oppenheimer, 2009). By 60 kya, however, the approaching ice age should have been associated with a drying climate in Africa, which leads to the suggestion that aridity drove the African exodus (Carto et al., 2009).

However, Lake Malawi sediments indicate a warm and wet eastern Africa at about 75 kya (Lane et al., 2013). Far from poor conditions driving dispersal, maybe good conditions and an increasing human population did so. Currently, we do not have enough information on African paleoclimates or the trajectory of African populations to distinguish these two scenarios (Harcourt, 2012, chap. 3; Henn et al., 2012).

Assuming no advanced rafting capabilities, the early dispersers from Africa must have entered the Arabian Peninsula across the marshes at the north end of the Red Sea, perhaps properly the Reed Sea (Drews and Han, 2010). By whatever route they arrived on the eastern coast of the Arabian

Peninsula, a northern dry-shod route round the north end of the Persian Gulf then seems likely.

However, at 60 kya, the Red Sea's narrowest point at the southern end might have been less than 20 km across. Ethiopians there could have seen Arabia across the Bab el Mandeb Strait. After they arrived in southeast Arabia, the dispersing humans would have found the Strait of Hormuz between them and Iran to have been only about 50 km across. Because maybe only 15,000 years later, at 45 kya, humans had reached Australia across 100 km of ocean (whether they arrived through Indonesia or New Guinea), maybe a southern route into Arabia, and then Iran is not impossible (Oppenheimer, 2012). Indeed, modeling that took into account parameters such as lowered sea levels, climate, the nature of the environment, and a preference for coastal travel, suggests one or both southern crossings as possible, even likely (Field and Lahr, 2005; Eriksson et al., 2012).

The dates given so far and the evidence for Africa as the phylogeographic origin of the human species come largely from archeology and paleontology. Genetics confirm humans' African origin. A recent origin from a single small population is indicated by the fact that the human species is less genetically diverse than some populations of chimpanzees (Gagneux et al., 1999; Stone et al., 2002). The size of the world's founding diaspora out of Africa is debatable, but genetic analyses indicate a figure of a few hundred (Macaulay et al., 2005). Humans' geographic origin in Africa is indicated by the fact that human molecular diversity is greatest in Africa (Cann et al., 1987; Campbell and Tishkoff, 2008). For instance, although four main mtDNA forms exist in African populations, only two exist outside, both from just one of the African forms (Oppenheimer, 2003; Stoneking and Harvati, 2013). Also, people of almost any region in the world are more genetically different from African populations than they are from all other populations (Cavalli-Sforza et al., 1994, chap. 2). Such lines of evidence of humans' African origin have been repeatedly confirmed, including with morphological analyses (Relethford, 2004; Li et al., 2008).

Humans arriving in Eurasia from Africa could have met Neanderthals and even Denisovans. Indeed, some genetic analyses indicate that humans mated with both of these other hominins (Abi-Rached et al., 2011; Lohse and Frantz, 2014; Prüfer et al., 2014; Fu et al., 2015). The findings are disputed, however, with the suggestion that, instead or in part, they could indicate descent from a common ancestor, the population of which was structured (i.e., genetically heterogeneous geographically), which realistically will almost always be the case over any appreciable area (Eriksson and Manica, 2012).

Old World Outside Africa

The oldest, largely accepted dates of humans outside of Arabia all fall around 50–40 kya regardless of whether in Asia, Australia, or Europe (O'Connell and Allen, 2004; Barker et al., 2007; Pope and Terrell, 2008; Demeter et al., 2012; Hiscock, 2013; Bosch et al., 2015). However, as already indicated, not only have several studies suggested older dates, but the chances of finding the oldest evidence are so slim that oldest dates must inevitably get older.

New Guinea, Australia, and Tasmania at 50 kya were a single continent, Sahul. One geographic source for the peopling of both New Guinea and Australia is thus a possibility (indeed, on genetic grounds, perhaps a likelihood) (Oppenheimer, 2012). Nevertheless, a separate migration into each is also possible (Kayser, 2010; van Holst Pellekaan, 2013). The shortest sea crossings in either case would have been, to repeat, about 100 km.

Subsequent to the earliest arrivals in New Guinea and Australia, these two populations could have remained separate from each other and the rest of the Old World for the next 40,000 years (Hudjashov et al., 2007). In New Guinea, a significant late introduction was that of the sweet potato a little over 300 years ago, which allowed expansion of agriculture, and hence populations, to regions too high or poor for the previous main food crop, taro (Allen, 2013).

By around 5 kya, the Pama–Nyungan language family originating in northeast Australia covered most of Australia, except the far northwest and the continent's dry interior (Clendon, 2006; McConvell, 2013). The Pama–Nyungan-speaking peoples entered the latter in the first millennium AD, about 3,000 years after Australia's 40,000-year isolation ended with an immigration that brought the dingo with it, probably from India (Pugach et al., 2013).

Tasmania was part of Sahul until around 8 kya, when rising seas finally formed the Bass Strait. Despite no sea barrier between Tasmania and Australia before that, Tasmania's first inhabitants apparently arrived only 35 kya, 10,000 years after first confirmed arrival in Australia (Jones, 1995).

The deliberate near eradication two centuries ago of native Tasmanians by arriving Europeans, largely British, means a paucity of genetic or linguistic information from which to judge origins, migrations, and settlement in the region. Nevertheless, linguists have identified five main groupings apparently so different from one another that the only words in common are largely those applying to European introductions, for example, of livestock (Bowern, 2012).

In Europe itself, the earliest dates in the west are for sites in the heel of Italy at around 44 kya (Benazzi et al., 2011), and about 43 kya in southern Britain (Higham et al., 2011). In other words, people were moving north as the peak of the ice age was rapidly approaching. However, the drop in temperature in the Northern Hemisphere from the peak interglacial at about 125 kya to the peak ice age at around 25 kya was erratic. Several times temperatures rose to nearly match the interglacial maximum, including at about 45 kya (Oppenheimer, 2012).

At the height of the ice age, however, the European ice cap covered Scandinavia and northern Europe down to northern Germany and the middle of Britain and Ireland. Evidence so far indicates that humans retreated from almost all of northern Europe. They returned to Britain from the Basque region of northern Spain about 10 kya (Oppenheimer, 2006) or around 1,000 years after the beginning of the Neolithic revolution and the expansion of agriculture into Europe from the Near East (Larson et al., 2007a). Celts and then Anglo-Saxons followed in historical times from mainland western Europe (Pitulko et al., 2016).

The words Basque, Celt, and Anglo-Saxon that I have just used describe both cultures and peoples, as with descriptors of cultures and peoples all over the world. In theory, a culture could move independently of its originating population. In practice, cultures usually move only because people move (Cavalli-Sforza et al., 1994; Cavalli-Sforza, 2000). This coordinated movement is a specific example of a general pattern (Pagel and Mace, 2004). In jargon, cultures are transmitted vertically and not horizontally. However, exceptions exist. Of 19 skeletons at a burial site in Oxford, England, the isotope ratios in the teeth of 18 of the people were the same as in the local soil. In other words, they were native-born. However, the burial site was culturally Anglo-Saxon (Hughes et al., 2014). In this case, the residents had adopted the culture of the invaders.

Four thousand years before humans were in Italy, they were on the central north coast of Siberia at Yenisei Bay hundreds of kilometers above the Arctic Circle. We know this from clear signs of killing and dismemberment of a mammoth and dating of the find (Pitulko et al., 2016). Sources differ on whether the region was eventually covered by the northern ice cap. However, farther to the east, at the famous Yana Rhinoceros Horn site in far northeast Siberia, the humans there at 32 kya (Pitulko et al., 2004) could have remained through the ice age, because the climate in eastern Siberia was too dry to support an ice sheet.

Genetic evidence indicates that people in northeastern Siberia arrived there from south central Russia (Schurr, 2004), although some linguistic evidence indicates north central Russia also (Vajda, 2010). A recent surprise is evidence from a 24-kya site near Lake Baikal in central south Siberia of a large genetic component of western European origin (Raghavan et al., 2014). The same genetic signature was detected at a 17-kya site 600 km northwest of the Lake Baikal site, implying that people remained in at least southern Siberia through the peak of the last ice age (Raghavan et al., 2014).

The earliest signs so far of humans elsewhere in eastern Asia are from around 40 kya, although a genetic analysis indicates 60 kya for eastern Asia and a little under 40 kya for Japan (Shi et al., 2008). Humans in Hokkaido, the northernmost island of Japan, could be from southeast Siberia, but those in the southern islands are from Korea (Adachi et al., 2011). That double phylogeographic origin is the pattern also for Japan's other mammals. Within Japan, the two geographic origins are indicated by the so-called Blakiston's Line through the Tsuguru Strait between Hokkaido and Honshu. With the approaching ice age's lowering of sea levels, the original human immigrants could have arrived overland. However, movements to outlying islands from maybe 9 kya were over sea. Japan experienced several subsequent immigrations, with one at nearly 3 kya producing perhaps the largest genetic contribution to the islands (Hudson, 2013; Matsumura and Oxenham, 2013).

Elsewhere in Asia, topography as well as cold might have been a barrier to movement, because the earliest signs of humans in Tibet date to 31 kya (Qi et al., 2013). The Himalayas as a barrier is nicely shown by the greater genetic distance between people on either side of them than the average for the same geographic distance elsewhere (Rosenberg, 2011).

Movement Within and into Africa

Of course, as humans spread across the Old World outside of Africa, they also spread within Africa. For instance, Y-chromosome analyses indicate links of west and east African pygmoid peoples with the ancient southern African Khoisan/San peoples (Behar et al., 2008) and between west African pygmoid peoples and east African Hadza peoples (Batini et al., 2011). Although the field of phylogeography usually uses genetics as its evidence, genetics need not be its only tool. Corresponding to genetic indications of links between eastern and southern African peoples, the rock art of Ituri pygmoid peoples in eastern Zaire is apparently similar to that of Khoi peoples in southern Africa (Batini et al., 2011) (Khoisan, Khoi, San, Khoe, and Khoenkhoen are sometimes distinguished in the literature and sometimes not). One of the latest and largest movements within Africa was the expansion of the Bantu peoples from around maybe 3 kya judged from archeology, genetics, and linguistics (Ehret, 2013; Robertshaw, 2013).

In historic times, the slave trade produced probably the fastest movement of large numbers of Africans from sub-Saharan Africa into North Africa. People also moved back into Africa from the rest of the Old World (Reed and Tishkoff, 2006; Gallego Llorente et al., 2015) [hence, for example, the surprising finding that the San Khomani peoples of South Africa, outwardly similar to the adjacent San Namibia, have about 10% admixture of western European genes (Hellenthal et al., 2014)].

Americas

Until the last ice age, the Bering Strait was a barrier between the Old and New Worlds. However, although 80 km across now, the strait was a broad flat expanse of land about 50 m above sea level during the peak of the ice age. Then or later, Siberians crossed into Alaska.

An allele and a single blood group, both almost ubiquitous among Native American and Amerindian populations and both present in eastern Siberians, indicate that a very small population from Siberia peopled the New World (Cavalli-Sforza et al., 1994; Schroeder et al., 2007). For instance, although northeastern Siberians have all three major blood groups (A, B, and O), more than 95% of native Amerindians are O (Cavalli-Sforza et al., 1994). As said, the northeastern Siberians are, in part, genetically western European. That being the case and given that Native Americans of both Americas originated in northeast Siberia, then western European genes in Native Americans is perhaps inevitable. The extent of contribution might be surprising, however. One calculation has the western European complement constituting perhaps one-quarter of the Native American genome (Raghavan et al., 2014).

As elsewhere in the world, these earliest arrivals could well have traveled first down the coast of both continents (Fagundes et al., 2008; Marangoni et al., 2014). Regarding timing of their entry into the Americas, previously accepted estimates of earliest arrivals now need revision. The quite widely accepted earliest date for the presence of humans in North America was 15.5 kya for a site in central Texas (Waters et al., 2011). South America's earliest date was 14.5 kya from the famous Monte Verde site in southern Chile (Dillehay, 1999; Dillehay et al., 2008). However, those finds are now superseded by a 2015 report of at least 18 kya at Monte Verde (Dillehay et al., 2015) and a 2014 report of dates of roughly 30–20 kya in northeast Brazil (Kinoshita et al., 2014). This latter finding makes a formerly largely rejected claim of around 32 kya in northeast Brazil (Guidon, 1986) now not so incredible.

These dates in both Americas along with several others before them must silence all future mention of the "Clovis-first" hypothesis, namely that the people of the Clovis culture were the first into North America at about 13 kya. Nevertheless, the Clovis people and their culture were phylogeographically successful. The genotype of an infant from 12.6 kya found in western Montana tightly associated with Clovis tools indicates not only a strong Siberian heritage, but also a strong genetic link to most Native North American peoples as if Clovis replaced the previous North

Americans (Rasmussen et al., 2014). Either the prior immigrants stuck to the coast, not moving inland, or before they could densely populate North America, the Clovis peoples entered and superseded them.

Monte Verde in southern Chile is about 16,000 km from Siberia. How long it took these first Americans to reach southern South America, we have little to no idea, given that the oldest South American dates are now older than the oldest North American dates. One line of evidence could be the distances traveled by relatively modern nonsedentary traditional societies. In the Americas, they moved camp a median of about 100 km/y (Binford, 2001). At that rate and with straight-line travel down the coast, humans would have taken less than two centuries to travel from Alaska to Monte Verde.

Alternatively, humans moved through the world by demographic expansion. Moving into continents empty of humans and moving into prey populations unused to humans and hence, easily hunted, the expansion could have been rapid. A calculation that accounted for number and distance of camp movements, frequency of foraging forays, reproductive rates, rate of population increase, costs of carrying children, hunting and foraging returns, and potential for children to contribute to the diet concluded that the 16,000-km journey in 2,000 years was possible even with a high rate of reproduction and hence the necessity to feed and carry children (Surovell, 2000). After all, that time for that distance translates to less than 25 m/d.

A main reason why fast movement could be compatible with high rates of reproduction was that women at the vanguard of the advance into virgin territory might have needed to forage just one-quarter of the distance of those following behind and entering depleted land. Indeed, across hunter–gatherer societies, those with more camp moves per year can reproduce more than those with fewer (Surovell, 2000).

Similarly, an analysis of eastern Canadian genealogies indicated that the ancestors of the current population who were at the front of the population expanding through the region in the 300 years to the mid-1900s contributed more to the present-day gene pool than did people from the core of the population's range (Moreau et al., 2011). An explanation for the phenomenon comes from the genealogies of females in the region, which indicate that those at the front had 15% more offspring than those behind and 20% more married children (Moreau et al., 2011).

We know little about original human movement through the Amazon, in part because archeology across large expanses of forest is difficult. Not only that but Amazonia might have lost nearly two-thirds of its languages as a result of European invasion (Aikhenvald, 2013). The same is true of the Caribbean. Earliest archeological remains in that region are in Cuba from about 7 kya and might be associated with a chert quarrying industry tied to Belize (Keegan, 2013). The next signs of movement into the Caribbean, judged by archeology, are immigrations from the north of South America about 2.5 kya, and they continue from then (Keegan, 2013).

By about 12.5 kya, humans were in the high Andes: at over 3,000 m in Chile and Argentina and over 4,000 m in Peru (Rademaker et al., 2014). These high-altitude dates are several thousand years after humans arrived in coastal South America. The dates come from the collagen of large bones of camelids directly associated with stone tool workshop sites. The significance of large bones is that they are less likely to have sunk to layers lower than those in which they first fell.

Later, in the middle of the last millennium, the Incas provide an example of colonization by migration of people as opposed to slow demographic expansion. Within a century, they conquered a multiethnic region 4,000-km long controlled with over 20,000 km of paths and roads, most of them paved (Beresford-Jones and Heggart, 2013). Furthermore, the control apparently involved relocation of entire native communities for both economic and defense purposes. A little later, the Spanish did the same, moving tens of thousands to work in and around the Bolivian silver mines (Beresford-Jones and Heggart, 2013).

Six hundred years before the Spanish arrival, the Wari of Peru were trading products from nearly 1,000 km away (Beresford-Jones and Heggart, 2013). The huge cities then and later—a Wari one of about 15 km^2, for example—also indicate great movements of peoples, forced or voluntary, into main centers of production, just as happens today (Currie and Mace, 2012).

Movement east from eastern Beringia into the north of North America had to wait for the near disappearance of the northern ice cap, which extended south to Washington in the west and New York in the east. That movement seemingly did not happen until about 5 kya, with the diaspora of the Eskimo–Aleut peoples from first Siberia and then Alaska across to Newfoundland and west coast Greenland, while the Inuit occupied Alaska (Reich et al., 2012; Fortescue, 2013; Rasmussen et al., 2010).

Oceans

By the time that humans had reached northern Siberia 32 kya, they were beginning to move across the Pacific north and east of New Guinea. They had reached the Bismarck archipelago by then, east of New Guinea (Wickler and Spriggs, 1988). Assuming island hopping along the way, the longest sea crossing would have been about 75 km.

Not until maybe 4,000 years ago did they get any farther. Then, they rapidly expanded across the western Pacific, north into Micronesia, and southeast as far as Fiji in southeastern Melanesia by about 3,200 years ago

(Kirch and Kahn, 2007; Wollstein et al., 2010; Wilmshurst et al., 2011). Fiji is about 1,000 km east of New Caledonia and 800 km east of New Hebrides. Large ocean-going seacrafts were necessary for such a voyage.

As in the Americas, an unexploited fauna and flora could well have allowed the rapid expansion, perhaps especially a fauna unused to being hunted (in the case of many birds, flightless). Archeologically, this expansion is indicated by finely decorated Lapita pottery (Spriggs, 2013). The name comes from the site in New Caledonia halfway between Australia and Fiji, where it was first identified.

A little under 2,000 years ago humans reached Tahiti, and then Hawai'i by maybe 1,500 years ago or maybe even as late as 800 years ago (Wilmshurst et al., 2011). They got to New Zealand 750 years ago from the Cook Islands 2,500 km to the northeast (Oppenheimer, 2003; Wilmshurst et al., 2011). Given the length of the voyage to New Zealand, a figure of only about 70 founding females (Murray-McIntosh et al., 1998) might not be surprising but even so, implies several successful crossings along with the degree of variation in mtDNA in the teeth and bones of probable founding individuals (Knapp et al., 2012).

Taiwan seems to be a main origin of the dispersal across the Pacific given that it is the current center of origin of 9 of 10 of the main Austronesian (essentially western and island Pacific) languages (Gray et al., 2009; Blust, 2013). Archeology substantiates that view (Bellwood, 2013). Genetically, a plausible scenario is movement of peoples from Taiwan to eastern Indonesia, perhaps through the Philippines, and thence out into the Pacific. However, the story is complex, with, for instance, different Asian origins for different areas of the Pacific (Cox, 2013) and some contrasting results from Y-chromosome and mtDNA analyses (Kirch and Kahn, 2007; Kayser, 2010; Cox, 2013).

At about the same time as humans were moving across the last stretches of the Pacific, they were crossing the Indian Ocean. The first successful settlers in Madagascar might have arrived 2,000 years ago (Dewar et al., 2013), even if the first signs of presence are from 3,400 years ago (Gommery et al., 2011). Genetic studies indicate equal contributions to the Malagasy gene pool from Africa and Indonesia (Hurles et al., 2005; Razafindrazaka et al., 2010; Cox et al., 2012). Even so, the number of Indonesian founding females in Madagascar might have been one-half the number of female founders of New Zealand: about 30 instead of 70 (Cox et al., 2012).

The Indonesian contribution to the origins of the people of Madagascar is strongly supported by the fact that the language of Madagascar, Malagasy, is mainly Bornean (indeed, from the Barito region of southern Borneo). However, not only were the Barito peoples seafarers, but also, Malagasy has several other Indonesian sources as well as some Malaysian roots as if the original contingent of immigrants boarded from a variety of the islands west of Borneo (Serva et al., 2012). The most likely route is not directly across the Indian Ocean but coastal and probably by coastal trading (Anderson, 2013).

Modern Migrations and Their Barriers

Movements of genes and peoples continue to this day, of course, in response to repression and opportunity. China, for instance, has seen massive migrations. Some have been organized by the Chinese government; others, especially in the surrounding regions, have occurred in response to events in the regions (LaPolla, 2013). These movements include migrations of tens of thousands into the cities. Indeed, as already indicated, large flourishing cities are some of the most ethnically dense regions, at least in the United States (Ottaviano and Peri, 2005).

Despite such movements, geographic and cultural barriers operate now as in the past. The barriers exist even in western Europe, with its dense transport network (Barbujani and Sokal, 1990; Novembre et al., 2008). Populations either side of mountains or waters are genetically distinct. Indeed, culture itself or language itself is often a barrier because of both difficulty of communication and also, to coin a word, “xenantipathy” or dislike of strangers (Rabbie, 1992). Thus, a map of the distribution of genotypes across Europe is almost identical to the geographic map of Europe (Novembre et al., 2008). Similarly, across a sample of the Solomon Islands of the Pacific, insular differences in peoples’ anatomy correlated more strongly with linguistic differences between the islands than with geographic distance between them (Dow et al., 1987).

Sex-Specific Phylogeography

So far, I have written as if everything phylogeographic that one of the human sexes did, so also did the other. That uniformity is not necessarily the case.

The *Mayflower*, one of the first ships from Europe to bring future residents to the future United States, apparently carried three times as many men as women (Johnson, 1994–2013). Among migrants and traders, that ratio is usual, with the result that Y-chromosome genotypes are sometimes more widely spread than mtDNA genotypes (Tumonggor et al., 2013). The result is evident in the Pacific, for example (Hage and Marck, 2003), and in the widespread presence of Mongol Y-chromosome genes through Asia and eastern Europe, the legacy of the Khan invasions of the 13th and 14th centuries CE (Zerjal et al., 2003).

However, among agriculturalists, males are usually the landowners and hence tend to be the resident sex, whereas females move in marriage. The result is regional differentiation in Y-chromosome genotypes but regional ubiquity of mtDNA genotypes (Forster and Renfrew, 2011). Furthermore, because a language of a people tends to be that of the invading dominant sex, male, a frequent outcome is a match between regional differentiation in language and Y-chromosome genotype but a mismatch between language and mtDNA genotypes (Forster and Renfrew, 2011).

Iceland epitomizes the contrast (Barbujani and Sokal, 1990; Helgason et al., 2001). All Icelanders speak Icelandic, which is essentially Scandinavian (indeed, originally Norwegian). Scandinavia, especially Norway, is also the genetic origin of the people of eastern Iceland and the males of western Iceland. However, western Iceland's mtDNA profile (i.e., female genetic profile) is quite largely of Gaelic British origin. The first millennium's Viking raids into western Britain and the abduction of women from there to western Iceland would explain that pattern.

Phylogeography of Human Commensals

I have so far indicated how archeology, genetics, and linguistics elucidate human phylogeography and in so doing, indicated how genetics and culture can determine human phylogeographic patterns. Human phylogeography uses an additional source of information—the phylogeography of the organisms that live with us, on us, and in us.

The global phylogeography of several organisms responsible for human disease and infection matches that of their human hosts (Morand, 2012). For instance, the human malaria organism, *Plasmodium falciparum*, is most genetically diverse in Africa, loses genetic diversity the farther that it is from Africa, loses similarity to African *P. falciparum* the farther it is from Africa, and probably left Africa 60–50 kya (Tanabe et al., 2010). The same pattern occurs in *Helicobacter pylori*, the human stomach bacterium (Falush et al., 2003b). Its phylogeography not only maps onto human immigration into the Americas more than 12 kya but also, maps onto the Neolithic introduction of farming into Europe, the Bantu expansion within Africa, European immigration into the Americas, and the slave trade from Africa into the Americas (Falush et al., 2003b) as well as onto the spread of humans across the Pacific (Moodley et al., 2009).

The diaspora across the Pacific of the paper mulberry *Broussonetia papyrifera*, a lizard *Lipinia noctua*, pigs, and the Pacific rat *Rattus exulans* as judged genetically must also be a result of the diaspora of humans, because none of these terrestrial species spread over the ocean on their own (Austin, 1999; Matisoo-Smith and Robins, 2004; Larson et al., 2007b; Chang et al., 2015). The phylogenetics of the lizard even show the initial human colonization of the western Pacific, the long pause there, and then, the rapid expansion across the central and eastern Pacific (Austin, 1999).

Remains of the Pacific rat, both its bones and the seeds that it gnawed, show human arrival in New Zealand in 1280 CE (Wilmshurst et al., 2008). Additionally, as a final example for the Pacific, sweet potato genetics reveal human travel from South America to Polynesia (Roullier et al., 2013), probably by returning Polynesians.

Animal phylogeographic genetics indicate human movement to and through Europe as well. Land snails moved from Iberia to Ireland (Grindon and Davison, 2013), mice moved from Norway into Britain (Searle et al., 2009), and pigs moved into Europe from Turkey by both inland and coastal routes (Larson et al., 2007a).

CULTURAL DIVERSITY AND THE ENVIRONMENT

The aspect of phylogeography that I have addressed so far is the spread of a species from its region of origin. As other species spread and subsequently, geographically differentiate, so do humans. Some of the differentiation is, in effect, random genetic drift unrelated to the nature of the environment. In modern terms, "genetic drift," is how Darwin explained regional differences in the human form, because he saw no correlation between environment and anatomy (Darwin, 1859, chap. 11). Of course, we now know that the environment, in fact, strongly influences our anatomy and physiology as many anthropological text books describe.

Far less well known is another aspect of human phylogeography, namely the relation between the nature of the environment and the geographic distribution of cultures (Harcourt, 2012, chaps. 5 and 6). The phylogeographic pattern of diasporas from several single origins—northeast Africa to Eurasia, southeast Eurasia to Australia, and northeast Siberia to the Americas—allows the equivalent of a repeat experiment on the relation between geography and phylogenetic and cultural diversity. In this section, I review evidence for an influence of the environment on cultural diversity, indeed evidence that the environment influences cultural diversity in the same way as it influences biological diversity or taxonomic diversity.

Tropical Diversity

Humans originated in Africa and eventually spread across much of the world. We did not spread evenly. Throughout the world, in Africa, Eurasia, and the Americas, tropical latitudes are more culturally diverse than nontropical ones, regardless of whether the data are number of hunter–gatherer cultures or number of languages (Harcourt, 2012, chap. 5). Most plants and nonhuman animal taxa also show this latitudinal pattern of greater tropical diversity (Hillebrand, 2004). More than 30 biogeographic hypotheses exist for the relation, almost none of which originated with analysis of human cultures (Lomolino et al., 2010, chap. 15). Nevertheless, a long-argued environmentally based hypothesis seems to explain the latitudinal distribution of human cultures, at least in part.

The main facts behind the hypothesis include the well-known year-round high productivity of the tropics where water is sufficient along with the less well-known fact of smaller geographic ranges of tropical cultures, hunter–gatherer cultures, and languages (Harcourt, 2012, chaps. 5 and 6). The hypothesis then is that the high year-round tropical productivity allows a population of sustainable size to persist in a small area, a consequence of which is dense packing of cultures (Nettle, 1998; Cashdan, 2001; Collard and Foley, 2002; Moore et al., 2002; Currie and Mace, 2012; Harcourt, 2012, chaps. 5 and 6). The packing could come about by overlap of large ranges. However, human cultures are territorial. Ranges do not overlap, except at high latitudes, where the necessarily large ranges are impossible to defend (Harcourt, 2012, chap. 6). The gradient of cultural diversity with latitude and the matching gradient of range size with latitude are not unique to a tropical–nontropical comparison. It occurs within North America too (Mace and Pagel, 1995; Harcourt, 2012, chap. 5).

Another hypothesis for regional contrasts in density of cultures is contrast in diversity of habitat (Mace and Pagel, 1995; Maffi, 2005), which is the case for diversity of species (Lomolino et al., 2010, chap. 15). It is certainly easy to imagine cultures specializing in exploitation of particular habitats and therefore not extending outside those habitats (Codding and Jones, 2013). Indeed, cultures may deliberately separate themselves by exploiting different habitats or exploiting them differentially (Milton, 1991).

This hypothesis and the other hypotheses to explain gradients in diversity are largely correlational. However, humans have themselves in effect performed a phylogeographic experiment to test the productivity hypothesis. They have intensified productivity in cities—in the form of jobs and hence salaries—that allows purchase of food in a concentrated area instead of its production over an extended area. If level of productivity determines phylogenetic diversity and if phylogeographic concentrations are determined by concentrations of high productivity, rich cities should by one measure of phylogeographic diversity (cultural diversity) be some of the most phylogeographically diverse regions (Ottaviano and Peri, 2005).

They are, at least in the United States. Thus, change over time in wages and jobs significantly matched change in diversity of languages in 160 metropolitan areas in the United States over the analyzed 20 years from 1970 controlling for a variety of factors, such as size of the city, sex ratio of workers, and level of schooling (Ottaviano and Peri, 2005). As a specific example of the effect, over 300 languages might be spoken in one of the world's largest cities, London (Currie and Mace, 2012).

Conclusion

If productivity affects human cultural diversity, it is far from the only influence (Nettle, 1998; Cashdan, 2001; Collard and Foley, 2002; Moore et al., 2002; Manne, 2003). Many other factors can affect diversity, including history. Tropical South America, for example, has an unusually low density of cultures by comparison with tropical Asia and Africa. One reason might be the relatively short period that humans have been in the American tropics. However, another could be the devastating effect of the Old World diseases (smallpox, for example) brought to the New World by invading Europeans from the 15th century on (McNeill, 1998, 2010). The consequent mass mortality of Native Americans and Amerindians provides a foretaste, perhaps, of both the current global phylogeographic interaction of disease organisms and their human carriers and recipients and also the disappearance of indigenous languages as majority cultures expand (Nettle and Romaine, 2000; Harrison, 2007; Harcourt, 2012, chap. 8; Harcourt, 2015, chap. 12; Lewis et al., 2016).

ACKNOWLEDGMENTS

I thank John Avise and Francisco Ayala for inviting my contribution to the "In the Light of Evolution X: Comparative Phylogeography" Colloquium; John Avise, Francisco Ayala, and Brian Bowen for organizing the colloquium; and Kelly Stewart and two anonymous referees for most helpful commentary on my paper.

Part IV

COMPARATIVE PHYLOGEOGRAPHY IN A CONCEPTUAL SENSE

One final dimension in which phylogeography can be considered "comparative" is through its interactions with various other biodiversity disciplines. This general topic has already been addressed by several of this colloquium's authors. For example, Bowen et al. (Chapter 1) explicitly link modern marine phylogeography to traditional biogeographic perspectives on biodiversity in the sea; Wakely et al. (Chapter 5), Mehta et al. (Chapter 6), and others relate phylogeography to historical population demography, coalescent theory, and traditional population genetics; Edwards et al. (Chapter 9) explicitly weighed and contrasted comparative phylogeography with molecular phylogenetics and phylogenomics; Heintzman et al. (Chapter 13) and Prates et al. (Chapter 3), among others, advocated the incorporation of more paleoecology into phylogeography; and several authors touched upon the special relevance of phylogeographic findings for the field of conservation biology. The final chapter in this book provides one further example of how comparative phylogeography relates to yet another emerging biodiversity field.

The term "landscape genetics" has blossomed in recent years into a recognizable and popular research arena, purportedly forming a needed bridge between landscape ecology and population genetics (much as the field of phylogeography attempted to bridge phylogenetics and population genetics). In Chapter 16, Leslie Rissler recounts the relatively recent (as well as deeper) conceptual histories of both phylogeography and landscape genetics, strives to highlight different empirical and intellectual emphases in these two fields, and speculates on some major research areas ripe for further investigation in the emerging era of genomics. She concludes by advocating a union of comparative phylogeography and landscape genetics under the broader umbrella of biogeography, which has always sought to connect ecology and evolutionary biology. The union of comparative phylogeography and biogeography is a theme that emerged in the Chapter 1, was elaborated throughout this book (Chapters 8 and 11, and elsewhere), and is eloquently summarized in this final contribution.

16

Union of Phylogeography and Landscape Genetics

LESLIE J. RISSLER

Phylogeography and landscape genetics have arisen within the past 30 years. Phylogeography is said to be the bridge between population genetics and systematics, and landscape genetics the bridge between landscape ecology and population genetics. Both fields can be considered as simply the amalgamation of classic biogeography with genetics and genomics; however, they differ in the temporal, spatial, and organismal scales addressed and the methodology used. I begin by briefly summarizing the history and purview of each field and suggest that, even though landscape genetics is a younger field (coined in 2003) than phylogeography (coined in 1987), early studies by Dobzhansky on the "microgeographic races" of *Linanthus parryae* in the Mojave Desert of California and *Drosophila pseudoobscura* across the western United States presaged the fields by over 40 years. Recent advances in theory, models, and methods have allowed researchers to better synthesize ecological and evolutionary processes in their quest to answer some of the most basic questions in biology. I highlight a few of these novel studies and emphasize three major areas ripe for investigation using spatially explicit genomic-scale data: the biogeography of speciation, lineage divergence and species delimitation, and understanding adaptation through time and space. Examples of areas in need of study are highlighted, and I end by advocating a union of phylogeography and landscape genetics under the more general field: biogeography.

Division of Environmental Biology, Directorate of Biological Sciences, National Science Foundation, Arlington, VA 22230. Email: lrissler@nsf.gov.

Two similar fields have arisen within the past 30 years: phylogeography (Avise et al., 1987) and landscape genetics (Manel et al., 2003). Both fields, to varying degrees, integrate theory and methods from population genetics, biogeography, and ecology. It is my contention that these fields, though somewhat independent in terms of history, methods, scale, and application, do not address fundamentally distinct questions. As such, ecologists and evolutionary biologists who combine genealogical and spatially explicit approaches to the study of biodiversity patterns and processes would benefit from a conceptual unification of phylogeography and landscape genetics. Both fields can be considered as an amalgamation of classic biogeography and genetics. I begin by briefly summarizing the history of each field, especially landscape genetics because phylogeography has been discussed at length elsewhere (e.g., Hickerson et al., 2010), and highlighting a few examples typical of each field. I then summarize the supposed differences between them and move on to a discussion of what I consider the fundamental questions that fall under a general biogeographic framework encompassing both phylogeography and landscape genetics. I end by advocating for a unification of the fields and point out basic areas of inquiry that are ripe for future investigation by biogeographers at all spatial and temporal scales.

A HISTORY OF THE FIELDS OF PHYLOGEOGRAPHY AND LANDSCAPE GENETICS

Phylogeography (including comparative phylogeography) and landscape genetics are considered separate fields: But are they? One of the earliest papers examining phenotypic variation (albeit not genetic variation, although this was assumed) in an explicitly spatial way was by Epling and Dobzhansky (1942), who focused on the “microgeographic races” of *Linanthus parryae* in the Mojave Desert of California. The distribution of the relative frequencies of blue and white flowers were mapped every half mile along a 200-mile roadway as well as every 25 feet along a 750-foot transect. These data were used to test Wright’s shifting balance theory of evolution (Wright, 1931, 1937), in particular, whether random drift could act in large populations distributed over uniform environments. Lewontin et al. (2003) in their synthesis of Dobzhansky’s works note that, in personal correspondence between Sewall Wright and Theodosius Dobzhansky, Wright told Dobzhansky how to analyze the data to test for isolation by distance (IBD) before Wright had even published a formal statistical model for IBD. Wright later used the *Linanthus* data in his development of *F*-statistics (Wright, 1943); these metrics are central to describing the structure of genetic variation within and among natural populations and can be used to infer demographic history. More recently, they have been used to pinpoint genomic regions under strong selection (reviewed in Holsinger and Weir, 2009). Thus, this paper (Epling and Dobzhansky, 1942) can be considered the birth of landscape genetics although, at the time of Wright, "landscape" rarely meant a literal landscape but rather a figurative one describing the relation between the genotypes of individuals and their fitness or the relation between the allele frequencies in a population and its mean fitness (reviewed in Coyne et al., 1997). That said, those blue and white flowers of the *Linanthus* system studied by Epling, Dobzhansky, and Wright helped the leaders of the Modern Synthesis to merge theory with empirical data and spawned many studies investigating evolutionary processes (reviewed in Schemske and Bierzychudek, 2001, 2007; Lewontin et al., 2003). Putting those processes on a map and understanding how selection, drift, and gene flow operate on a changing landscape, varying across space and time, remain a focus of evolutionary biology today.

Another paper that presages the fields of phylogeography and landscape genetics is Dobzhansky et al. (1966) on the geographic distribution of chromosomal inversions in *Drosophila pseudoobscura* in the western United States. That 1966 paper summarizes Dobzhansky's earlier work in the 1940s and 1950s on the system and compares temporal and spatial patterns among populations of gene arrangements on the third chromosome. Surprisingly, the frequencies of different inversions showed a consistent geographic break that separated populations in the Pacific Coast states (e.g., California, Oregon, and Washington) from those in the Great Basin, Rocky Mountains, and Texas (Fig. 16.1). This biogeographic pattern was not the focus of the paper, but, today, due to comparative phylogeographic and biogeographic studies across the western United States, we recognize this region as a major suture zone (i.e., an area where multiple hybridizing species and/or phylogeographic lineages come into contact) (Anderson, 1948, 1953; Remington, 1968; Swenson and Howard, 2004, 2005; Rissler and Smith, 2010). Remington, who first mapped suture zones across North America, was strongly influenced by Dobzhansky's ideas on reinforcement during speciation (Dobzhansky, 1940), and later researchers like Hewitt studying species in Europe (e.g., Hewitt, 1999, 2001), Avise in the southeastern United States (Avise, 2000), and Moritz et al. (2009) in the Australian Wet Tropics rainforest also suggested that shared suture zones were ideal natural laboratories to study the processes driving divergence, reproductive isolation, and speciation. Unknown for most suture zones is whether their location is due to strong selection or ecological processes arising from abiotic conditions that vary between physiographic areas, or to historical processes as midpoints between glacial refugia where secondary contact between lineages resulted in either introgression or reinforcement (e.g., Swenson, 2010). In fact, the last sentence in Dobzhansky et al.'s 1966 paper was, "The causation of the genetic changes observed remains

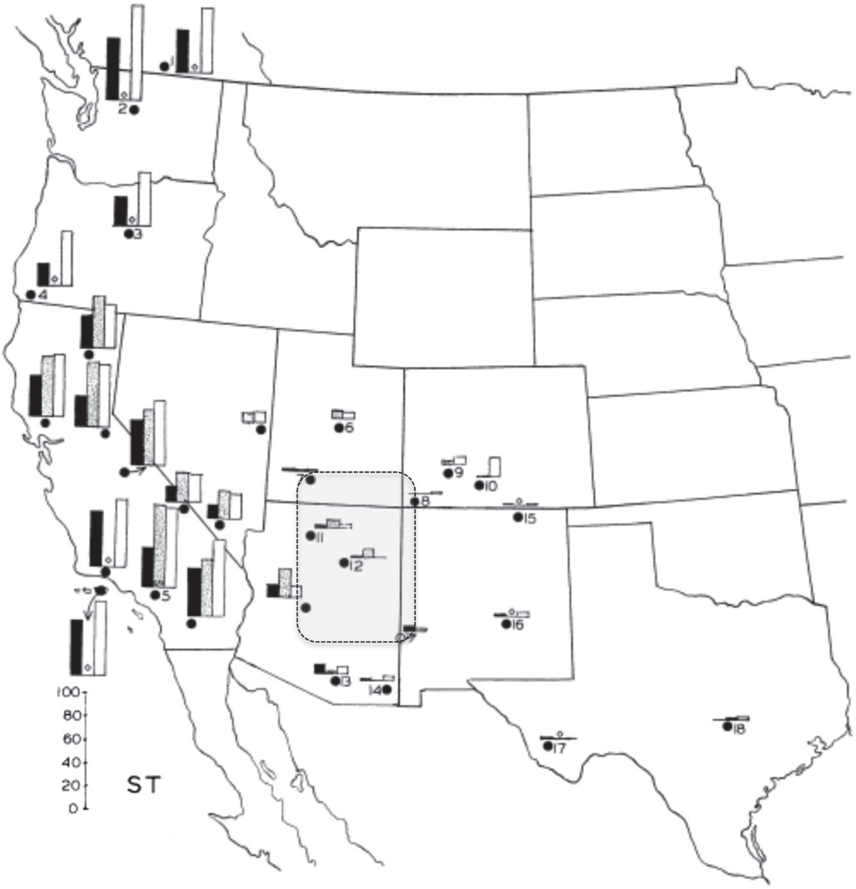


FIGURE 16.1 An early "landscape genetics" study by Dobzhansky et al. (1966) where the spatial and temporal distribution of *D. pseudoobscura* chromosome types was mapped across the western United States. ST, standard gene arrangement of the third chromosome. Black bars represent the frequency of the ST chromosomes in ~1940, stippled in 1957, and white in the period 1963–1965. The dotted black rectangle is the approximate location of a suture zone later predicted by Remington (1968) and confirmed by Rissler and Smith (2010) and Dyer (2015) separating populations in the Great Plains from populations in the Pacific Coastal regions (figure without suture zone is figure 1 from Dobzhansky et al., 1966).

problematic." It is in answering this question and similar questions about the relative influence of ecological and evolutionary processes on biogeographic patterns that a union of phylogeography and landscape genetics has much to offer.

Phylogeography derives from a synergy between systematics (specifically, molecular phylogenetics and population genetics) because phylogeographers study the geographic distributions of genealogical lineages, rather than species, per se. This integration was possible because of the technical developments of the past half-century that allow the measurement of genetic variation at fine spatial scales, along with the growing realization that the diversification of species can be explained by population-level evolutionary processes (reviewed in Coyne and Orr, 2004). Landscape genetics, on the other hand, grew up mostly as a subdiscipline of ecology and macroecology, with the goal of tracking the movement of organisms across space and over generations, but not deep time (Manel et al., 2003). The proliferation and ubiquity of molecular genetics created an opportunity for ecologists (and conservation biologists) to incorporate population genetics into their studies of the distribution and abundance of organisms. This union spawned landscape genetics (Manel et al., 2003; Storfer et al., 2007; Holderegger and Wagner, 2008; Belkenhol et al., 2009; Sork and Waits, 2010; Dyer, 2015).

Although the use of molecular genetics within ecology goes back decades, the moniker "landscape genetics" was not coined until 2003 (Manel et al., 2003). The history of the field is one that looks primarily to ecology and biogeography, rather than evolution and the Modern Synthesis, as sources of inspiration (but see Dyer, 2015). As Manel et al. (2003) pointed out, landscape genetic studies focus on individuals (in preference to populations, lineages, or species) to find genetic discontinuities in space, which are then correlated with landscape or environmental features. This phylogeny-free view, at least in the early days, was distinct from phylogeography and comparative phylogeography (Bermingham and Avise, 1986; Bermingham and Moritz, 1998; Avise, 2008; Gutierrez-Garcia and Vazquez-Dominguez, 2011). Most landscape genetic studies focus on explaining contemporary rather than historic causes of intraspecific genetic variation (Storfer et al., 2010; Wang, 2010; but see Bohonak and Vandergast, 2011) although this focus is a recurring problem for the field (known as the "time lag" problem) because genetic variation may not be reflective of very recent changes in population size or connectivity (reviewed in Epps and Keyghobadi, 2015).

It is only recently and rarely that researchers have attempted to combine phylogeography and landscape genetics, specifically to account for historical processes when explaining responses of organisms to contemporary ecological perturbations. For example, Swaegers et al. (2014, 2015), studying genetic diversity and structure of the damselfly *Coenagrion scitulum,* used a combination of mitochondrial gene sequencing, microsatellite analyses, and single-nucleotide polymorphisms across the entire species' range to disentangle the role of historical range expansions from the effect of recent range shifts due to global warming. Their multilocus polygenic analysis identified candidate loci potentially under thermal selection, and their studies highlight the usefulness of integrating genomic, phenotypic, and environmental data to disentangle historical and contemporary evolutionary and ecological processes.

THE WAYS IN WHICH PHYLOGEOGRAPHY AND LANDSCAPE GENETICS ARE THOUGHT TO DIFFER

Both phylogeography and landscape genetics fall squarely within the realm of biogeography—"the science that attempts to document and understand spatial patterns of biological diversity" (Lomolino et al., 2010). Species numbers can increase due to both dispersal into an area and in situ speciation. Therefore, phylogeography, with its emphasis on geographic patterns of genealogical lineages (often incipient species), and landscape genetics, with its emphasis on dispersal and gene flow across space, are two sides of the biogeography coin. Today, we know that explaining patterns of genetic variation and organismal traits is both a historical (evolutionary) and an ecological question and an ecological question that lies at the junction of major microevolutionary and macroevolutionary disciplines (summarized in Avise, 2000). Unifying population genetics with explicitly spatial dimensions of time and space is what landscape genetics and phylogeography are all about. It is the recognition that the same ecological and evolutionary processes that cause lineage divergence can also drive speciation that links population genetics and ecology with systematics and biogeography. Thus, the fields of phylogeography and landscape genetics differ not because the questions they address are fundamentally different (Dyer, 2015), but because they arose from different subdisciplines in biology, with one emphasizing history and one ecology. Their union is inevitable as molecular genetics further permeates all of the biological sciences, enabling us to distinguish ancient and contemporary influences on the genome and the resulting consequences for the phenotype, populations of organisms, and their divergence (i.e., speciation). Because of these differences in emphases and disciplinary roots, phylogeography and landscape genetics are considered to differ in three major ways: temporally, methodologically, and spatially.

Temporal Scale

The temporal scale is one of the more common ways of differentiating the fields, with phylogeography examining timescales on the order of millions of years (more typical of lineage divergence and speciation events), but landscape genetics examining more shallow timescales closer to thousands of years BP. However, the advent and proliferation of next-generation sequencing and nonmodel organism genomics now allow evolutionary biologists to investigate both neutral and adaptive genetic diversity at a variety of temporal scales (Bohonak and Vandergast, 2011; Epps and Keyghobadi, 2015). Both historic and contemporary environmental, geographic, and ecological factors influence patterns of genetic variation (e.g., Garrick et al., 2008; Duncan et al., 2015) so the artificial separation of two fields by these criteria creates a situation in which ecologists (who more often focus on current patterns of diversity) and evolutionary biologists (who more often focus on historic mechanisms of diversity) fail to integrate their research (e.g., Knowles and Alvarado-Serrano, 2010; He et al., 2013).

Molecular Markers and Methodology

At the time of phylogeography's inception (the mid-1980s), the state of the art for genetic analysis was mtDNA. This material was useful for intraspecific studies, but mtDNA has many peculiarities (e.g., incomplete lineage sorting between hybridizing taxa, predominately maternal inheritance, ease of introgression compared with nuclear genes, etc.) that limit its relevance to speciation research. In contrast, landscape genetics studies tended to rely on microsatellite data. The growing use of multilocus and single-nucleotide polymorphisms (SNPs) will help bridge these fields. Modern phylogeography uses all manner of genetic diversity data to ascertain relationships between lineages and geography, and landscape genetics uses highly variable genetic data to test the relative influence of landscape composition and configuration, most often in contemporary settings (Storfer et al., 2007). Therefore, the choice of marker is no longer a useful way to distinguish the fields.

Although a detailed discussion of analytical methods used in the two fields is beyond the scope of this chapter (Storfer et al., 2007), I note that phylogeographic studies tend to emphasize genealogical lineages with a more explicit appreciation of history whereas landscape genetics studies focus on spatially explicit population genetic analyses, with little regard for genealogical relationships of closely related lineages or species (but see Vandergast et al., 2007; Koscinski et al., 2009; Swaegers et al., 2014, 2015). As our ability improves to (*i*) model past environmental and landscape changes, (*ii*) describe the tree of life, and (*iii*) estimate demographic responses of populations and lineages to spatial and temporal perturbations, the union of phylogeography and landscape genetics will strengthen—one more reason to consider these fields under the single rubric of biogeography.

Space: Which Landscape?

The "adaptive landscape" from the Modern Synthesis is essentially geography-free. Phylogeography, on the other hand, is as much about understanding the physiography of an area as the process of lineage diversification, so the "landscape" is a particular region of interest, like California (e.g., Lapointe and Rissler, 2005), or watersheds in the eastern United States (e.g., Avise et al., 1987), or mountains in Australia (Moritz et al., 2009), or the North Atlantic intertidal (e.g., Wares and Cunningham, 2001). The landscape for landscape genetics studies is also any geographic space, albeit usually at a much smaller scale, encompassing only a portion of any species' range. Typical landscape genetics studies are less interested in physiography, history, climate, and geology and more focused on a particular species' or population's dispersal across space and how those factors influence movement and gene flow. These spaces are often distilled into distance matrices that can be ecological (e.g., environmental climate data—past, present, or future), habitat types (e.g., wetlands, agricultural fields, roads), or any other data layer that represents how an organism may disperse across areas that are more or less permeable to movement. Landscape genetics studies often assess the level of isolation by distance (IBD) (Wright, 1943), isolation by environment (IBE) (Wang and Summers, 2010; Wang and Bradburd, 2014), or isolation by adaptation (IBA) (reviewed in Wang, 2013). Fine-scale ecological data are more likely to provide clear association among ecology, geography, and local adaptation, if it exists, because the spatial scale of genetic variation matches that of the landscape of interest, thus linking ecological and evolutionary pattern and process (e.g., Bermingham and Avise, 1986; Manel et al., 2010; Eckert and Dyer, 2012).

EXAMPLES OF RECENT ADVANCES IN UNDERSTANDING THE BIOGEOGRAPHY OF LIFE

Biogeography, broadly construed, is an integrative science that spans vast temporal and spatial scales encompassing the purview of both phylogeography and landscape genetics. Recent advances in theory, models, and methods have allowed researchers to better synthesize ecological and evolutionary processes in their quest to answer some of the most basic questions in biology, such as why species have range limits in the absence of geographic barriers. For example, niche theory (Chase, 2011) and ecological niche modeling (e.g., Elith et al., 2006; Kozak et al., 2008; Warren, 2012; Alvarado-Serrano and Knowles, 2014) have been responsible for an exponential increase in studies that use Geographic Information Systems (GIS) and various algorithms (e.g., Phillips et al., 2004, 2006; Elith et al., 2011) to quantify the relationship between species' distributions and associated environmental parameters (past, current, or future) and project it onto geographic space. Ecological niche models (ENMs) use as input georeferenced specimen data, typically from natural history collections (e.g., www.gbif.org/), and spatially explicit environmental data, often from publically accessible sites like WorldClim (www.worldclim.org/; Hijmans et al., 2005), to examine where organisms should be distributed based solely on the input data, that is, without biogeographic barriers, ecological interactions, or adaptation. Additional investigations using reciprocal transplants (e.g., between allopatric sister species) can illuminate whether ecogeographic isolation mirrors geographic separation, providing insight into the relative roles of ecological adaptation and divergence vs. niche conservatism (discussed in *Divergence, Speciation, and Species Delimitation*) (Sobel et al., 2010). These kinds of studies that mesh niche theory and modeling are relevant to a vast array of questions in biogeography, ecology, evolutionary biology, and even systematics (e.g., Rissler and Apodaca, 2007).

Another basic question in biogeography that has recently benefited from an interdisciplinary approach is the following: Why are there more species in the tropics? Wiens and coworkers in several publications (e.g., Kozak and Wiens, 2007; Wiens et al., 2009; Hua and Wiens, 2010; Pyron and Wiens, 2013; Chejanovski and Wiens, 2014) have highlighted the importance of investigating the three mechanisms responsible for increased species richness in a region—increased speciation rates, decreased extinction rates, and/or increased dispersal rates (Ricklefs, 1987, 2004; Wiens and Donoghue, 2004). Pyron and Wiens (2013) used phylogenetic comparative methods on a tree of 2,871 species of amphibians and found that tropical regions had higher speciation rates and lower extinction rates, which were strongly linked to ecological factors such as climate. Even within just the temperate zone, Chejanovski and Wiens (2014) found that areas with higher species richness had species with narrower climatic niches, which follows the hypothesis of Janzen (1967), where regions with reduced seasonality (specifically temperature) promote the evolution of species with narrow climatic tolerances. The explicit link among niche breadth, dispersal, and speciation rates (reviewed in Ghalambor et al., 2006) is centered on the idea that species with narrower niche breadths are less able to disperse across environmental gradients, resulting in increased opportunities for allopatric speciation. In summary, the increased availability of large-scale phylogenetic and phylogenomic data, global environmental data, georeferenced specimen data, and spatial statistics has allowed researchers to more robustly address fundamental and long-standing questions in ecology, evolution, and biogeography.

Additional advances have come in the realm of better and faster computer algorithms. For example, understanding current species distributions and genetic diversity requires accounting for variation in population sizes and migration rates across a species′ temporal and spatial range, such as in the studies of Knowles and Alvarado-Serrano (2010) and Brown and Knowles (2012). Through the use of approximate Bayesian computation (ABC) (Beaumont et al., 2002; Wegmann et al., 2010), they tested the likelihood of different dispersal and demographic scenarios [see He et al. (2013) for an example using the Australian lizard *Lerista lineopunctulata*], resulting in explanations for current population genetic patterns. He et al. (2013) argue that this approach can result in completely different conclusions than those from more common correlative approaches in landscape genetics, like those of IBD or associated matrices that do not consider the temporal variation in habitat suitability. Such variation can affect population sizes and dispersal patterns resulting from organisms tracking environmental shifts due to niche requirements or moving out of glacial refugia during climate change.

Approaches that consider demographic history have recently been expanded to community-level assemblages (Ghalambor et al., 2006). Again, the advance is mainly a technological one using computationally intensive model testing [e.g., hierarchical approximate Bayesian computation (hABC)] and high-resolution genomic data to estimate demographic scenarios across multiple codistributed species. For example, Xue and Hickerson (2015) developed a new way to leverage single-nucleotide polymorphism datasets to estimate the extent of temporal synchronicity in range expansions across multiple taxa. But simply acknowledging that species interactions within a community can influence genetic diversity of co-occurring species (e.g., Criscione et al., 2006; James et al., 2011) is an important advance in biogeography. Hand et al. (2015) even advocate for a new field—"landscape community genomics"—which would explicitly consider both the biotic and abiotic factors that affect genomic variation. But the same basic questions in biogeography remain. For example, how stable are species associations through space and time? Have multiple species in a community responded similarly to past and current geographic and environmental conditions? Can we predict how species individually, and communities collectively, will respond to climate change? To better answer these questions, we need to integrate ideas, theory, models, and methods across ecology and evolutionary biology rather than balkanize and separate subdisciplines, as if understanding biogeography is possible without a holistic, interdisciplinary approach (e.g., Riddle et al., 2008).

ECOLOGICAL AND EVOLUTIONARY QUESTIONS THAT WOULD BENEFIT FROM A MORE HOLISTIC VIEW OF BIOGEOGRAPHY

Phylogeography is loosely analogous to historical biogeography (Arbogast and Kenagy, 2001), relying more heavily on history than current ecological conditions, and landscape genetics is loosely analogous to ecological biogeography, relying more heavily on the latter than the former. Modern biogeography, on the other hand, must necessarily consider both historic and current ecological and evolutionary processes. As such, all of the processes that drive the distribution and abundance of genes, organisms, traits, populations, species, and communities are the purview of biogeography. Below are three major, but related, topics that would benefit from synergy among ecologists, evolutionary biologists, and biogeographers.

The Biogeography of Speciation

The modern evolutionary synthesis of the late 1930s made it clear that understanding genetic variation within species is essential for understanding the evolution of higher taxa. Speciation is a predominately geographic phenomenon (Coyne and Orr, 2004). Understanding the relative roles of isolation, gene flow, divergence, and selection on the process of speciation remains a promising area of research. Geographically isolated populations are often viewed as incipient species (Mayr, 1963), and the extent of gene flow between populations is a central focus of landscape genetics. Although Avise in his classic text *Phylogeography* (Avise, 2000) has a section titled "Genealogical Concordance: Toward Speciation and Beyond," surprisingly few empirical studies in biogeography explicitly investigate speciation.

The biogeography of speciation is of relatively recent interest (e.g., Coyne and Price, 2000; Losos and Schluter, 2000). For example, Kisel and Barraclough (2010) quantified the speciation-area relationship by examining the probability of in situ speciation on islands of angiosperms, bats, birds, carnivorous mammals, ferns, lizards, butterflies and moths, and land snails. The minimum island size necessary to support speciation was found to scale linearly with the strength of gene flow across almost all of these groups. Snails have speciated on islands smaller than 1 km^2, but bats show no in situ speciation on any island except for Madagascar (>500,000 km^2) (Kisel and Barraclough, 2010). More recent studies show even stronger associations between island area and the probability of speciation by using phylogenetic information rather than taxonomy-based approaches (Igea et al., 2015). Despite extensive effort, there are still only a handful of cases of substantial genetic divergence in the face of strong gene flow (e.g., *Rhagoletis*, Crater Lake cichlids) and even fewer of divergence proceeding to full speciation (reviewed in Coyne, 2011). Accurate assessments of gene flow and phylogenies, increasingly made possible by molecular genetics, will play a crucial role in this endeavor. Spatial statistical methods common to landscape genetics (e.g., isolation by environment while controlling for isolation by distance) have been used to assess the relative influence of divergent selection vs. geographic isolation on speciation. Thus, comparing the relative influence of geography, history (isolation), and ecology (divergent selection) in speciation and biogeography remains an exciting and current question in evolutionary biology at the confluence of phylogeography and landscape genetics (e.g., Papadopulos et al., 2011, 2014; Nosil, 2012; Shafer and Wolf, 2013; Wang, 2013).

Divergence, Speciation, and Species Delimitation

The important role of ecology in evolution was recognized by both Mayr and Dobzhansky (reviewed in Sobel et al., 2010), but, as pointed out by Coyne and Orr (2004), it was not until the 1980s that new tools (e.g., molecular genetics, phylogenetic methods, comparative analyses) allowed molecular systematists to observe fine-scaled genetic differences within species and allowed evolutionary ecologists to apply knowledge of natural selection and biogeographic barriers in nature to questions about speciation. Today, there are many studies of ecological speciation (e.g., Hendry, 2009) defined as reproductive isolation driven by ecological selection, in the presence or absence of gene flow (Schluter, 2001, 2009; Sobel et al., 2010; Nosil, 2012). Importantly, landscape genetics has added many methods that allow one to examine isolation by distance using different metrics (e.g., resistance, geographic barriers, environmental factors, community interactions, and local adaptation) (Nosil et al., 2008; Shafer and Wolf, 2013; Wang, 2013; Wang et al., 2013; Papadopulos et al., 2014) although the link between these measures and broader ideas about how adaptation and natural selection in different environments drive speciation is less appreciated in the field (but see Papadopulos et al., 2014). That said, isolation due to local adaptation is something that links landscape genetics with work on speciation and is often suggested to be a signature of sympatric speciation (e.g., Nosil et al., 2008; Funk et al., 2011; Shafer and Wolf, 2013). But how different abiotic conditions must be to ensure divergent selection even in the face of gene flow, and at what spatial scale, is an open and interesting question (Nosil, 2008).

Ecology and geography may also play a role in speciation, not by causing divergent selection in separated populations, but by the inability of separated populations to adapt to intervening, unsuitable habitat due to strong "niche conservatism" (Wiens, 2004). One study that shows the power of combining phylogeographic analyses with spatially explicit climate data to address questions regarding speciation is that of Cadena et al. (2012). They examined 93 pairs of sister species (mammals, birds, amphibians, and reptiles) that were restricted either to the New World tropics or North Temperate Zone to understand whether niche conservatism or divergence was more likely to explain higher speciation rates in the tropics. They predicted that, if niche conservatism was more important, then they should see greater overlap of the thermal niches of sister species in the tropics, relative to sister species in the temperate zone. Alternatively, if ecological divergence was more important, then the greater climatic stratification along elevational gradients in the tropics would create sister species with less overlap in their thermal niche. Using the WorldClim database (Hijmans et al., 2005), the authors extracted climate data from over 33,000 georeferenced collection localities. They found that tropical sister species had narrower thermal regimes and were more evolutionarily conserved than sister species in the temperate zone, suggesting that populations in the tropics should experience more opportunities for isolation and subsequent speciation via niche conservatism. Recent modeling (Hua and Wiens, 2013) has also advanced our understanding of the link between climate and speciation by making specific predictions about when and where you might expect niche conservatism vs. divergence to be more important in speciation. These studies and future studies that investigate the relative importance of mechanisms (e.g., geographic isolation, strong natural selection, sexual selection) (reviewed in Richardson et al., 2014) capable of initiating and enhancing lineage divergence, at various spatial scales under different environmental conditions, will be key to a better appreciation of the role of ecology and geography in the origination of organismal and trait diversity.

Even systematics has been infused with more biogeography in recent years because of the ubiquity of spatially explicit environmental data and publicly accessible georeferenced specimen data from natural history collections. An early comparative phylogeographic analysis of California by Lapointe and Rissler (2005) took multiple species' phylogenies (bird, plant, insect, mammal, amphibian, and reptile) and developed a new statistical method to assess concordancy of multiple phylogeographic trees and synthesize the data into a regional supertree. They found that the geographic location of breaks between lineages was correlated with sharp changes in climatic conditions. Later ecological studies in the same geographic region found that the contact zones between parapatrically separated lineages of an endemic Californian salamander and potential incipient species were inhospitable for migrants and thus barriers to gene flow (Rissler and Apodaca, 2007). These kinds of phylogeographic and landscape genetics studies, which integrate geographic and ecological analyses, have spawned new methods in species delimitation that explicitly consider the extent of ecological divergence when diagnosing species (Raxworthy et al., 2007; Rissler and Apodaca, 2007; Leaché et al., 2009; Pelletier et al., 2015). Thus, understanding the role of ecological and evolutionary processes across a landscape can be illuminating to the study of both the processes and the resulting biogeographic patterns of speciation.

Understanding Adaptation Through Time and Space

Much current research is focused on understanding patterns of adaptive genetic diversity and the ecological and evolutionary processes that cause those patterns (Hancock et al., 2011; Schoville et al., 2012; Jones et al., 2013; Savolainen et al., 2013). Landscape genetics and genomics (Joost et al., 2007) have been touted as providing a unique perspective on local adaptation because space is considered explicitly (Holderegger and Wagner, 2008). A recent example by Vincent et al. (2013) studied gene–environment interactions driving adaptation in 54 North American populations of Atlantic salmon. They found that regional genetic structure and ecological structure (described by 49 environmental variables) were correlated, and, in particular, they noted a signature of thermal local adaptation linked to counter gradient selection imposed by growing season length. However, because landscape genomic studies are generally only correlative, additional studies are needed to confirm adaptation.

New spatial modeling methods have improved our ability to assess the geography of adaptation. For example, Fitzpatrick and Keller (2015), using balsam poplar as a case study, took methods more common to ecology—generalized dissimilarity modeling (Ferrier et al., 2007) and gradient forests (Ellis et al., 2012) [which are regression-based models that map turnover in biological composition (e.g., species) using nonlinear functions of environmental gradients]—and instead mapped the turnover of thousands of single-nucleotide polymorphisms. They found several threshold gene–environment relationships along temperature gradients (e.g., circadian clock gene *GIGANTEA-5*), suggesting strong local adaptation to temperature. These methods also can be used to forecast temporal changes in genetic composition due to climate change.

SPECIFIC AREAS IN NEED OF GROWTH TO ADVANCE INTEGRATION

Theory and Analytic Methods

Theory and analytic methods are underdeveloped aspects of the growing unification of ecology and evolutionary biology, given the diffusion of genetics into individual, population, community, and even ecosystem questions. The analysis of evolutionary processes should continue to better integrate spatial and ecological data (e.g., Fitzpatrick and Keller, 2015). True integration may require the development of new theory, as well as the development of new analytical methods (Petren, 2013). Populations comprising a single species can vary in multiple traits (e.g., demography, morphology, physiology, connectivity to other populations, genetic diversity, etc.) across a species' range, a condition that should be more formally integrated into theories and methods. Integrating theory of species range dynamics (Scheiner and Willig, 2011) with microevolutionary processes (genetic drift, mutation, gene flow, and natural selection) is challenging but would advance fields interested in understanding the distribution of organisms (e.g., biogeography, ecology, evolution, and conservation) (Vellend, 2010).

Data: What We Need, Where to Get It, and How to Add to It

A lack of sufficient data plagues many fields, but it is particularly insidious for biogeography because species are currently experiencing an extinction rate at least 1,000 times the background rate (De Vos et al., 2015). We are losing species before they are even formally described (the Linnean shortfall), and, for those that are described (<10% of species on Earth), we have little to no knowledge of their geographic distributions, and even less knowledge of life history characteristics and traits of populations, lineages, and species (the Wallacean shortfall) (Lomolino, 2004; Whittaker et al., 2005). Another impediment is a lack of fine-scale environmental and ecological data, which severely limits our understanding of the biogeography of life on this planet and how it may change given human-caused perturbations.

Large-scale mapping initiatives (e.g., The Global Map project, www.iscgm.org/index.html) have been ongoing for the past 50 years, but those focused on biodiversity data are only about a decade old (e.g., Global Biodiversity Information Facility, www.gbif.org). Linking this information with natural history information is an ongoing challenge but is being tackled by databases like iDigBio (www.idigbio.org), Morphobank (www.morphobank.org), and Phenoscape (www.phenoscape.org). Phylogenetic information is increasing through programs like the National Science

Foundation's Assembling, Visualizing, and Analyzing the Tree of Life (Avatol, www.avatol.org) and the newer Genealogy of Life, whose aim is an open-access tree of life that encompasses other information, such as traits, geographic distributions, and associated environmental parameters for comparative biological questions. Contributing to the acquisition and synthesis of genomic, phenotypic, and ecological data across populations and lineages within a species, and across species through time and space, is an important endeavor for biodiversity informatics and is critical for advancing our knowledge of life on an ever-changing planet (Losos et al., 2013).

WHAT EXCITING QUESTIONS STILL REMAIN?

Following are a few fundamental questions that span the biogeography–phylogeography–landscape genetics spectrum (Fig. 16.2); they are not new but are examples of important areas of inquiry bridging ecology and evolutionary biology that are ripe for investigation using spatially explicit analyses of genomic-scale data: (*i*) How much reduction of gene flow, if any, is required to generate new species, and is that gene flow most often reduced by geographic distance, natural selection, or other factors? (*ii*) Why do species have range limits in the absence of

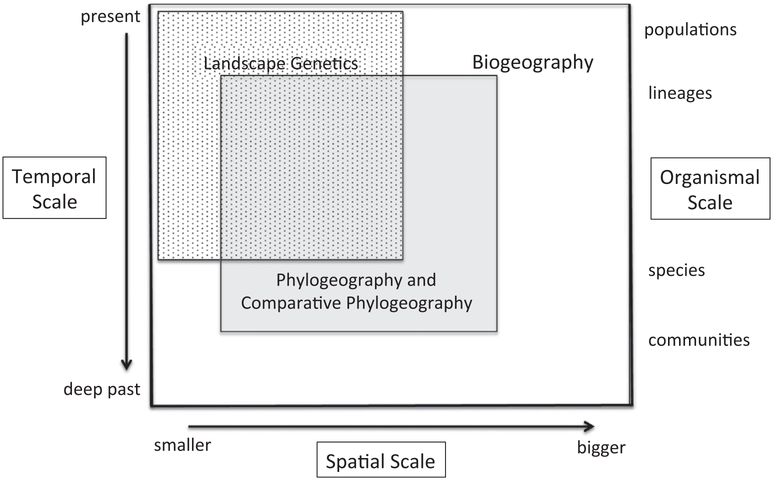


FIGURE 16.2 Landscape genetics (stippled) and phylogeography and comparative phylogeography (gray) differ somewhat in temporal, spatial, and organismal extent. These fields fall within the broader discipline of biogeography (white), which encompasses all scales.

geographic barriers? (*iii*) What is the relative role of history vs. contemporary processes on genetic diversity patterns of species and thus communities? (*iv*) Can adaptive evolution and genetic diversity keep pace with the scale of current environmental variability and shifts? (*v*) Can we predict species' responses to climate change (including degree of adaptation or phenotypic plasticity), and how might species in a community differ and thus influence coevolutionary or eco-evolutionary dynamics? (*vi*) How constrained by phylogenetic history and ecology are the evolution of species' traits, and what explains convergent evolution among suites of traits (functional or otherwise) along environmental gradients across codistributed species?

SUMMARY

Biogeography is the union of ecology and evolutionary biology. It can provide insight into speciation, species delimitations, adaptation, and the future of species in a rapidly changing world. For studies linking geography, ecology, and history in their understanding of genetic variation in nature, we should be using all methods that allow us to gain access to, and understanding of, the processes influencing biodiversity. Why continue to separate fields by the scale of study (temporal or spatial) or the method of analysis? This false separation leads to a real separation in thinking, in training, in practice, and in effort. We owe it to the next generation of scientists to show them the links between the fields and how an integrative approach can yield stronger inferences about some of the most fundamental questions in biology.

ACKNOWLEDGMENTS

I thank J. Avise and F. Ayala for the invitation to contribute to this colloquium; S. Scheiner and two anonymous reviewers for careful reviews of the manuscript; and J. Coyne for discussions on speciation and comments on an earlier draft of this manuscript. The views expressed in this paper do not necessarily reflect those of the U.S. National Science Foundation or U.S. government.

References

Abbott IA (1999) *Marine Red Algae of the Hawaiian Islands* (Bishop Museum, Honolulu).

Abellán P, Millán A, Ribera I (2009) Parallel habitat-driven differences in the phylogeographical structure of two independent lineages of Mediterranean saline water beetles. *Mol Ecol* 18(18):3885–3902.

Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, Gragert L, Babrzadeh F, Gharizadeh B, Luo M, Plummer FA, Kimani J, Carrington M, Middleton D, Rajalingam R, Beksac M, Marsh SG, Maiers M, Guethlein LA, Tavoularis S, Little AM, Green RE, Norman PJ, Parham P (2011) The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* 334(6052):89–94.

Achilli A, Perego UA, Lancioni H, Olivieri A, Gandini F, Kashani BH, Battaglia V, Grugni V, Angerhofer N, Rogers MP, Herrera RJ, Woodward SR, Labuda D, Smith DG, Cybulski JS, Semino O, Malhi RS, Torroni A (2013) Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proc Natl Acad Sci USA* 110(35):14308–14313.

Adachi N, Shinoda K, Umetsu K, Kitano T, Matsumura H, Fujiyama R, Sawada J, Tanaka M (2011) Mitochondrial DNA analysis of Hokkaido Jomon skeletons: Remnants of archaic maternal lineages at the southwestern edge of former Beringia. *Am J Phys Anthropol* 146(3):346–360.

Adams CC (1905) The postglacial dispersal of the North American biota. *Biol Bull* 9(1):53–71.

Addison JA, Hart MW (2005) Colonization, dispersal, and hybridization influence phylogeography of North Atlantic sea urchins (*Strongylocentrotus droebachiensis*). *Evolution* 59(3):532–543.

Adkins RM, Honeycutt RL (1994) Evolution of the primate cytochrome *c* oxidase subunit II gene. *J Mol Evol* 38(3):215–231.

Ahti PA, Coleman RR, DiBattista JD, Berumen ML, Rocha LA, Bowen BW (2016) Phylogeography of Indo-Pacific reef fishes: Sister wrasses *Coris gaimard* and *C. cuvieri* in the Red Sea, Indian Ocean and Pacific Ocean. *J Biogeogr* 43(6):1103–1115.

Aikhenvald AY (2013) Amazonia: Linguistic history. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 384–391.

Akaike H (1973) Information theory as an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*, eds Petrov BN, Csaki F (Akademiai Kiado, Budapest), pp 267–281.

Akashi H (1999) Inferring the fitness effects of DNA mutations from polymorphism and divergence data: Statistical power to detect directional selection under stationarity and free recombination. *Genetics* 151(1):221–238.

Alison Kay E, Palumbi SR (1987) Endemism and evolution in Hawaiian marine invertebrates. *Trends Ecol Evol* 2(7):183–186.

Allen B (2013) Papua New Guinea: Indigenous migrations in the recent past. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 302–307.

Allen GR (2008) Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. *Aquatic Conserv: Mar Freshw Ecosyst* 18(5):541–556.

Allison SD, Martiny JBH (2008) Resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci USA* 105(Suppl 1):11512–11519.

Alström P, Rasmussen PC, Zhao C, Xu J, Dalvi S, Cai T, Guan Y, Zhang R, Kalyakin MV, Lei F, Olsson U (2016) Integrative taxonomy of the plain-backed thrush (*Zoothera mollissima*) complex (Aves, Turdidae) reveals cryptic species, including a new species. *Avian Res* 7(1):1–39.

Alvain S, Moulin C, Dandonneau Y, Bréon F-M (2005) Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. *Deep-Sea Res* 52(11):1989–2004.

Alvarado-Serrano DF, Knowles LL (2014) Ecological niche models in phylogeographic studies: Applications, advances and precautions. *Mol Ecol Resour* 14(2):233–248.

Alvarez G, Ceballos FC, Quinteiro C (2009) The role of inbreeding in the extinction of a European royal dynasty. *PLoS One* 4(4):e5174.

Alvarez N, Thiel-Egenter C, Tribsch A, Holderegger R, Manel S, Schönswetter P, Taberlet P, Brodbeck S, Gaudeul M, Gielly L, Küpfer P, Mansion G, Negrini R, Paun O, Pellecchia M, Rioux D, Schüpfer F, Van Loo M, Winkler M, Gugerli F, IntraBioDiv Consortium (2009) History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecol Lett* 12(7):632–640.

Amick DS (2016) Evolving views on the Pleistocene colonization of North America *Quat Int*, doi:10.1016/j.quaint.2015.12.030.

Amorim IR, Emerson BC, Borges PAV, Wayne RK (2012) Phylogeography and molecular phylogeny of Macaronesian island *Tarphius* (Coleoptera: Zopheridae): Why are there so few species in the Azores? *J Biogeogr* 39(9):1583–1595.

Anderson A (2013) Trans-Indian Ocean migration. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 254–258.

Anderson DR (2008) *Model Based Inference in the Life Sciences* (Springer, New York).

Anderson E (1948) Hybridization of the habitat. *Evolution* 2(1):1–9.

Anderson E (1953) Introgressive hybridization. *Biol Rev Camb Philos Soc* 28(3):280–307.

Anderson MJ, Gorley RN, Clarke KR (2008) *PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods* (PRIMER-E, Plymouth, UK).

Andrew RL, Ostevik KL, Ebert DP, Rieseberg LH (2012) Adaptation with gene flow across the landscape in a dune sunflower. *Mol Ecol* 21(9):2078–2091.

Andrews KR (2014) Population genetics in the conservation of cetaceans and primates. *Primates and Cetaceans*, eds Yamagiwa J, Karczmarski L (Springer, New York), pp 289–308.

Andrews KR, Karczmarski L, Au WW, Rickards SH, Vanderlip CA, Bowen BW, Gordon Grau E, Toonen RJ (2010) Rolling stones and stable homes: Social structure, habitat diversity and population genetics of the Hawaiian spinner dolphin (*Stenella longirostris*). *Mol Ecol* 19(4):732–748.

Andrews KR, Perrin WF, Oremus M, Karczmarski L, Bowen BW, Puritz JB, Toonen RJ (2013) The evolving male: Spinner dolphin (*Stenella longirostris*) ecotypes are divergent at Y chromosome but not mtDNA or autosomal markers. *Mol Ecol* 22(9):2408–2423.

Andrews KR, Hohenlohe PA, Miller MR, Hand BK, Seeb JE, Luikart G (2014a) Trade-offs and utility of alternative RADseq methods: Reply to Puritz et al. *Mol Ecol* 23(24):5943–5946.

Andrews KR, Norton EL, Fernandez-Silva I, Portner E, Goetze E (2014b) Multilocus evidence for globally distributed cryptic species and distinct populations across ocean gyres in a mesopelagic copepod. *Mol Ecol* 23(22):5462–5479.

Angelis K, dos Reis M (2015) The impact of ancestral population size and incomplete lineage sorting on Bayesian estimation of species divergence times. *Curr Zool* 61(5):874–885.

Antonelli A, Zizka A, Silvestro D, Scharn R, Cascales-Miñana B, Bacon CD (2015) An engine for global plant diversity: Highest evolutionary turnover and emigration in the American tropics. *Front Genet* 6:130.

Antonovics J (1971) The effects of a heterogeneous environment on the genetics of natural populations. *Am Sci* 59(5):593–599.

Aplin KP, Rhind SG, Have JT, Chesser RT (2015) Taxonomic revision of *Phascogale tapoatafa* (Meyer, 1793) (Dasyuridae; Marsupialia), including descriptions of two new subspecies and confirmation of *P. pirata* Thomas, 1904 as a "Top End" endemic. *Zootaxa* 4055(1):1–73.

Arbogast BS, Kenagy G (2001) Comparative phylogeography as an integrative approach to historical biogeography. *J Biogeogr* 28(7):819–825.

Arbogast BS, Drovetski SV, Curry RL, Boag PT, Seutin G, Grant PR, Grant BR, Anderson DJ (2006) The origin and diversification of Galapagos mockingbirds. *Evolution* 60(2): 370–382.

Armitage SJ, Jasim SA, Marks AE, Parker AG, Usik VI, Uerpmann HP (2011) The southern route "out of Africa": Evidence for an early expansion of modern humans into Arabia. *Science* 331(6016):453–456.

Árnason E (2004) Mitochondrial cytochrome *B* DNA variation in the high-fecundity Atlantic cod: Trans-Atlantic clines and shallow gene genealogy. *Genetics* 166(4):1871–1885.

Arnedo MA, Gillespie RG (2006) Species diversification patterns in the Polynesian jumping spider genus *Havaika* Prószyński, 2001 (Araneae, Salticidae). *Mol Phylogenet Evol* 41(2):472–495.

Arnedo MA, Oromí P, Ribera C (2001) Radiation of the spider genus *Dysdera* (Araneae, Dysderidae) in the Canary Islands: Cladistic assessment based on multiple data sets. *Cladistics* 17(4):313–353.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Mol Ecol* 22(11):3179–3190.

Ashley MV, Abraham ST, Backs JR, Koenig WD (2015) Landscape genetics and population structure in valley oak (*Quercus lobata* Née). *Am J Bot* 102(12):2124–2131.

Auler AS, Wang X, Edwards RL, Cheng H, Cristalli PS, Smart PL, Richards DA (2004) Quaternary ecological and geomorphic changes associated with rainfall events in presently semi-arid northeastern Brazil. *J Quaternary Sci* 19(7):693–701.

Austin CC (1999) Lizards took express train to Polynesia. *Nature* 397(6715):113–114.

Avila-Pires TCS, Mulcahy DG, Werneck FP, Sites JW (2012) Phylogeography of the teiid lizard *Kentropyx calcarata* and the sphaerodactylid *Gonatodes humeralis* (Reptilia: Squamata): Testing a geological scenario for the lower Amazon-Tocantins basins, Amazonia, Brazil. *Herpetologica* 68(2):272–287.

Avise JC (1992) Molecular population structure and the biogeographic history of a regional fauna: A case history with lessons for conservation biology. *Oikos* 63(1):62–76.

Avise JC (1996) Toward a regional conservation genetics perspective: Phylogeography of faunas in the southeastern United States. *Conservation Genetics: Case Histories from Nature*, eds Avise JC, Hamrick J (Chapman and Hall, New York), pp 431–470.

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

Avise JC (2008) Phylogeography: Retrospect and prospect. *J Biogeogr* 36(1):3–15.

Avise JC, Ayala FJ (2007) In the light of evolution I: Adaptation and complex design. *Proc Natl Acad Sci USA* 104(Suppl 1):8563–8676.

Avise JC, Ayala FJ (2009) In the light of evolution III: Two centuries of Darwin. *Proc Natl Acad Sci USA* 106(Suppl 1):9933–10066.

Avise JC, Ayala FJ (2010) In the light of evolution IV: The human condition. *Proc Natl Acad Sci USA* 107(Suppl 2):8897–9022.

Avise JC, Giblin–Davidson C, Laerm J, Patton JC, Lansman RA (1979a) Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. *Proc Natl Acad Sci USA* 76:6694–6698.

Avise JC, Lansman RA, Shade RO (1979b) The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus *Peromyscus*. *Genetics* 92:279–295.

Avise JC, Helfman GS, Saunders NC, Hales LS (1986) Mitochondrial DNA differentiation in North Atlantic eels: Population genetic consequences of an unusual life history pattern. *Proc Natl Acad Sci USA* 83(12):4350–4354.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst* 18:489–522.

Avise JC, Hubbell SP, Ayala FJ (2008) In the light of evolution II: Biodiversity and extinction. *Proc Natl Acad Sci USA* 105(Suppl 1):11453–11586.

Baas Becking LGM (1934) *Geobiologie: Of Inleiding tot de Milieukunde* (Van Stockum and Zoon, The Hague, The Netherlands).

Badgley C, Smiley TM, Finarelli JA (2014) Great Basin mammal diversity in relation to landscape history. *J Mammal* 95(6):1090–1106.

Bagley JC, Johnson JB (2014) Testing for shared biogeographic history in the lower Central American freshwater fish assemblage using comparative phylogeography: Concerted, independent, or multiple evolutionary responses? *Ecol Evol* 4(9):1686–1705.

Baker AJ, Tavares ES, Elbourne RF (2009) Countering criticisms of single mitochondrial DNA gene barcoding in birds. *Mol Ecol Resour* 9(Suppl s1):257–268.

Baker PA, Fritz SC, Dick CW, Eckert AJ, Horton BK, Manzoni S, Ribas C, Garzione C, Battisi D (2014) The emerging field of Geogenomics: Constraining geological problems with genetic data. *Earth Sci Rev* 135:38–47.

Balakrishnan CN, Lee JY, Edwards SV (2010) Phylogeography and phylogenetics in the nuclear age. *Searching for the Causes of Evolution: From Field Observations to Mechanisms*, eds Grant P, Grant R (Princeton Univ Press, Princeton, NJ), pp 65–88.

Balaresque P, Poulet N, Cussat-Blanc S, Gerard P, Quintana-Murci L, Heyer E, Jobling MA (2015) Y-chromosome descent clusters and male differential reproductive success: Young lineage expansions dominate Asian pastoral nomadic populations. *Eur J Hum Genet* 23(10):1413–1422.

Baldwin BG (2014) Origins of plant diversity in the California floristic province. *Annu Rev Ecol Evol Syst* 45:347–369.

Baldwin BG, Robichaux RH (1995) Historical biogeography and ecology of the Hawaiian silversword alliance (Asteraceae): New molecular phylogenetic perspectives. *Hawaiian Biogeography: Evolution on a Hot Spot Archipelago*, eds Wagner WL, Funk VA (Smithsonian Institution Press, Washington, DC), pp 259–287.

Ball M, Neigel JE, Avise JC (1990) Gene genealogies within organismal pedigrees of random-mating populations. *Evolution* 44:360–370.

Barber PH, Cheng SH, Erdmann MV, Tengardjaja K (2011) Evolution and conservation of marine biodiversity in the Coral Triangle: Insights from stomatopod Crustacea. *Crustac Issues* 19:129–156.

Barberán A, Fernández-Guerra A, Bohannan BJM, Casamayor EO (2012) Exploration of community traits as ecological markers in microbial metagenomes. *Mol Ecol* 21(8):1909–1917.

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: How, where and how many? *Methods Ecol Evol* 3(2):327–338.

Barbujani G, Sokal RR (1990) Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc Natl Acad Sci USA* 87(5):1816–1819.

Barker G, Barton H, Bird M, Daly P, Datan I, Dykes A, Farr L, Gilbertson D, Harrisson B, Hunt C, Higham T, Kealhofer L, Krigbaum J, Lewis H, McLaren S, Paz V, Pike A, Piper P, Pyatt B, Rabett R, Reynolds T, Rose J, Rushworth G, Stephens M, Stringer C, Thompson J, Turney C (2007) The "human revolution" in lowland tropical Southeast Asia: The antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo). *J Hum Evol* 52(3):243–261.

Barley AJ, Monnahan PJ, Thomson RC, Grismer LL, Brown RM (2015) Sun skink landscape genomics: Assessing the roles of micro-evolutionary processes in shaping genetic and phenotypic diversity across a heterogeneous and fragmented landscape. *Mol Ecol* 24(8):1696–1712.

Barrett RD, Schluter D (2008) Adaptation from standing genetic variation. *Trends Ecol Evol* 23(1):38–44.

Barrow LN, Bigelow AT, Phillips CA, Lemmon EM (2015) Phylogeographic inference using Bayesian model comparison across a fragmented chorus frog species complex. *Mol Ecol* 24(18):4739–4758.

Barrowclough GF, Zink RM (2009) Funds enough, and time: mtDNA, nuDNA and the discovery of divergence. *Mol Ecol* 18(14):2934–2936.

Barton NH (1995) Linkage and the limits to natural selection. *Genetics* 140(2):821–841.

Barton NH, Etheridge AM (2011) The relation between reproductive value and genetic contribution. *Genetics* 188(4):953–973.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annu Rev Ecol Syst* 16:113–148.

Batalha-Filho H, Fjeldsa J, Fabre PH, Miyaki CY (2013) Connections between the Atlantic and the Amazonian Forest avifaunas represent distinct historical events. *J Ornithol* 154(1):41–50.

Batini C, Ferri G, Destro-Bisol G, Brisighelli F, Luiselli D, Sánchez-Diz P, Rocha J, Simonson T, Brehm A, Montano V, Elwali NE, Spedini G, D'Amato ME, Myres N, Ebbesen P, Comas D, Capelli C (2011) Signatures of the preagricultural peopling processes in sub-Saharan Africa as revealed by the phylogeography of early Y chromosome lineages. *Mol Biol Evol* 28(9):2603–2613.

Batjes NH (2000) Global data set of derived soil properties, 0.5-degree grid (ISRIC-WISE). Available at daac.ornl.gov.

Baum DA, Shaw KL (1995) Genealogical perspectives on the species problem. *Experimental and Molecular Approaches to Plant Biosystematics*, eds Hoch PC, Stephenson AG (Missouri Botanical Garden Press, St. Louis), pp 289–303.

Baums IB, Boulay JN, Polato NR, Hellberg ME (2012) No gene flow across the Eastern Pacific Barrier in the reef-building coral *Porites lobata*. *Mol Ecol* 21(22):5418–5433.

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* 162(4):2025–2035.

Beaumont MA, Nielsen R, Robert C, Hey J, Gaggiotti O, Knowles L, Estoup A, Panchal M, Corander J, Hickerson M, Sisson SA, Fagundes N, Chikhi L, Beerli P, Vitalis R, Cornuet J-M, Huelsenbeck J, Foll M, Yang Z, Rousset F, Balding D, Excoffier L (2010) In defense of model-based inference in phylogeography. *Mol Ecol* 19(3):436–446.

Beaumont W, Beverly R, Southon J, Taylor RE (2010) Bone preparation at the KCCAMS laboratory. *Nucl Instrum Methods Phys Res B* 268(7–8):906–909.

Beavis AS, Sunnucks P, Rowell DM (2011) Microhabitat preferences drive phylogeographic disparities in two Australian funnel web spiders. *Biol J Linn Soc Lond* 104(4):805–819.

Becquet C, Przeworski M (2007) A new approach to estimate parameters of speciation models with application to apes. *Genome Res* 17(10):1505–1519.

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152(2):763–773.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proc Natl Acad Sci USA* 98(8):4563–4568.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* 185(1):313–326.

Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith C, Wells RS, Rosset S; Genographic Consortium (2008) The dawn of human matrilineal diversity. *Am J Hum Genet* 82(5):1130–1140.

Beheregaray LB (2008) Twenty years of phylogeography: The state of the field and the challenges for the Southern Hemisphere. *Mol Ecol* 17(17):3754–3774.

Beheregaray LB, Gibbs JP, Havill N, Fritts TH, Powell JR, Caccone A (2004) Giant tortoises are not so slow: Rapid diversification and biogeographic consensus in the Galápagos. *Proc Natl Acad Sci USA* 101(17):6514–6519.

Belkenhol N, Gugerli F, Cushman SA, Waits LP, Coulon A, Arntzen JW, Holderegger R, Wagner HH, Arens PFP, Campagne P, Dale VH, Nicieza AG, Smulders MJM, Tedesco E, Wang H, Wasserman T (2009) Identifying future research needs in landscape genetics: Where to from here? *Landsc Ecol* 24(4):455–463.

Bell RC, MacKenzie JB, Hickerson MJ, Chavarría KL, Cunningham M, Williams S, Moritz C (2012) Comparative multi-locus phylogeography confirms multiple vicariance events in co-distributed rainforest frogs. *Proc Biol Sci* 279(1730):991–999.

Bellemain E, Ricklefs RE (2008) Are islands the end of the colonization road? *Trends Ecol Evol* 23(8):461–468.

Bellwood P (2013) Southeast Asian islands: Archaeology. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 284–292.

Benazzi S, Douka K, Fornai C, Bauer CC, Kullmer O, Svoboda J, Pap I, Mallegni F, Bayle P, Coquerelle M, Condemi S, Ronchitelli A, Harvati K, Weber GW (2011) Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* 479(7374):525–528.

Benkman CW (1996) Are the ratios of bill crossing morphs in crossbills the result of frequency-dependent selection? *Evol Ecol* 10(1):119–126.

Beresford-Jones D, Heggart P (2013) Andes: Archaeology. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 410–416.

Bergmann GT, Bates ST, Eilers KG, Lauber CL, Caporaso JG, Walters WA, Knight R, Fierer N (2011) The under-recognized dominance of *Verrucomicrobia* in soil bacterial communities. *Soil Biol Biochem* 43(7):1450–1455.

Bergsten J, Bilton DT, Fujisawa T, Elliott M, Monaghan MT, Balke M, Hendrich L, Geijer J, Herrmann J, Foster GN, Ribera I, Nilsson AN, Barraclough TG, Vogler AP (2012) The effect of geographical scale of sampling on DNA barcoding. *Syst Biol* 61(5):851–869.

Bermingham E, Avise JC (1986) Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics* 113(4):939–965.

Bermingham E, Moritz C (1998) Comparative phylogeography: Concepts and applications. *Mol Ecol* 7:367–369.

Bernardi G, Bucciarelli G, Costagliola D, Robertson DR, Heiser J (2004) Evolution of coral reef fish *Thalassoma* spp. (Labridae). 1. Molecular phylogeny and biogeography. *Mar Biol* 144(2):369–375.

Bernardo J, Spotila JR (2006) Physiological constraints on organismal response to global warming: Mechanistic insights from clinally varying populations and implications for assessing endangerment. *Biol Lett* 2(1):135–139.

Bernatchez L, Wilson CC (1998) Comparative phylogeography of Nearctic and Palearctic fishes. *Mol Ecol* 7(4):431–452.

Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, Moritz C (2013) Unlocking the vault: Next-generation museum population genomics. *Mol Ecol* 22(24):6018–6032.

Bickford D, Lohman DJ, Sodhi NS, Ng PK, Meier R, Winker K, Ingram KK, Das I (2007) Cryptic species as a window on diversity and conservation. *Trends Ecol Evol* 22(3):148–155.

Binford LR (2001) *Constructing Frames of Reference: An Analytical Method for Archaeological Theory Building Using Hunter-Gatherer and Environmental Data Sets* (Univ of California Press, Berkeley).

Bird CE, Holland BS, Bowen BW, Toonen RJ (2011) Diversification of sympatric broadcast-spawning limpets (*Cellana* spp.) within the Hawaiian archipelago. *Mol Ecol* 20(10): 2128–2141.

Birky CW Jr, Wolf C, Maughan H, Herbertson L, Henry E (2005) Speciation and selection without sex. *Hydrobiologia* 546(1):29–45.

Bivand RS, Pebesma E, Gomez-Rubio V (2013) *Applied Spatial Data Analysis with R* (Springer, New York), 2nd Ed.

Blair C, Campbell CR, Yoder AD (2015) Assessing the utility of whole genome amplified DNA for next-generation molecular ecology. *Mol Ecol Resour* 15(5):1079–1090.

Blair D, McMahon A, McDonald B, Tikel D, Waycott M, Marsh H (2014) Pleistocene sea level fluctuations and the phylogeography of the dugong in Australian waters. *Mar Mamm Sci* 30(1):104–121.

Blanco MB (2011) Timely estrus in wild brown mouse lemur females at Ranomafana National Park, southeastern Madagascar. *Am J Phys Anthropol* 145(2):311–317.

Blanco MB, Rasoazanabary E, Godfrey LR (2015) Unpredictable environments, opportunistic responses: Reproduction and population turnover in two wild mouse lemur species (*Microcebus rufus* and *M. griseorufus*) from eastern and western Madagascar. *Am J Primatol* 77(9):936–947.

Blome MW, Cohen AS, Tryon CA, Brooks AS, Russell J (2012) The environmental context for the origins of modern human diversity: A synthesis of regional variability in African climate 150,000–30,000 years ago. *J Hum Evol* 62(5):563–592.

Bloomquist EW, Lemey P, Suchard MA (2010) Three roads diverged? Routes to phylogeographic inference. *Trends Ecol Evol* 25(11):626–632.

Blum MG (2012) Approximate Bayesian computation: A nonparametric perspective. *J Am Stat Assoc* 105(491):1178–1187.

Blum MGB, Heyer E, François O, Austerlitz F (2006) Matrilineal fertility inheritance detected in hunter-gatherer populations using the imbalance of gene genealogies. *PLoS Genet* 2(8):e122.

Blust R (2013) Southeast Asian islands and Oceania: Austronesiam linguistic history. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 276–283.

Bohonak AJ, Vandergast AG (2011) The value of DNA sequence data for studying landscape genetics. *Mol Ecol* 20(12):2477–2479, reply 2480–2482.

Bolnick DI, Near TJ (2005) Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). *Evolution* 59(8):1754–1767.

Bonacum J, O'Grady PM, Kambysellis M, Desalle R (2005) Phylogeny and age of diversification of the *planitibia* species group of the Hawaiian *Drosophila*. *Mol Phylogenet Evol* 37(1):73–82.

Bond WJ, Silander JA, Ranaivonasy J, Ratsirarson J (2008) The antiquity of Madagascar's grasslands and the rise of C(4) grassy biomes. *J Biogeogr* 35(10):1743–1758.

Bonneville A, Le Suavé R, Audin L, Clouard V, Dosso L, Yves GP, Janney P, Jordahl K, Maamaatuaiahutapu K (2002) Arago Seamount: The missing hotspot found in the Austral Islands. *Geology* 30(11):1023–1026.

Bosch MD, Mannino MA, Prendergast AL, O'Connell TC, Demarchi B, Taylor SM, Niven L, van der Plicht J, Hublin JJ (2015) New chronology for Ksâr 'Akil (Lebanon) supports Levantine route of modern human dispersal into Europe. *Proc Natl Acad Sci USA* 112(25):7683–7688.

Boschi EE (2000) Species of decapod crustaceans and their distribution in the American marine zoogeographic provinces. *Revista de Investigación y Desarrollo Pesquero* 13:1–136.

Boulesteix AL, Strimmer K (2007) Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Brief Bioinform* 8(1):32–44.

Bourjea J, Lapègue S, Gagnevin L, Broderick D, Mortimer JA, Ciccione S, Roos D, Taquet C, Grizel H (2007) Phylogeography of the green turtle, *Chelonia mydas*, in the Southwest Indian Ocean. *Mol Ecol* 16(1):175–186.

Bowen BW (2016) The three domains of conservation genetics: Case histories from Hawaiian waters. *J Hered* 107(4):309–317.

Bowen BW, Grant WS (1997) Phylogeography of the sardines (*Sardinops* spp.): Assessing biogeographic models and population histories in temperate upwelling zones. *Evolution* 51(5):1601–1610.

Bowen BW, Karl SA (2007) Population genetics and phylogeography of sea turtles. *Mol Ecol* 16(23):4886–4907.

Bowen BW, Rocha LA, Toonen RJ, Karl SA; ToBo Laboratory (2013) The origins of tropical marine biodiversity. *Trends Ecol Evol* 28(6):359–366.

Bowen BW, Shanker K, Yasuda N, Celia M, Malay MCD, von der Heyden S, Paulay G, Rocha LA, Selkoe KA, Barber PH, Williams ST, Lessios HA, Crandall ED, Bernardi G, Meyer CP, Carpenter KE, Toonen RJ (2014) Phylogeography unplugged: Comparative surveys in the genomic era. *Bull Mar Sci* 90(1):13–46.

Bowern C (2012) The riddle of Tasmanian languages. *Proc Biol Sci* 279(1747):4590–4595.

Bowman DMJS, Brown GK, Braby MF, Brown JR, Cook LG, Crisp MD, Ford F, Haberle S, Hughes J, Isagi Y, Joseph L, McBride J, Nelson G, Ladiges PY (2010) Biogeography of the Australian monsoon tropics. *J Biogeogr* 37(2):201–216.

Bradburd GS, Ralph PL, Coop GM (2016) A spatial framework for understanding population structure and admixture. *PLoS Genet* 12(1):e1005703.

Bradshaw AD (1972) Some of the evolutionary consequences of being a plant. *Evol Biol* 5:25–47.

Bradshaw HD, Schemske DW (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* 426(6963):176–178.

Bragg JG, Potter S, Bi K, Moritz C (2016) Exon capture phylogenomics: Efficacy across scales of divergence. *Mol Ecol Resour* 16:1059–1068.

Braithwaite DT, Keegan KP (2013) matR: Metagenomics analysis tools for R. R package version 0.9.9.

Brandley MC, Bragg JG, Singhal S, Chapple DG, Jennings CK, Lemmon AR, Lemmon EM, Thompson MB, Moritz C (2015) Evaluating the performance of anchored hybrid enrichment at the tips of the tree of life: A phylogenetic analysis of Australian *Eugongylus* group scincid lizards. *BMC Evol Biol* 15(1):62.

Briggs JC (2003) Marine centres of origin as evolutionary engines. *J Biogeogr* 30(1):1–18.

Briggs JC, Bowen BW (2012) A realignment of marine biogeographic provinces with particular reference to fish distributions. *J Biogeogr* 39(1):12–30.

Brito PH, Edwards SV (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* 135(3):439–455.

Broennimann O, Fitzpatrick MC, Pearman PB, Petitpierre B, Pellissier L, Yoccoz NG, Thuiller W, Fortin M-J, Randin C, Zimmermann NE, Graham CH, Guisan A (2012) Measuring ecological niche overlap from occurrence and spatial environmental data. *Glob Ecol Biogeogr* 21(4):481–497.

Brown AM, Kopps AM, Allen SJ, Bejder L, Littleford-Colquhoun B, Parra GJ, Cagnazzi D, Thiele D, Palmer C, Frère CH (2014) Population differentiation and hybridisation of Australian snubfin (*Orcaella heinsohni*) and Indo-Pacific humpback (*Sousa chinensis*) dolphins in north-western Australia. *PLoS One* 9(7):e101427.

Brown JL (2014) SDMtoolbox: A python-based GIS toolkit for landscape genetic, biogeography, and species distribution model analyses. *Methods Ecol Evol* 5(7):694–700.

Brown JL, Knowles LL (2012) Spatially explicit models of dynamic histories: Examination of the genetic consequences of Pleistocene glaciation and recent climate change on the American pika. *Mol Ecol* 21(15):3757–3775.

Brown JL, Weber JJ, Alvarado-Serrano DF, Hickerson MJ, Franks SJ, Carnaval AC (2016) Predicting the genetic consequences of future climate change: The power of coupling spatial demography, the coalescent, and historical landscape changes. *Am J Bot* 103(1):153–163.

Brown R, Diesmos A (2009) Philippines, biology. *Encyclopedia of Islands,* eds Gillespie R, Clague D (Univ of California Press, Berkeley), pp 723–732.

Bru D, Ramette A, Saby NP, Dequiedt S, Ranjard L, Jolivet C, Arrouays D, Philippot L (2011) Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. *ISME J* 5(3):532–542.

Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, de Vargas C, Gasol JM, Gorsky G, Gregory AC, Guidi L, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Poulos BT, Schwenck SM, Speich S, Dimier C, Kandels-Lewis S, Picheral M, Searson S; Tara Oceans Coordinators, Bork P, Bowler C, Sunagawa S, Wincker P, Karsenti E, Sullivan MB (2015) Ocean plankton: Patterns and ecological drivers of ocean viral communities. *Science* 348(6237):1261498.

Brumfield RT, Capparella AP (1996) Historical diversification of birds in northwestern South America: A molecular perspective on the role of vicariant events. *Evolution* 50(4):1607–1624.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol Evol* 18(5):249–256.

Brunsfeld S, Sullivan J, Soltis D, Soltis P (2001) Comparative phylogeography of northwestern North America: A synthesis. *Integrating Ecological and Evolutionary Processes in a Spatial Context,* eds Silvertown J, Antonovics J (Blackwell Science, Oxford, UK), pp 319–340.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Mol Biol Evol* 29(8):1917–1932.

Buckley D (2009) Toward an organismal, integrative, and iterative phylogeography. *BioEssays* 31(7):784–793.

Buckley HL, Paterson AM, Cruickshank RH (2013) The founder space race: A response to Waters et al. *Trends Ecol Evol* 28(4):189–190.

Buckley TR, Attanayake D, Bradler S (2009) Extreme convergence in stick insect evolution: Phylogenetic placement of the Lord Howe Island tree lobster. *Proc Biol Sci* 276(1659):1055–1062.

Bucklin A, Wiebe PH, Smolenack SB, Copley NJ, Beaudet JG, Bonner KG, Färber-Lorda J, Pierson JJ (2007) DNA barcodes for species identification of euphausiids (Euphausiacea, Crustacea). *J Plankton Res* 29(6):483–493.

Budd AF, Pandolfi JM (2010) Evolutionary novelty is concentrated at the edge of coral species distributions. *Science* 328(5985):1558–1561.

Bulgarella M, Peters JL, Kopuchian C, Valqui T, Wilson RE, McCracken KG (2012) Multilocus coalescent analysis of haemoglobin differentiation between low- and high-altitude populations of crested ducks (*Lophonetta specularioides*). *Mol Ecol* 21(2):350–368.

Buonaccorsi VP, McDowell JR, Graves JE (2001) Reconciling patterns of inter-ocean molecular variance from four classes of molecular markers in blue marlin (*Makaira nigricans*). *Mol Ecol* 10(5):1179–1196.

Burban C, Petit RJ (2003) Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Mol Ecol* 12(6):1487–1495.

Buri P (1956) Gene frequency in small populations of mutant *Drosophila*. *Evolution* 10:367–402.

Burney DA (1986) *Late Quaternary Environmental Dynamics of Madagascar.* PhD Dissertation (Duke University, Durham, NC).

Burney DA (1987) Late Quaternary stratigraphic charcoal records from Madagascar. *Quat Res* 28:274–280.

Burney DA (1988) Modern pollen spectra from Madagascar. *Palaeogeogr Palaeoclimatol Palaeoecol* 66:63–75.

Burney DA (1997) Theories and facts regarding Holocene environmental change before and after human colonization. *Natural Change and Human Impact in Madagascar,* eds Goodman SM, Patterson BD (Smithsonian Institution, Washington, DC), pp 75–89.

Burney DA, James H, Grady F, Rafamantanantsoa J-G, Ramilisonina, Wright H, Cowart J (1997) Environmental change, extinction and human activity: Evidence from caves in NW Madagascar. *J Biogeogr* 24:755–767.

Burnham KP, Anderson DR (1998) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer, New York).

Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer, New York), 2nd Ed.

Burns JA (2010) Mammalian faunal dynamics in Late Pleistocene Alberta, Canada. *Quat Int* 217(1–2):37–42.

Burns SJ, Godfrey LR, Faina P, McGee D, Hardt B, Ranivoharimanana L, Randrianasy J (2016) Rapid human-induced landscape transformation in Madagascar at the end of the first millennium of the Common Era. *Quat Sci Rev* 134:92–99.

Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, Suh A, Dutoit L, Bureš S, Garamszegi LZ, Hogner S, Moreno J, Qvarnström A, RužićM, Sæther SA, Sætre GP, Török J, Ellegren H (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res* 25(11):1656–1665.

Burridge AK, Goetze E, Raes N, Huisman J, Peijnenburg KT (2015) Global biogeography and evolution of *Cuvierina* pteropods. *BMC Evol Biol* 15(1):39.

Butlin RK (2005) Recombination and speciation. *Mol Ecol* 14(9):2621–2635.

Caccone A, Gentile G, Gibbs JP, Frirts TH, Snell HL, Betts J, Powell JR (2002) Phylogeography and history of giant Galápagos tortoises. *Evolution* 56(10):2052–2066.

Cadena CD, Kozak KH, Gómez JP, Parra JL, McCain CM, Bowie RC, Carnaval AC, Moritz C, Rahbek C, Roberts TE, Sanders NJ, Schneider CJ, VanDerWal J, Zamudio KR, Graham CH (2012) Latitude, elevational climatic zonation and speciation in New World vertebrates. *Proc Roy Soc B* 279(1726):194–201.

Callmander MW, Phillipson PB, Schatz GE, Andriambololonera S, Rabarimanarivo M, Rakotonirina N, Raharimampionona J, Chatelain C, Gautier L, Lowry PP (2011) The endemic and non-endemic vascular flora of Madagascar updated. *Plant Ecol Evol* 144(2):121–125.

Calsbeek R, Thompson JN, Richardson JE (2003) Patterns of molecular evolution and diversification in a biodiversity hotspot: The California Floristic Province. *Mol Ecol* 12(4): 1021–1029.

Camargo A, Sinervo B, Sites JW Jr (2010) Lizards as model organisms for linking phylogeographic and speciation studies. *Mol Ecol* 19(16):3250–3270.

Camargo A, Morando M, Avila LJ, Sites JW Jr (2012) Species delimitation with ABC and other coalescent-based methods: A test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). *Evolution* 66(9):2834–2849.

Campbell MC, Tishkoff SA (2008) African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet* 9:403–433.

Canino MF, Spies IB, Cunningham KM, Hauser L, Grant WS (2010) Multiple ice-age refugia in Pacific cod, *Gadus macrocephalus. Mol Ecol* 19(19):4339–4351.

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325(6099):31–36.

Cannat M, Briais A, Deplus C, Escartín J, Georgen J, Lin J, Mercouriev S, Meyzen C, Muller M, Pouliquen G, Rabain A, da Silva P (1999) Mid-Atlantic Ridge–Azores hotspot interactions: Along-axis migration of a hotspot-derived event of enhanced magmatism 10 to 4 Ma ago. *Earth Planet Sci Lett* 173(3):257–269.

Cannon CH, Morley RJ, Bush ABG (2009) The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. *Proc Natl Acad Sci USA* 106(27):11188–11193.

Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, Kinzig AP, Daily GC, Loreau M, Grace JB, Larigauderie A, Srivastava DS, Naeem S (2012) Biodiversity loss and its impact on humanity. *Nature* 486(7401):59–67.

Cardoso P, Arnedo MA, Triantis KA, Borges PA (2010) Drivers of diversity in Macaronesian spiders and the role of species extinctions. *J Biogeogr* 37(6):1034–1046.

Carling MD, Brumfield RT (2007) Gene sampling strategies for multi-locus population estimates of genetic diversity (θ). *PLoS One* 2(1):e160.

Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues MT, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science* 323(5915):785–789.

Carnaval AC, Waltari E, Rodrigues MT, Rosauer D, VanDerWal J, Damasceno R, Prates I, Strangas M, Spanos Z, Rivera D, Pie MR, Firkowski CR, Bornschein MR, Ribeiro LF, Moritz C (2014) Prediction of phylogeographic endemism in an environmentally complex biome. *Proc R Soc B* 281(1792):20141461.

Carpenter KE, Barber PH, Crandall ED, Ablan-Lagman MCA, Ambariyanto, Mahardika IGN, Manjaji-Matsumoto M, Juinio-Meñez MA, Santos MD, Starger C, Toha AH (2010) Comparative phylogeography of the Coral Triangle and implications for marine management. *J Mar Biol* 2011:396982.

Carr SM, Kivlichan DS, Pepin P, Crutcher DC (1999) Molecular systematics of gadid fishes: Implications for the biogeographic origins of Pacific species. *Can J Zool* 77(1):19–26.

Carroll EL, Baker CS, Watson M, Alderman R, Bannister J, Gaggiotti OE, Gröcke DR, Patenaude N, Harcourt R (2015) Cultural traditions across a migratory network shape the genetic structure of southern right whales around Australia and New Zealand. *Sci Rep* 5:16182.

Carstens BC, Knowles LL (2007) Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: An example from *Melanoplus* grasshoppers. *Syst Biol* 56(3):400–411.

Carstens BC, Richards CL (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution* 61(6):1439–1454.

Carstens BC, Stevenson AL, Degenhardt JD, Sullivan J (2004) Testing nested phylogenetic and phylogeographic hypotheses in the *Plethodon vandykei* species group. *Syst Biol* 53(5):781–792.

Carstens BC, Brunsfeld SJ, Demboski JR, Good JM, Sullivan J (2005) Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: Hypothesis testing within a comparative phylogeographic framework. *Evolution* 59(8):1639–1652.

Carstens BC, Stoute HN, Reid NM (2009) An information-theoretical approach to phylogeography. *Mol Ecol* 18(20):4270–4282.

Carstens BC, Brennan RS, Chua V, Duffie CV, Harvey MG, Koch RA, McMahan CD, Nelson BJ, Newman CE, Satler JD, Seeholzer G, Posbic K, Tank DC, Sullivan J (2013) Model selection as a tool for phylogeographic inference: An example from the willow *Salix melanopsis*. *Mol Ecol* 22(15):4014–4028.

Carto SL, Weaver AJ, Hetherington R, Lam Y, Wiebe EC (2009) Out of Africa and into an ice age: On the role of global climate change in the late Pleistocene migration of early modern humans out of Africa. *J Hum Evol* 56(2):139–151.

Carvalho JC, Cardoso P, Rigal F, Triantis KA, Borges PA (2015) Modeling directional spatio-temporal processes in island biogeography. *Ecol Evol* 5(20):4671–4682.

Cashdan E (2001) Ethnic diversity and its environmental determinants: Effects of climate, pathogens, and habitat diversity. *Am Anthropol* 103(4):968–991.

Castro ALF, Stewart BS, Wilson SG, Hueter RE, Meekan MG, Motta PJ, Bowen BW, Karl SA (2007) Population genetic structure of Earth's largest fish, the whale shark (*Rhincodon typus*). *Mol Ecol* 16(24):5183–5192.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: An analysis tool set for population genomics. *Mol Ecol* 22(11):3124–3140.

Catullo RA, Lanfear R, Doughty P, Keogh JS (2014) The biogeographical boundaries of northern Australia: Evidence from ecological niche models and a multi-locus phylogeny of *Uperoleia* toadlets (Anura: Myobatrachidae). *J Biogeogr* 41(4):659–672.

Cavalli-Sforza LL (2000) *Genes, Peoples, and Languages* (Univ of California Press, Berkeley).

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The History and Geography of Human Genes* (Princeton Univ Press, Princeton, NJ).

Cela-Conde CJ, Lombardo RG, Avise JC, Ayala FJ (2013) In the light of evolution VII: The human mental machinery. *Proc Natl Acad Sci USA* 110(Suppl 2):10339–10462.

Chan LM, Zamudio KR (2009) Population differentiation of temperate amphibians in unpredictable environments. *Mol Ecol* 18(15):3185–3200.

Chan LM, Brown JL, Yoder AD (2011) Integrating statistical genetic and geospatial methods brings new power to phylogeography. *Mol Phylogenet Evol* 59(2):523–537.

Chan YL, Schanzenbach D, Hickerson MJ (2014) Detecting concerted demographic response across community assemblages using hierarchical approximate Bayesian computation. *Mol Biol Evol* 31(9):2501–2515.

Chang C-S, Liu HL, Moncada X, Seelenfreund A, Seelenfreund D, Chung KF (2015) A holistic picture of Austronesian migrations revealed by phylogeography of Pacific paper mulberry. *Proc Natl Acad Sci USA* 112(44):13537–13542.

Chang JT (1999) Recent common ancestors of all present-day individuals. *Adv Appl Probab* 31:1002–1026.

Chapin FS 3rd, Zavaleta ES, Eviner VT, Naylor RL, Vitousek PM, Reynolds HL, Hooper DU, Lavorel S, Sala OE, Hobbie SE, Mack MC, Díaz S (2000) Consequences of changing biodiversity. *Nature* 405(6783):234–242.

Chapman NH, Thompson EA (2003) A model for the length of tracts of identity by descent in finite random mating populations. *Theor Popul Biol* 64(2):141–150.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* 134(4):1289–1303.

Charmantier A, Doutrelant C, Dubuc-Messier G, Fargevieille A, Szulkin M (2015) Mediterranean blue tits as a case study of local adaptation. *Evol Appl* 9(1):135–152.

Chase JM (2011) Ecological niche theory. *The Theory of Ecology,* eds Scheiner SM, Willig MR (Univ of Chicago Press, Chicago), pp 93–108.

Chejanovski ZA, Wiens JJ (2014) Climatic niche breadth and species richness in temperate treefrogs. *J Biogeogr* 41(10):1936–1946.

Chen J-M, Werth S, Sork VL (2016) Comparison of phylogeographical structures of a lichen-forming fungus and its green algal photobiont in western North America. *J Biogeogr* 43(5):932–943.

Cheng H, Sinha A, Cruz FW, Wang X, Edwards RL, d'Horta FM, Ribas CC, Vuille M, Stott LD, Auler AS (2013) Climate change patterns in Amazonia and biodiversity. *Nat Commun* 4:1411.

Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE, Holland J, Hufford MB, Lai J, Li M, Liu X, Lu Y, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhäjärvi T, Rong T, Sekhon RS, Sun Q, Tenaillon MI, Tian F, Wang J, Xu X, Zhang Z, Kaeppler SM, Ross-Ibarra J, McMullen MD, Buckler ES, Zhang G, Xu Y, Ware D (2012) Maize HapMap2 identifies extant variation from a genome in flux. *Nat Genet* 44(7):803–807.

Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30(23):3317–3324.

Cibois A, Thibault JC, Pasquet E (2004) Biogeography of eastern Polynesian monarchs (*Pomarea*): An endemic genus close to extinction. *Condor* 106(54):837–851.

Cibois A, Beadell JS, Graves GR, Pasquet E, Slikas B, Sonsthagen SA, Thibault J-C, Fleischer RC (2011) Charting the course of reed-warblers across the Pacific islands. *J Biogeogr* 38(10):1963–1975.

Claridge EM (2006) The systematics and diversification of *Rhyncogonus* (Entiminae: Curculionidae: Coleoptera) in the central Pacific. PhD Thesis, University of California, Berkeley.

Claridge EM, Gillespie RG, Brewer MS, Roderick GK (2016) Stepping-stones across space and time: Repeated radiation of Pacific flightless broad-nosed weevils (Coleoptera: Curculionidae: Entiminae: *Rhyncogonus*). *J Biogeogr*, in press.

Clark JD, Beyene Y, WoldeGabriel G, Hart WK, Renne PR, Gilbert H, Defleur A, Suwa G, Katoh S, Ludwig KR, Boisserie J-R, Asfaw B, White TD (2003) Stratigraphic, chronological and behavioural contexts of Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature* 423(6941):747–752.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM (2009) The Last Glacial Maximum. *Science* 325(5941):710–714.

Clarke KR, Warwick RM (2001) *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation* (PRIMER-E, Plymouth, UK), 2nd Ed.

Clausen J, Keck DD, Hiesey WM (1947) Heredity of geographically and ecologically isolated races. *Am Nat* 81(797):114–133.

Cleland CA (2001) Historical science, experimental science, and the scientific method. *Geology* 29(11):987–990.

Clendon M (2006) Reassessing Australia's linguistic prehistory. *Curr Anthropol* 47(1):39–61.

Clouard V, Bonneville A (2005) Ages of seamounts, islands, and plateaus on the Pacific plate. *Plates*, *Plumes*, *and Paradigms*, eds Foulger GR, Natland JH, Presnall DC, Anderson DL (Geological Society of America, Boulder, CO), pp 71–90.

Codding BF, Jones TL (2013) Environmental productivity predicts migration, demographic, and linguistic patterns in prehistoric California. *Proc Natl Acad Sci USA* 110(36): 14569–14573.

Coleman RR, Eble JA, DiBattista JD, Rocha LA, Randall JE, Berumen ML, Bowen BW (2016) Regal phylogeography: Range-wide survey of the marine angelfish *Pygoplites diacanthus* reveals evolutionary partitions between the Red Sea, Indian Ocean, and Pacific Ocean. *Mol Phylogenet Evol* 100:243–253.

Collard IF, Foley RA (2002) Latitudinal patterns and environmental determinants of recent human cultural diversity: Do humans follow biogeographical rules? *Evol Ecol Res* 4(3):371–383.

Collevatti RG, Terribile LC, de Oliveira G, Lima-Ribeiro MS, Nabout JC, Rangel TF, Diniz-Filho JAF (2013) Drawbacks to palaeodistribution modelling: The case of South American seasonally dry forests. *J Biogeogr* 40:345–358.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G Jr, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM.(2005) Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307(5717):1928–1933.

Colwell RK, Rangel TF (2010) A stochastic, evolutionary model for range shifts and richness on tropical elevational gradients under Quaternary glacial cycles. *Philos Trans R Soc Lond B Biol Sci* 365(1558):3695–3707.

Comes HP, Kadereit JW (1998) The effect of Quaternary climatic changes on plant distribution and evolution. *Trends Plant Sci* 3(11):432–438.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185(4):1411–1423.

Corander J, Marttinen P, Sirén J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* 9(1):539.

Corbett-Detig RB, Hartl DL, Sackton TB (2015) Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol* 13(4):e1002112.

Corl A, Davis AR, Kuchta SR, Sinervo B (2010) Selective loss of polymorphic mating types is associated with rapid phenotypic evolution during morphic speciation. *Proc Natl Acad Sci USA* 107(9):4254–4259.

Costa LP (2003) The historical bridge between the Amazon and the Atlantic Forest of Brazil: A study of molecular phylogeography with small mammals. *J Biogeogr* 30(1):71–86.

Cowie RH, Holland BS (2008) Molecular biogeography and diversification of the endemic terrestrial fauna of the Hawaiian Islands. *Philos Trans R Soc Lond B Biol Sci* 363(1508): 3363–3376.

Cowman PF, Bellwood DR (2013) The historical biogeography of coral reef fishes: Global patterns of origination and dispersal. *J Biogeogr* 40(2):209–224.

Cox CL, Davis Rabosky AR (2013) Spatial and temporal drivers of phenotypic diversity in polymorphic snakes. *Am Nat* 182(2):E40–E57.

Cox MP (2013) Southeast Asian islands and Oceania: Human genetics. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 293–301.

Cox MP, Nelson MG, Tumonggor MK, Ricaut F-X, Sudoyo H (2012) A small cohort of Island Southeast Asian women founded Madagascar. *Proc Biol Sci* 279(1739):2761–2768.

Coyne JA (2011) Speciation in a small space. *Proc Natl Acad Sci USA* 108(32):12975–12976.

Coyne JA, Orr HA (2004) *Speciation* (Sinauer, Sunderland, MA).

Coyne JA, Price TD (2000) Little evidence for sympatric speciation in island birds. *Evolution* 54(6):2166–2171.

Coyne JA, Barton NH, Turelli M (1997) Perspective: A critique of Sewal Wright's shifting balance theory of evolution. *Evolution* 51(3):643–671.

Cracraft J (1986) Origin and evolution of continental biotas: Speciation and historical congruence within the Australian avifauna. *Evolution* 40(5):977–996.

Craig DA (2003) Geomorphology, development of running water habitats, and evolution of black flies on Polynesian islands. *Bioscience* 53(11):1079–1093.

Craig MT, Eble JA, Bowen BW, Robertson DR (2007) High genetic connectivity across the Indian and Pacific Oceans in the reef fish *Myripristis berndti* (Holocentridae). *Mar Ecol Prog Ser* 334(3):245–254.

Crandall ED, Frey MA, Grosberg RK, Barber PH (2008) Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. *Mol Ecol* 17(2):611–626.

Crandall ED, Sbrocco EJ, Deboer TS, Barber PH, Carpenter KE (2012) Expansion dating: Calibrating molecular clocks in marine species from expansions onto the Sunda Shelf following the Last Glacial Maximum. *Mol Biol Evol* 29(2):707–719.

Craul M, Zimmermann E, Rasoloharijaona S, Randrianambinina B, Radespiel U (2007) Unexpected species diversity of Malagasy primates (*Lepilemur* spp.) in the same biogeographical zone: A morphological and molecular approach with the description of two new species. *BMC Evol Biol* 7:83.

Crawford AJ, Bermingham E, Carolina PS (2007) The role of tropical dry forest as a long-term barrier to dispersal: A comparative phylogeographical analysis of dry forest tolerant and intolerant frogs. *Mol Ecol* 16(22):4789–4807.

Criscione CD, Cooper B, Blouin MS (2006) Parasite genotypes identify source populations of migratory fish more accurately than fish genotypes. *Ecology* 87(4):823–828.

Crispo E (2008) Modifying effects of phenotypic plasticity on interactions among natural selection, adaptation and gene flow. *J Evol Biol* 21(6):1460–1469.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* 23(13):3133–3157.

Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian computation (ABC) in practice. *Trends Ecol Evol* 25(7):410–418.

Csilléry K, François O, Blum MGB (2012) ABC: An R package for approximate Bayesian computation. *Methods Ecol Evol* 3(3):475–479.

Currie TE, Mace R (2012) The evolution of ethnolinguistic diversity. *Adv Complex Syst* 15(1–2):1150006–1150026.

Cutter AD (2013) Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. *Mol Phylogenet Evol* 69(3):1172–1185.

Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga JL, Meyer M (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci USA* 110(39):15758–15763.

Dana JD (1853) On the isothermal oceanic chart: Illustrating the geographical distribution of marine animals. *Am J Sci* 16:314–327.

Darlington PJ Jr (1959) Area, climate, and evolution. *Evolution* 13(4):488–510.

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).

Darwin C, Wallace A (1858) On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Proc Linn Soc London* 3(9):45–62.

Dasmahapatra KK, Lamas G, Simpson F, Mallet J (2010) The anatomy of a "suture zone" in Amazonian butterflies: A coalescent-based test for vicariant geographic divergence and speciation. *Mol Ecol* 19(19):4283–4301.

Dauby G, Duminil J, Heuertz M, Koffi GK, Stévart T, Hardy OJ (2014) Congruent phylogeographical patterns of eight tree species in Atlantic Central Africa provide insights into the past dynamics of forest cover. *Mol Ecol* 23(9):2299–2312.

Dawson MN (2014) Natural experiments and meta-analyses in comparative phylogeography. *J Biogeogr* 41(1):52–65.

De Maio N, Wu C-H, O'Reilly KM, Wilson D (2015) New routes to phylogeography: A Bayesian structured coalescent approximation. *PLoS Genet* 11(8):e1005421.

de Oliveira PE, Barreto AMF, Suguio K (1999) Late Pleistocene/Holocene climatic and vegetational history of the Brazilian Caatinga: The fossil dunes of the middle São Francisco River. *Palaeogeogr Palaeoclimatol Palaeoecol* 152:319–337.

de Queiroz A (2005) The resurrection of oceanic dispersal in historical biogeography. *Trends Ecol Evol* 20(2):68–73.

De Queiroz K (2007) Species concepts and species delimitation. *Syst Biol* 56(6):879–886.

De Vos JM, Joppa LN, Gittleman JL, Stephens PR, Pimm SL (2015) Estimating the normal background rate of species extinction. *Conserv Biol* 29(2):452–462.

Deagle BE, Jones FC, Absher DM, Kingsley DM, Reimchen TE (2013) Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Mol Ecol* 22(7):1917–1932.

Debiasse MB, Nelson BJ, Hellberg ME (2014) Evaluating summary statistics used to test for incomplete lineage sorting: Mito-nuclear discordance in the reef sponge *Callyspongia vaginalis*. *Mol Ecol* 23(1):225–238.

DeChaine EG, Martin AP (2005) Historical biogeography of two alpine butterflies in the Rocky Mountains: Broad-scale concordance and local-scale discordance. *J Biogeogr* 32(3):1943–1956.

Degnan JH (2010) Probabilities of gene trees with intraspecific sampling given a species tree. *Estimating Species Trees: Practical and Theoretical Aspects*, eds Knowles LL, Kubatko LS (Wiley-Blackwell, Hoboken, NJ), pp 53–78.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol* 24(6):332–340.

Degnan JH, Salter LA (2005) Gene tree distributions under the coalescent process. *Evolution* 59(1):24–37.

Dellicour S, Michez D, Mardulyn P (2015) Comparative phylogeography of five bumblebees: Impact of range fragmentation, range size and diet specialization. *Biol J Linn Soc Lond* 116(4):926–939.

Delmore KE, Hübner S, Kane NC, Schuster R, Andrew RL, Câmara F, Guigó R, Irwin DE (2015) Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Mol Ecol* 24(8):1873–1888.

Delrieu-Trottin E, Williams JT, Bacchet P, Kulbicki M, Mourier J, Galzin R, Lison de Loma T, Mou-Tham G, Siu G, Planes S (2015) Shore fishes of the Marquesas Islands: An updated checklist with new records and new percentage of endemic species. *Check List* 11(5):1–13.

Demesure B, Comps B, Petit RJ (1996) Chloroplast DNA phylogeography of the common beech *Fagus sylvatica* L. in Europe. *Evolution* 50(6):2515–2520.

Demeter F, Shackelford LL, Bacon AM, Duringer P, Westaway K, Sayavongkhamdy T, Braga J, Sichanthongtip P, Khamdalavong P, Ponche JL, Wang H, Lundstrom C, Patole-Edoumba E, Karpoff AM (2012) Anatomically modern human in Southeast Asia (Laos) by 46 ka. *Proc Natl Acad Sci USA* 109(36):14375–14380.

Demos TC, Kerbis Peterhans JC, Agwanda B, Hickerson MJ (2014) Uncovering cryptic diversity and refugial persistence among small mammal lineages across the Eastern Afromontane biodiversity hotspot. *Mol Phylogenet Evol* 71(1):41–54.

Dentener F, Drevet J, Lamarque JF, Bey I, Eickhout B, Fiore AM, Hauglustaine D, Horowitz LW, Krol M, Kulshrestha UC, Lawrence M, Galy-Lacaux C, Rast S, Shindell D, Stevenson D, Van Noije T, Atherton C, Bell N, Bergman D, Butler T, Cofala J, Collins B, Doherty R, Ellingsen K, Galloway J, Gauss M, Montanaro V, Müller JF, Pitari G, Rodriguez J, Sanderson M, Solmon F, Strahan S, Schultz M, Sudo K, Szopa S, Wild O (2006) Nitrogen and sulfur deposition on regional and global scales: A multimodel evaluation. *Global Biogeochem Cy* 20(4):GB4003.

Derrida B, Manrubia SC, Zanette DH (2000) On the genealogy of a population of biparental individuals. *J Theor Biol* 203(3):303–315.

Dethmers KE, Broderick D, Moritz C, Fitzsimmons NN, Limpus CJ, Lavery S, Whiting S, Guinea M, Prince RI, Kennett R (2006) The genetic structure of Australasian green turtles (*Chelonia mydas*): Exploring the geographical scale of genetic exchange. *Mol Ecol* 15(13):3931–3946.

Dewar RE (2014) Early human settlers and their impacts on Madagascar's landscapes. *Conservation and Environmental Management in Madagascar,* ed Scales IR (Earthscan/Routledge, Oxen, NY), pp 44–64.

Dewar RE, Richard AF (2007) Evolution in the hypervariable environment of Madagascar. *Proc Natl Acad Sci USA* 104(34):13723–13727.

Dewar RE, Wright HT (1993) The culture history of Madagascar. *J World Prehist* 7(4):417–466.

Dewar RE, Radimilahy C, Wright HT, Jacobs Z, Kelly GO, Berna F (2013) Stone tools and foraging in northern Madagascar challenge Holocene extinction models. *Proc Natl Acad Sci USA* 110(31):12583–12588.

d'Horta FM, Cuervo AM, Ribas CC, Brumfield RT, Miyaki CY (2013) Phylogeny and comparative phylogeography of *Sclerurus* (Aves: Furnariidae) reveal constant and cryptic diversification in an old radiation of rain forest understorey specialists. *J Biogeogr* 40(1):37–49.

Diaz S, Cabido M (2001) Vive la difference: Plant functional diversity matters to ecosystem processes. *Trends Ecol Evol* 16(11):646–655.

Diaz-Jaimes P, Uribe-Alcocer M, Rocha-Olivares A, García-de-León FJ, Nortmoon P, Durand JD (2010) Global phylogeography of the dolphinfish (*Coryphaena hippurus*): The influence of large effective population size and recent dispersal on the divergence of a marine pelagic cosmopolitan species. *Mol Phylogenet Evol* 57(3):1209–1218.

Díaz-Pérez A, Sequeira M, Santos-Guerra A, Catalán P (2008) Multiple colonizations, in situ speciation, and volcanism-associated stepping-stone dispersals shaped the phylogeography of the Macaronesian red fescues (*Festuca* L., Gramineae). *Syst Biol* 57(5):732–749.

DiBattista J, Wilcox C, Craig M, Rocha L, Bowen B (2011) Phylogeography of the Pacific blueline surgeonfish, *Acanthurus nigroris*, reveals high genetic connectivity and a cryptic endemic species in the Hawaiian archipelago. *J Mar Biol* 2011:839134.

DiBattista JD, Rocha LA, Craig MT, Feldheim KA, Bowen BW (2012) Phylogeography of two closely related Indo-Pacific butterflyfishes reveals divergent evolutionary histories and discordant results from mtDNA and microsatellites. *J Hered* 103(5):617–629.

DiBattista JD, Berumen ML, Gaither MR, Rocha LA, Eble JA, Choat JH, Craig MT, Skillings DJ, Bowen BW (2013) After continents divide: Comparative phylogeography of reef fishes from the Red Sea and Indian Ocean. *J Biogeogr* 40(6):1170–1181.

DiBattista JD, Waldrop E, Rocha LA, Craig MT, Berumen ML, Bowen BW (2015) Blinded by the bright: A lack of congruence between colour morphs, phylogeography and taxonomy for a cosmopolitan Indo-Pacific butterflyfish, *Chaetodon auriga. J Biogeogr* 42(10):1919–1929.

DiBattista JD, Choat JH, Gaither MR, Hobbs J-PA, Lozano-Cortés DF, Myers RF, Paulay G, Rocha LA, Toonen RJ, Westneat MW, Berumen ML (2016a) On the origin of endemic species in the Red Sea. *J Biogeogr* 43(1):13–30.

DiBattista JD, Roberts MB, Bouwmeester J, Bowen BW, Coker DJ, Lozano-Cortés DF, Choat JH, Gaither MR, Hobbs J-PA, Khalil MT, Kochzius M, Myers RF, Paulay G, Robitzch VSN, Saenz-Agudelo P, Salas E, Sinclair-Taylor TH, Toonen RJ, Westneat MW, Williams ST, Berumen ML (2016b) A review of contemporary patterns of endemism for shallow water reef fauna in the Red Sea. *J Biogeogr* 43(3):423–439.

Dickinson WR (1998) Geomorphology and geodynamics of the Cook-Austral Island Seamount chain in the South Pacific Ocean: Implications for hotspots and plumes. *Int Geol Rev* 40(12):1039–1075.

Dierickx EG, Shultz AJ, Sato F, Hiraoka T, Edwards SV (2015) Morphological and genomic comparisons of Hawaiian and Japanese black-footed albatrosses (*Phoebastria nigripes*) using double digest RADseq: Implications for conservation. *Evol Appl* 8(7):662–678.

Dillehay TD (1999) The Late Pleistocene cultures of South America. *Evol Anthropol* 7(6):206–216.

Dillehay TD, Ramírez C, Pino M, Collins MB, Rossen J, Pino-Navarro JD (2008) Monte Verde: Seaweed, food, medicine, and the peopling of South America. *Science* 320(5877):784–786.

Dillehay TD, Ocampo C, Saavedra J, Sawakuchi AO, Vega RM, Pino M, Collins MB, Scott Cummings L, Arregui I, Villagran XS, Hartmann GA, Mella M, González A, Dix G (2015) New archaeological evidence for an early human presence at Monte Verde, Chile. *PLoS One* 10(11):e0141923.

Dimitrov D, Arnedo MA, Ribera C (2008) Colonization and diversification of the spider genus *Pholcus* Walckenaer, 1805 (Araneae, Pholcidae) in the Macaronesian archipelagos: Evidence for long-term occupancy yet rapid recent speciation. *Mol Phylogenet Evol* 48(2):596–614.

Dobzhansky TH (1940) Speciation as a stage in evolutionary divergence. *Am Nat* 74(753):312–321.

Dobzhansky T (1973) Nothing in biology makes sense except in the light of evolution. *Am Biol Teach* 35(3):125–129.

Dobzhansky T, Anderson WW, Pavlovsky O (1966) Genetics of natural populations. XXXVIII. Continuity and change in populations of *Drosophila pseudoobscura* in western United States. *Evolution* 20(3):418–427.

Dolby GA, Bennett SEK, Lira-Noriega A, Wilder BT, Muguia-Vega A (2015) Assessing the geological and climatic forcing of biodiversity and evolution surrounding the Gulf of California. *J Southwest* 57(2-3):391–456.

Dolman G, Hugall AF (2008) Combined mitochondrial and nuclear data enhance resolution of a rapid radiation of Australian rainbow skinks (Scincidae: *Carlia*). *Mol Phylogenet Evol* 49(3):782–794.

Dolman G, Joseph L (2012) A species assemblage approach to comparative phylogeography of birds in southern Australia. *Ecol Evol* 2(2):354–369.

Dolman G, Moritz C (2006) A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*). *Evolution* 60(3):573–582.

Donoghue MJ, Moore BR (2003) Toward an integrative historical biogeography. *Integr Comp Biol* 43(2):261–270.

Dopman EB, Pérez L, Bogdanowicz SM, Harrison RG (2005) Consequences of reproductive barriers for genealogical discordance in the European corn borer. *Proc Natl Acad Sci USA* 102(41):14706–14711.

Dow MM, Cheverud JM, Friedlaender JS (1987) Partial correlation of distance matrices in studies of population structure. *Am J Phys Anthropol* 72(3):343–352.

Drew JA, Barber PH (2012) Comparative phylogeography in Fijian coral reef fishes: A multitaxa approach towards marine reserve design. *PLoS One* 7(10):e47710.

Drews C, Han W (2010) Dynamics of wind setdown at Suez and the Eastern Nile Delta. *PLoS One* 5(8):e12481.

Driver JC, Handly M, Fladmark K, Nelson D, Sullivan G, Preston R. (1996). Stratigraphy, radiocarbon dating, and culture history of Charlie Lake Cave, British Columbia. *Arctic* 49(3):265–277.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8):1969–1973.

Dumont BL, Payseur BA (2008) Evolution of the genomic rate of recombination in mammals. *Evolution* 62(2):276–294.

Duncan KM, Martin AP, Bowen BW, De Couet HG (2006) Global phylogeography of the scalloped hammerhead shark (*Sphyrna lewini*). *Mol Ecol* 15(8):2239–2251.

Duncan SI, Crespi EJ, Mattheus NM, Rissler LJ (2015) History matters more when explaining genetic diversity within the context of the core-periphery hypothesis. *Mol Ecol* 24(16):4323–4336.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Mol Biol Evol* 28(8):2239–2252.

Dutton PH, Bowen BW, Owens DW, Barragan A, Davis SK (1999) Global phylogeography of the leatherback turtle (*Dermochelys coriacea*). *J Zool* (*Lond*) 248(3):397–409.

Dwyer GS, Cronin TM, Baker PA, Raymo ME, Buzas JS, Correge T (1995) North Atlantic deepwater temperature change during late Pliocene and late Quaternary climatic cycles. *Science* 270(5240):1347–1351.

Dyer RJ (2015) Is there such a thing as landscape genetics? *Mol Ecol* 24(14):3518–3528.

Dyer RJ, Nason JD, Garrick RC (2010) Landscape modelling of gene flow: Improved power using conditional genetic distance derived from the topology of population networks. *Mol Ecol* 19(17):3746–3759.

Dyke AS (2004) An outline of North American deglaciation with emphasis on central and northern Canada. *Quaternary Glaciations: Extent and Chronology Part II: North America*, eds Ehlers J, Gibbard JE (Elsevier, Amsterdam), pp 373–424.

Dyke AS (2005) Late Quaternary vegetation history of northern North America based on pollen, macrofossil, and faunal remains. *Geogr Phys Quat* 59(2-3):211–262.

Eastman JT (2005) The nature of the diversity of Antarctic fishes. *Polar Biol* 28(2):93–107.

Eaton DAR (2014) PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* 30(13):1844–1849.

Eberle M, Kappeler PM (2004) Sex in the dark: Determinants and consequences of mixed male mating tactics in *Microcebus murinus*, a small solitary nocturnal primate. *Behav Ecol Sociobiol* 57(1):77–90.

Eble JA, Bowen BW, Bernardi G (2015) Phylogeography of coral reef fishes. *Ecology of Fishes on Coral Reefs,* ed Mora C (Univ of Hawaii Press, Honolulu), pp 64–75.

Eble JA, Rocha LA, Craig MT, Bowen BW (2011) Not all larvae stay close to home: Insights into marine population connectivity with a focus on the brown surgeonfish (*Acanthurus nigrofuscus*). *J Mar Biol* 2011:518516.

Eckert AJ, Dyer RJ (2012) Defining the landscape of adaptive genetic diversity. *Mol Ecol* 21(12):2836–2838.

Edwards CJ, Suchard MA, Lemey P, Welch JJ, Barnes I, Fulton TL, Barnett R, O'Connell TC, Coxon P, Monaghan N, Valdiosera CE, Lorenzen ED, Willerslev E, Baryshnikov GF, Rambaut A, Thomas MG, Bradley DG, Shapiro B (2011) Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr Biol* 21(15):1251–1258.

Edwards SV (1993) Long-distance gene flow in a cooperative breeder detected in genealogies of mitochondrial DNA sequences. *Proc Biol Sci* 252(1335):177–185.

Edwards SV (2016) Inferring species trees. *Encyclopedia of Evolutionary Biology,* ed Kliman R (Elsevier, New York).

Edwards SV, Beerli P (2000) Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* 54(6):1839–1854.

Edwards S, Bensch S (2009) Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Mol Ecol* 18(14):2930–2933, discussion 2934–2936.

Edwards SV, Shultz AJ, Campbell-Staton C (2015) Next-generation sequencing and the expanding domain of phylogeography. *Folia Zool (Brno)* 64(3):187–206.

Edwards SV, Xi Z, Janke A, Faircloth BC, McCormack JE, Glenn TC, Zhong B, Wu S, Lemmon EM, Lemmon AR, Leaché AD, Liu L, Davis CC (2016) Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Mol Phylogenet Evol* 94(Pt A):447–462.

Efromovich S, Kubatko LS (2008) Coalescent time distributions in trees of arbitrary size. *Stat Appl Genet Mol Biol* 7(1):2.

Ehret C (2013) Sub-Saharan Africa: Linguistics. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 96–106.

Eldon B, Degnan JH (2012) Multiple merger gene genealogies in two species: Monophyly, paraphyly, and polyphyly for two examples of Lambda coalescents. *Theor Popul Biol* 82(2):117–130.

Eldridge MD, Potter S, Cooper SJ (2011) Biogeographic barriers in north-western Australia: An overview and standardisation of nomenclature. *Aust J Zool* 59(4):270–272.

Eldridge MD, Potter S, Johnson CN, Ritchie EG (2014) Differing impact of a major biogeographic barrier on genetic structure in two large kangaroos from the monsoon tropics of Northern Australia. *Ecol Evol* 4(5):554–567.

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson KS, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29(2):129–151.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. *Div Distr* 17(1):43–57.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A, Uebbing S, Wolf JB (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* 491(7426):756–760.

Ellis N, Smith SJ, Pitcher CR (2012) Gradient forests: Calculating importance gradients on physical predictors. *Ecology* 93(1):156–168.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6(5):e19379.

Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proc Natl Acad Sci USA* 107:16196–16200.

Ence DD, Carstens BC (2011) SpedeSTEM: A rapid and accurate method for species delimitation. *Mol Ecol Resour* 11(3):473–480.

Endler JA (1973) Gene flow and population differentiation. *Science* 179(4070):243–250.

Epling C, Dobzhansky T (1942) Genetics of natural populations. VI. Microgeographic races of *Linanthus parryae. Genetics* 27(3):317–332.

Epps CW, Keyghobadi N (2015) Landscape genetics in a changing world: Disentangling historical and contemporary influences and inferring change. *Mol Ecol* 24(24):6021–6040.

Eriksson A, Manica A (2012) Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proc Natl Acad Sci USA* 109(35):13956–13960.

Eriksson A, Betti L, Friend AD, Lycett SJ, Singarayer JS, von Cramon-Taubadel N, Valdes PJ, Balloux F, Manica A (2012) Late Pleistocene climate change and the global expansion of anatomically modern humans. *Proc Natl Acad Sci USA* 109(40):16089–16094.

Erlandson JM (2013) After Clovis-first collapsed: Reimagining the peopling of the Americas. *Paleoamerican Odyssey,* eds Graf KE, Ketron CV, Waters MR (Texas A&M Univ Press, College Station), pp 127–131.

Espíndola A, Alvarez N (2011) Comparative phylogeography in a specific and obligate pollination antagonism. *PLoS One* 6(12):e28662.

Espindola A, Carstens BC, Alvarez N (2014) Comparative phylogeography of mutualists and the effect of the host on the genetic structure of its partners. *Biol J Linn Soc Lond* 113:1021–1035.

Ewens WJ (1990) Population genetics theory—the past and the future. *Mathematical and Statistical Developments of Evolutionary Theory*, ed Lessard S (Kluwer, Amsterdam), pp 177–227.

Ewens WJ (2004) *Mathematical Population Genetics, Volume I: Theoretical Foundations* (Springer, Berlin).

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564–567.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genet* 9(10):e1003905.

Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L (2007) Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci USA* 104(45):17614–17619.

Fagundes NJR, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM, Smith DG, Silva WA Jr, Zago MA, Ribeiro-dos-Santos AK, Santos SE, Petzl-Erler ML, Bonatto SL (2008) Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* 82(3):583–592.

Faivovich J, Ferraro DP, Basso NG, Haddad CFB, Rodrigues MT, Wheeler WC, Lavilla EO (2012) A phylogenetic analysis of *Pleurodema* (Anura: Leptodactylidae: Leiuperinae) based on mitochondrial and nuclear gene sequences, with comments on the evolution of anuran foam nests. *Cladistics* 28(5):460–482.

Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. *Science* 320(5879):1034–1039.

Falush D, Stephens M, Pritchard JK (2003a) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164(4):1567–1587.

Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, Yamaoka Y, Mégraud F, Otto K, Reichard U, Katzowitsch E, Wang X, Achtman M, Suerbaum S (2003b) Traces of human migrations in *Helicobacter pylori* populations. *Science* 299(5612):1582–1585.

Faria CMA, Machado A, Amorim IR, Gage MJG, Borges PAV, Emerson BC (2016) Evidence for multiple founding lineages and genetic admixture in the evolution of species within an oceanic island weevil (Coleoptera, Curculiondae) super-radiation. *J Biogeogr* 43(1):178–191.

Farnelid H, Andersson AF, Bertilsson S, Al-Soud WA, Hansen LH, Sørensen S, Steward GF, Hagström Å, Riemann L (2011) Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. *PLoS One* 6(4):e19223.

Faulks L, Svanbäck R, Eklöv P, Östman Ö (2015) Genetic and morphological divergence along the littoral–pelagic axis in two common and sympatric fishes: Perch, *Perca fluviatilis* (Percidae) and roach, *Rutilusrutilus* (Cyprinidae). *Biol J Linn Soc Lond* 114(4):929–940.

Fay JC, Wu C-I (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155(3): 1405–1413.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends Genet* 28(7):342–350.

Fedje DW, White J, Wilson M, Nelson D, Vogel J, Southon J (1995) Vermilion Lakes site: Adaptations and environments in the Canadian Rockies during the latest Pleistocene and early Holocene. *Am Antiq* 60(1):81–108.

Feldman CR, Brodie ED Jr, Brodie ED 3rd, Pfrender ME (2009) The evolutionary origins of beneficial alleles during the repeated adaptation of garter snakes to deadly prey. *Proc Natl Acad Sci USA* 106(32):13415–13420.

Felsenstein J (2006) Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? *Mol Biol Evol* 23(3):691–700.

Fenner JN (2005) Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol* 128(2):415–423.

Fernandez-Silva I, Andrews KR, Baums IB, Bird CE, Concepcion GT, Daly-Engel TS, Eble JA, Faucci A, Gaither MR, Iacchei M, Puritz JB, Schultz JK, Skillings DJ, Timmers MA, Bowen BW (2015) Yellow tails in the Red Sea: Phylogeography of the Indo-Pacific goatfish *Mulloidichthys flavolineatus* reveals isolation in peripheral provinces and cryptic evolutionary lineages. *J Biogeogr* 42(12):2402–2413.

Ferrier S, Manion G, Elith J, Richardson K (2007) Using generalized dissimilarity modeling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers Distrib* 13(3):252–264.

Field JS, Lahr MM (2005) Assessment of the southern dispersal: GIS-based analyses of potential routes at oxygen isotopic stage 4. *J World Prehist* 19(1):1–45.

Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* 103(3):626–631.

Fierer N, Strickland MS, Liptzin D, Bradford MA, Cleveland CC (2009) Global patterns in belowground communities. *Ecol Lett* 12(11):1238–1249.

Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA, Wall DH, Caporaso JG (2012) Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc Natl Acad Sci USA* 109(52):21390–21395.

Fierer N, Ladau J, Clemente JC, Leff JW, Owens SM, Pollard KS, Knight R, Gilbert JA, McCulley RL (2013) Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. *Science* 342(6158):621–624.

Fierer N, Barberán A, Laughlin DC (2014) Seeing the forest for the genes: Using metagenomics to infer the aggregated traits of microbial communities. *Front Microbiol* 5:614.

Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ (2012) Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484(7393):186–194.

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

Fitzpatrick BM, Johnson JR, Kump DK, Smith JJ, Voss SR, Shaffer HB (2010) Rapid spread of invasive genes into a threatened native species. *Proc Natl Acad Sci USA* 107(8):3606–3610.

Fitzpatrick MC, Keller SR (2015) Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecol Lett* 18(1):1–16.

Fladmark KR (1986) The first Americans: Getting one's Berings. *Nat Hist* 95(11):8–17.

Fleischer RC, McIntosh CE (2001) Molecular systematics and biogeography of the Hawaiian avifauna. *Stud Avian Biol* 22:51–60.

Floeter SR, Rocha LA, Robertson DR, Joyeux JC, Smith-Vaniz WF, Wirtz P, Edwards AJ, Barreiros JP, Ferreira CEL, Gasparini JL, Brito A, Falcón JM, Bowen BW, Bernardi G (2008) Atlantic reef fish biogeography and evolution. *J Biogeogr* 35(1):22–47.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* 180(2):977–993.

Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, Jiang X, Hall AB, Catteruccia F, Kakani E, Mitchell SN, Wu YC, Smith HA, Love RR, Lawniczak MK, Slotman MA, Emrich SJ, Hahn MW, Besansky NJ (2015) Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 347(6217):1258524.

Forbes E (1859) *The Natural History of the European Seas* (John Van Voorst, London).

Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Mol Ecol* 25(1):104–120.

Forsman Z, Wellington GM, Fox GE, Toonen RJ (2015) Clues to unraveling the coral species problem: Distinguishing species from geographic variation in *Porites* across the Pacific with molecular markers and microskeletal traits. *PeerJ* 3:e751.

Forster P, Renfrew C (2011) Evolution. Mother tongue and Y chromosomes. *Science* 333(6048): 1390–1391.

Fortescue M (2013) North America: Eskimo-Aleut linguistic history. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 340–345.

Fouquet A, Noonan BP, Rodrigues MT, Pech N, Gilles A, Gemmell NJ (2012) Multiple Quaternary refugia in the eastern Guiana Shield revealed by comparative phylogeography of 12 frog species. *Syst Biol* 61(3):461–489.

Fowler D, Coyle M, Skiba U, Sutton MA, Cape JN, Reis S, Sheppard LJ, Jenkins A, Grizzetti B, Galloway JN, Vitousek P, Leach A, Bouwman AF, Butterbach-Bahl K, Dentener F, Stevenson D, Amann M, Voss M (2013) The global nitrogen cycle in the twenty-first century. *Philos Trans R Soc Lond B Biol Sci* 368(1621):20130164.

Frichot E, Mathieu F, Trouillon T, Bouchard G, François O (2014) Fast and efficient estimation of individual ancestry coefficients. *Genetics* 196(4):973–983.

Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, Patterson N, Rohland N, Lazaridis I, Nickel B, Viola B, Prüfer K, Meyer M, Kelso J, Reich D, Pääbo S (2015) An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 524(7564):216–219.

Fuchs J, Parra JL, Goodman SM, Raherilalao MJ, Vanderwal J, Bowie RCK (2013) Extending ecological niche models to the past 120,000 years corroborates the lack of strong phylogeographic structure in the crested drongo (*Dicrurus forficatus forficatus*) on Madagascar. *Biol J Linn Soc Lond* 108(3):658–676.

Fukami T (2015) Historical contingency in community assembly: Integrating niches, species pools, and priority effects. *Annu Rev Ecol Evol Syst* 46:1–23.

Fukami T, Beaumont HJE, Zhang XX, Rainey PB (2007) Immigration history controls diversification in experimental adaptive radiation. *Nature* 446(7134):436–439.

Funk D, Omland K (2003) Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu Rev Ecol Evol Syst* 34:397–423.

Funk DJ, Egan SP, Nosil P (2011) Isolation by adaptation in Neochlamisus leaf beetles: Host-related selection promotes neutral genomic divergence. *Mol Ecol* 20(22):4671–4682.

Funk VA, Wagner WL (1995a) Biogeographic patterns in the Hawaiian Islands. *Hawaiian Biogeography: Evolution on a Hot Spot Archipelago*, eds Wagner WL, Funk VA (Smithsonian Institution Press, Washington, DC), pp 379–419.

Funk VA, Wagner WL, eds (1995b) *Hawaiian Biogeography Evolution on a Hot Spot Archipelago* (Smithsonian Institution Press, Washington, DC).

Gagneux P, Wills C, Gerloff U, Tautz D, Morin PA, Boesch C, Fruth B, Hohmann G, Ryder OA, Woodruff DS (1999) Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proc Natl Acad Sci USA* 96(9):5077–5082.

Gaither MR, Toonen RJ, Robertson DR, Planes S, Bowen BW (2010) Genetic evaluation of marine biogeographic barriers: Perspectives from two widespread Indo-Pacific snappers (*Lutjanus* spp.). *J Biogeogr* 37(1):133–147.

Gaither MR, Bowen BW, Bordenave T-R, Rocha LA, Newman SJ, Gomez JA, van Herwerden L, Craig MT (2011a) Phylogeography of the reef fish *Cephalopholis argus* (Epinephelidae) indicates Pleistocene isolation across the Indo-Pacific Barrier with contemporary overlap in the Coral Triangle. *BMC Evol Biol* 11:189.

Gaither MR, Jones SA, Kelley C, Newman SJ, Sorenson L, Bowen BW (2011b) High connectivity in the deepwater snapper *Pristipomoides filamentosus* (Lutjanidae) across the Indo-Pacific with isolation of the Hawaiian archipelago. *PLoS One* 6(12):e28913.

Gaither MR, Bernal MA, Coleman RR, Bowen BW, Jones SA, Simison WB, Rocha LA (2015a) Genomic signatures of geographic isolation and natural selection in coral reef fishes. *Mol Ecol* 24(7):1543–1557.

Gaither MR, Bernal MA, Fernandez-Silva I, Mwale M, Jones SA, Rocha C, Rocha L (2015b) Two deep evolutionary lineages in the circumtropical glasseye *Heteropriacanthus cruentatus* (Teleostei, Priacanthidae) with admixture in the south-western Indian Ocean. *J Fish Biol* 87(3):715–727.

Gaither MR, Bowen BW, Rocha LA, Briggs JC (2016) Fishes that rule the world: Circumtropical distributions revisited. *Fish Fish* 17:664–679.

Gallego Llorente M, Jones ER, Eriksson A, Siska V, Arthur KW, Arthur JW, Curtis MC, Stock JT, Coltorti M, Pieruccini P, Stretton S, Brock F, Higham T, Park Y, Hofreiter M, Bradley DG, Bhak J, Pinhasi R, Manica A (2015) Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* 350(6262):820–822.

Ganzhorn JU, Schmid J (1998) Different population dynamics of *Microcebus murinus* in primary and secondary deciduous dry forests of Madagascar. *Int J Primatol* 19(5):785–796.

Garb JE, Gillespie RG (2006) Island hopping across the central Pacific: Mitochondrial DNA detects sequential colonization of the Austral Islands by crab spiders (Araneae: Thomisidae). *J Biogeogr* 33(2):201–220.

Garrick RC, Rowell DM, Simmons CS, Hillis DM, Sunnucks P (2008) Fine-scale phylogeographic congruence despite demographic incongruence in two low-mobility saproxylic springtails. *Evolution* 62(5):1103–1118.

Garrick RC, Rowell DM, Sunnucks P (2012) Phylogeography of saproxylic and forest floor invertebrates from Tallaganda, South-eastern Australia. *Insects* 3(1):270–294.

Garrick RC, Bonatelli IA, Hyseni C, Morales A, Pelletier TA, Perez MF, Rice E, Satler JD, Symula RE, Thomé MT, Carstens BC (2015) The evolution of phylogeographic data sets. *Mol Ecol* 24(6):1164–1171.

Gasse F, Van Campo E (2001) Late Quaternary environmental changes from a pollen and diatom record in the southern tropics (Lake Tritrivakely, Madagascar). *Palaeogeogr Palaeoclimatol Palaeoecol* 167(3–4):287–308.

Gasse F, Cortijo E, Disnar JR, Ferry L, Gibert E, Kissel C, Laggoun-Defarge F, Lallier-Vergès E, Miskovsky JC, Ratsimbazafy B, Ranaivo F, Robison L, Tucholka P, Saos J-L, Sifeddine A, Taieb M, Van Campo E, Williamson D (1994) A 36 ka environmental record in the southern tropics—Lake Tritrivakely (Madagascar). *C R Seances Acad Sci IIa* 318(11):1513–1519.

Gautier L, Goodman SM (2003) Introduction to the flora of Madagascar. *The Natural History of Madagascar,* eds Goodman SM, Benstead J (Univ of Chicago Press, Chicago, IL), pp 229–250.

Gautier M, Foucaud J, Gharbi K, Cézard T, Galan M, Loiseau A, Thomson M, Pudlo P, Kerdelhué C, Estoup A (2013) Estimation of population allele frequencies from next-generation sequencing data: Pool-versus individual-based genotyping. *Mol Ecol* 22(14):3766–3779.

Gavin DG, Fitzpatrick MC, Gugger PF, Heath KD, Rodríguez-Sánchez F, Dobrowski SZ, Hampe A, Hu FS, Ashcroft MB, Bartlein PJ, Blois JL, Carstens BC, Davis EB, de Lafontaine G, Edwards ME, Fernandez M, Henne PD, Herring EM, Holden ZA, Kong WS, Liu J, Magri D, Matzke NJ, McGlone MS, Saltré F, Stigall AL, Tsai YH, Williams JW (2014) Climate refugia: Joint inference from fossil records, species distribution models and phylogeography. *New Phytol* 204(1):37–54.

Gehara M, Summers K, Brown JL (2013) Population expansion, isolation and selection: Novel insights on the evolution of color diversity in the strawberry poison frog. *Evol Ecol* 27(4):797–824.

George R (2005) Tethys Sea fragmentation and speciation of *Panulirus* spiny lobsters. *Crustaceana* 78(11):1281–1309.

Gerard A, Ganzhorn JU, Kull CA, Carriere SM (2015) Possible roles of introduced plants for native vertebrate conservation: The case of Madagascar. *Restor Ecol* 23(6):768–775.

Ghalambor CK, Huey RB, Martin PR, Tewksbury JJ, Wang G (2006) Are mountain passes higher in the tropics? Janzen's hypothesis revisited. *Integr Comp Biol* 46(1):5–17.

Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA (2013) Improving Bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol Biol Evol* 30(3):713–724.

Gillespie RG (2002) Colonization of remote oceanic islands of the Pacific: Archipelagos as stepping stones? *J Biogeogr* 29(5–6):655–662.

Gillespie RG (2003) Marquesan spiders of the genus *Tetragnatha*. *J Arachnol* 31(1):62–77.

Gillespie RG (2015) Island time and the interplay between ecology and evolution in species diversification. *Evol Appl* 9(1):53–73.

Gillespie RG, Baldwin BG (2009) Island biogeography of remote archipelagos: Interplay between ecological and evolutionary processes. *The Theory of Island Biogeography Revisited*, eds Losos J, Ricklefs R (Princeton Univ Press, Princeton, NJ), pp 358–387.

Gillespie RG, Roderick GK (2002) Arthropods on islands: Colonization, speciation, and conservation. *Annu Rev Entomol* 47:595–632.

Gillespie RG, Claridge EM, Goodacre SL (2008) Biogeography of the fauna of French Polynesia: Diversification within and between a series of hot spot archipelagos. *Philos Trans R* Soc *Lond B Biol Sci* 363(1508):3335–3346.

Gillespie RG, Baldwin BG, Waters JM, Fraser CI, Nikula R, Roderick GK (2012) Long-distance dispersal: A framework for hypothesis testing. *Trends Ecol Evol* 27(1):47–56.

Givnish TJ, Millam KC, Mast AR, Paterson TB, Theim TJ, Hipp AL, Henss JM, Smith JF, Wood KR, Sytsma KJ (2009) Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proc Biol Sci* 276(1656):407–416.

Goebel T, Buvit I (2011) *From the Yenisei to the Yukon: Interpreting Lithic Assemblage Variability in Late Pleistocene/Early Holocene Beringia* (Texas A&M Univ Press, College Station).

Goebel T, Waters MR, O'Rourke DH (2008) The late Pleistocene dispersal of modern humans in the Americas. *Science* 319(5869):1497–1502.

Goebel T, Smith HL, DiPietro L, Waters MR, Hockett B, Graf KE, Gal R, Slobodin SB, Speakman RJ, Driese SG, Rhode D (2013) Serpentine Hot Springs, Alaska: Results of excavations and implications for the age and significance of northern fluted points. *J Archaeol Sci* 40(12):4222–4233.

Goetze E (2011) Population differentiation in the open sea: Insights from the pelagic copepod *Pleuromamma xiphias*. *Integr Comp Biol* 51(4):580–597.

Gommery D, Ramanivoso, B, Faure M, Guérin C, Kerloc'h P, Sénégas F, Randrianantenaina H (2011) Oldest evidence of human activities in Madagascar on subfossil hippopotamus bones from Anjohibe (Mahajanga Province). *C R Palevol* 10(4):271–278.

Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. Mol *Ecol* 18(6):1207–1224.

Goodman SM, Jungers WL (2014) *Extinct Madagascar: Picturing the Island's Past* (Univ of Chicago Press, Chicago).

Gottscho AD, Marks SB, Jennings WB (2014) Speciation, population structure, and demographic history of the Mojave fringe-toed lizard (*Uma scoparia*), a species of conservation concern. *Ecol Evol* 4(12):2546–2562.

Gould SJ, Johnston RF (1972) Geographic variation. *Annu Rev Ecol Syst* 3:457–498.

Gowan EJ (2013) An assessment of the minimum timing of ice free conditions of the western Laurentide Ice Sheet. *Quat Sci Rev* 75:100–113.

Graham CH, Ferrier S, Huettman F, Moritz C, Peterson AT (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trends Ecol Evol* 19(9):497–503.

Graham MR, Jaeger JR, Prendini L, Riddle BR (2013) Phylogeography of Beck's desert scorpion, *Paruroctonus becki*, reveals Pliocene diversification in the Eastern California Shear Zone and postglacial expansion in the Great Basin Desert. *Mol Phylogenet Evol* 69(3):502–513.

Grant P-R, Grant B-R (2008) How and why species multiply. *The Radiation of Darwin's Finches*, eds Grant P-R, Grant B-R (Princeton Univ Press, Princeton, NJ).

Grant WS (1987) Genetic divergence between congeneric species of Atlantic and Pacific ocean fishes. *Population Genetics and Fishery Management*, eds Ryman N, Utter F (Univ of Washington Press, Seattle), pp 225–246.

Grant WS, Leslie RW, Bowen BW (2005) Molecular genetic assessment of bipolarity in the anchovy genus *Engraulis*. *J Fish Biol* 67(5):1242–1265.

Grant WS, Zelenina DA, Mugue NS (2014) Phylogeography of red king crab: Implications for management and stock enhancement. *The King Crabs*, ed Stevens B (CRC, Boca Raton, FL), pp 47–72.

Gravel S, Steel M (2015) The existence and abundance of ghost ancestors in biparental populations. *Theor Popul Biol* 101:47–53.

Gray RD, Drummond AJ, Greenhill SJ (2009) Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323(5913):479–483.

Green GM, Sussman RW (1990) Deforestation history of the eastern rain forests of Madagascar from satellite images. *Science* 248(4952):212–215.

Green JL, Bohannan BJM, Whitaker RJ (2008) Microbial biogeography: From taxonomy to traits. *Science* 320(5879):1039–1043.

Green RE, Malaspinas AS, Krause J, Briggs AW, Johnson PL, Uhler C, Meyer M, Good JM, Maricic T, Stenzel U, Prüfer K, Siebauer M, Burbano HA, Ronan M, Rothberg JM, Egholm M, Rudan P, Brajković D, Kućan Z, Gušić I, Wikström M, Laakkonen L, Kelso J, Slatkin M, Pääbo S (2008) A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134(3):416–426.

Greene HW (2005) Organisms in nature as a central focus for biology. *Trends Ecol Evol* 20(1):23–27.

Griffiths HJ, Barnes DK, Linse K (2009) Towards a generalized biogeography of the Southern Ocean benthos. *J Biogeogr* 36(1):162–177.

Grindon AJ, Davison A (2013) Irish *Cepaea nemoralis* land snails have a cryptic Franco-Iberian origin that is most easily explained by the movements of Mesolithic humans. *PLoS One* 8(6):e65792.

Grinnell J (1914) Barriers to distribution as regards birds and mammals. *Am Nat* 48(568): 248–254.

Grivet D, Deguilloux MF, Petit RJ, Sork VL (2006) Contrasting patterns of historical colonization in white oaks (*Quercus* spp.) in California and Europe. *Mol Ecol* 15(13):4085–4093.

Grivet D, Sork VL, Westfall RD, Davis FW (2008) Conserving the evolutionary potential of California valley oak (*Quercus lobata* Née): A multivariate genetic approach to conservation planning. *Mol Ecol* 17(1):139–156.

Group GSDT (2000) Global gridded surfaces of selected soil characteristics (IGBP-DIS) (Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, TN).

Gruenstaeudl M, Reid NM, Wheeler GL, Carstens BC (2015) Posterior predictive checks of coalescent models: P2C2M, an R package. *Mol Ecol Resour* 16(1):193–205.

Gruhn R, Bryan A (2011) A current view of the initial peopling of the Americas. *Peuplements et Préhistoire en Amériques*, ed Vialou D (Éditions du Comité des Travaux Historiques et Scientific, Paris), pp 17–30.

Grummer JA, Calderón-Espinosa ML, Nieto-Montes de Oca A, Smith EN, Méndez-de la Cruz FR, Leaché AD (2015) Estimating the temporal and spatial extent of gene flow among sympatric lizard populations (genus *Sceloporus*) in the southern Mexican highlands. *Mol Ecol* 24(7):1523–1542.

Gubry-Rangin C, Hai B, Quince C, Engel M, Thomson BC, James P, Schloter M, Griffiths RI, Prosser JI, Nicol GW (2011) Niche specialization of terrestrial archaeal ammonia oxidizers. *Proc Natl Acad Sci USA* 108(52):21206–21211.

Gugger PF, Sugita S, Cavender-Bares J (2010) Phylogeography of Douglas-fir based on mitochondrial and chloroplast DNA sequences: Testing hypotheses from the fossil record. *Mol Ecol* 19(9):1877–1897.

Gugger PF, Ikegami M, Sork VL (2013) Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Née. *Mol Ecol* 22(13): 3598–3612.

Gugger PF, Cokus SJ, Pellegrini M, Sork VL (2016) Association of transcriptome-wide sequence variation with climate gradients in valley oak (*Quercus lobata*). *Tree Genet Genomes* 12(2):1–14.

Guidon NGD (1986) Carbon-14 dates point to man in the Americas 32,000 years ago. *Nature* 321(6072):769–771.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5(10):e1000695.

Guthrie RD (1990) *Frozen Fauna of the Mammoth Steppe* (Univ of Chicago Press, Chicago).

Gutierrez-Garcia TA, Vazquez-Dominguez E (2011) Comparative phylogeography: Designing studies while surviving the process. *Bioscience* 61(11):857–868.

Haffer J (1969) Speciation in Amazonian forest birds. *Science* 165(3889):131–137.

Hage P, Marck J (2003) Matrilineality and the Melanesian origin of Polynesian Y chromosomes. *Curr Anthropol* 44(S5):121–127.

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—A baiting and iterative mapping approach. *Nucleic Acids Res* 41(13):e129.

Haines WP, Schmitz P, Rubinoff D (2014) Ancient diversification of *Hyposmocoma* moths in Hawaii. *Nat Commun* 5:3502.

Hallström BM, Janke A (2010) Mammalian evolution may not be strictly bifurcating. *Mol Biol Evol* 27(12):2804–2816.

Hämälainen A, Dammhahn M, Aujard F, Eberle M, Hardy I, Kappeler PM, Perret M, Schliehe-Diecks S, Kraus C (2014) Senescence or selective disappearance? Age trajectories of body mass in wild and captive populations of a small-bodied primate. *Proc R Soc Lond Biol Sci* 281(1791):20140830.

Hamilton MB, Miller JR (2002) Comparing relative rates of pollen and seed gene flow in the island model using nuclear and organelle measures of population structure. *Genetics* 162(4):1897–1909.

Hamner RM, Pichler FB, Heimeier D, Constantine R, Baker CS (2012) Genetic differentiation and limited gene flow among fragmented populations of New Zealand endemic Hector’s and Maui’s dolphins. *Conserv Genet* 13(4):987–1002.

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philos Trans R Soc Lond B Biol Sci* 351(1345):1291–1298.

Hancock AM, Rienzo AD (2008) Detecting the genetic signature of natural selection in human populations: Models, methods, and data. *Annu Rev Anthropol* 37:197–217.

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334(6052):83–86.

Hand BK, Lowe WH, Kovach RP, Muhlfeld CC, Luikart G (2015) Landscape community genomics: Understanding eco-evolutionary processes in complex environments. *Trends Ecol Evol* 30(3):161–168.

Hanski I (1999) *Metapopulation Ecology*. *Oxford Series in Ecology and Evolution* (Oxford Univ Press, Oxford, UK).

Harcourt AH (2012) *Human Biogeography* (Univ of California Press, Berkeley).

Harcourt AH (2015) *Humankind. How Biology and Geography Shape Human Diversity* (Pegasus Books, New York).

Hardy OJ (2009) How fat is the tail? *Heredity (Edinb)* 103(6):437–438.

Hare M (2001) Prospects for nuclear gene phylogeography. *Trends Ecol Evol* 16:700–706.

Hare MP, Weinberg JR (2005) Phylogeography of surfclams, *Spisula solidissima*, in the western North Atlantic based on mitochondrial and nuclear DNA sequences. *Mar Biol* 146(4):707–716.

Hare MP, Karl SA, Avise JC (1996) Anonymous nuclear DNA markers in the American oyster and their implications for the heterozygote deficiency phenomenon in marine bivalves. *Mol Biol Evol* 13(2):334–345.

Harmon LJ, Matthews B, Des Roches S, Chase JM, Shurin JB, Schluter D (2009) Evolutionary diversification in stickleback affects ecosystem functioning. *Nature* 458(7242):1167–1170.

Harper GJ, Steininger MK, Tucker CJ, Juhn D, Hawkins F (2007) Fifty years of deforestation and forest fragmentation in Madagascar. *Environ Conserv* 34(4):325–333.

Harrison KD (2007) *When Languages Die* (Oxford Univ Press, Oxford, UK).

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. *J Hered* 105(Suppl 1):795–809.

Harvey MG, Brumfield RT (2015) Genomic variation in a widespread Neotropical bird (*Xenops minutus*) reveals divergence, population expansion, and gene flow. *Mol Phylogenet Evol* 83:305–316.

He Q, Edwards DL, Knowles LL (2013) Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution* 67(12):3386–3402.

Heaney LR (2007) Is a new paradigm emerging for oceanic island biogeography? *J Biogeogr* 34(5):753–757.

Heckman KL, Mariani CL, Rasoloarison R, Yoder AD (2007) Multiple nuclear loci reveal patterns of incomplete lineage sorting and complex species history within western mouse lemurs (*Microcebus*). *Mol Phylogenet Evol* 43(2):353–367.

Hedrick PW (2013) Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol Ecol* 22(18): 4606–4618.

Heintzman PD, Zazula GD, Cahill JA, Reyes AV, MacPhee RD, Shapiro B (2015) Genomic data from extinct North American *Camelops* revise camel evolutionary history. *Mol Biol Evol* 32(9):2433–2440.

Helgason A, Hickey E, Goodacre S, Bosnes V, Stefánsson K, Ward R, Sykes B (2001) mtDNA and the islands of the North Atlantic: Estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* 68(3):723–737.

Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487(7405):94–98.

Hellenthal G, Busby GB, Band G, Wilson JF, Capelli C, Falush D, Myers S (2014) A genetic atlas of human admixture history. *Science* 343(6172):747–751.

Hendry AP (2009) Ecological speciation! Or the lack thereof? *Can J Fish Aquat Sci* 66:1383–1398.

Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM, Kidd JM, Rodríguez-Botigué L, Ramachandran S, Hon L, Brisbin A, Lin AA, Underhill PA, Comas D, Kidd KK, Norman PJ, Parham P, Bustamante CD, Mountain JL, Feldman MW (2011) Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc Natl Acad Sci USA* 108(13):5154–5162.

Henn BM, Cavalli-Sforza LL, Feldman MW (2012) The great human expansion. *Proc Natl Acad Sci USA* 109(44):17758–17764.

Hennig W (1966) *Phylogenetic Systematics* (Univ of Illinois Press, Champaign).

Hermansen JS, Haas F, Trier CN, Bailey RI, Nederbragt AJ, Marzal A, Saetre GP (2014) Hybrid speciation through sorting of parental incompatibilities in Italian sparrows. *Mol Ecol* 23(23):5831–5842.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol J Linn Soc Lond* 58(3):247–276.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biol J Linn Soc Lond* 68(1–2): 87–112.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature* 405(6789):907–913.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography: Or seeing genes in space and time. *Mol Ecol* 10(3):537–549.

Hewitt GM (2004) The structure of biodiversity: Insights from molecular phylogeography. *Front Zool* 1(1):4.

Hewitt GM (2011) Quaternary phylogeography: The roots of hybrid zones. *Genetica* 139(5): 617–638.

Hey J (2010) Isolation with migration models for more than two populations. *Mol Biol Evol* 27(4):905–920.

Hey J, Machado CA (2003) The study of structured populations: New hope for a difficult and divided science. *Nat Rev Genet* 4(7):535–543.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci USA* 104(8): 2785–2790.

Hickerson MJ, Meyer CP (2008) Testing comparative phylogeographic models of marine vicariance and dispersal using a hierarchical Bayesian approach. *BMC Evol Biol* 8:322.

Hickerson MJ, Meyer CP, Moritz C (2006a) DNA barcoding will often fail to discover new animal species over broad parameter space. *Syst Biol* 55(5):729–739.

Hickerson MJ, Stahl EA, Lessios HA (2006b) Test for simultaneous divergence using approximate Bayesian computation. *Evolution* 60(12):2435–2453.

Hickerson MJ, Stahl E, Takebayashi N (2007) msBayes: Pipeline for testing comparative phylogeographic histories using hierarchical approximate Bayesian computation. *BMC Bioinformatics* 8(1):268.

Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. *Mol Phylogen Evol* 54(6):291–301.

Hickerson MJ, Stone GN, Lohse K, Demos TC, Xie X, Landerer C, Takebayashi N (2014) Recommendations for using msBayes to incorporate uncertainty in selecting an ABC model prior: A response to Oaks et al. *Evolution* 68(1):284–294.

Higham T, Martinón-Torres M, Cai YJ, Xing S, Tong HW, Pei SW, Sier MJ, Wu XH, Edwards RL, Cheng H, Li YY, Yang XX, de Castro JM, Wu XJ (2011) The earliest evidence for anatomically modern humans in northwestern Europe. *Nature* 479(7374):521–524.

Hijmans RJ, Graham CH (2006) The ability of climate envelope models to predict the effect of climate change on species distributions. *Glob Change Biol* 12(12):2272–2281.

Hijmans RJ, van Etten J (2012) raster: Geographic analysis and modeling with raster data. Available at http://CRAN.R-project.org/package=raster.

Hijmans RJ, Cameron SE, Parra JL, Jones G, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 1978(15):1965–1978.

Hillebrand H (2004) On the generality of the latitudinal diversity gradient. *Am Nat* 163(2):192–211.

Hirai J, Tsuda A, Goetze E (2015) Extensive genetic diversity and endemism across the global range of the oceanic copepod *Pleuromamma abdominalis*. *Prog Oceanogr* 138:77–90.

Hiscock P (2013) The human colonization of Australia. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 55–60.

Hodge JR, Read CI, van Herwerden L, Bellwood DR (2012) The role of peripheral endemism in species diversification: Evidence from the coral reef fish genus *Anampses* (Family: Labridae). *Mol Phylogenet Evol* 62(2):653–663.

Hoekstra HE, Nachman MW (2003) Different genes underlie adaptive melanism in different populations of rock pocket mice. *Mol Ecol* 12(5):1185–1194.

Hoekstra HE, Drumm KE, Nachman MW (2004) Ecological genetics of adaptive color polymorphism in pocket mice: Geographic variation in selected and neutral genes. *Evolution* 58(6):1329–1341.

Hoekstra HE, Krenz JG, Nachman MW (2005) Local adaptation in the rock pocket mouse (*Chaetodipus intermedius*): Natural selection and phylogenetic history of populations. *Heredity (Edinb)* 94(2):217–228.

Hoelzel AR (1998) Genetic structure of cetacean populations in sympatry, parapatry, and mixed assemblages: Implications for conservation policy. *J Hered* 89(5):451–458.

Hoelzel AR, Dahlheim M, Stern SJ (1998) Low genetic variation among killer whales (*Orcinus orca*) in the eastern North Pacific and genetic differentiation between foraging specialists. *J Hered* 89(2):121–128.

Holderegger R, Wagner HH (2008) Landscape genetics. *Bioscience* 58(3):199–207.

Holmes MW, Hammond TT, Wogan GO, Walsh RE, LaBarbera K, Wommack EA, Martins FM, Crawford JC, Mack KL, Bloch LM, Nachman MW (2016) Natural history collections as windows on evolutionary processes. *Mol Ecol* 25(4):864–881.

Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: Defining, estimating and interpreting F_{ST}. *Nat Rev Genet* 10(9):639–650.

Hooper D-U, Mauconduit F, Nissou MF, Nugue G, Lahrech H, Berger F, Mossuz P, Wion D (2005) Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecol Monogr* 75(1):3–35.

Hope AG, Takebayashi N, Galbreath KE, Talbot SL, Cook JA (2013) Temporal, spatial and ecological dynamics of speciation among amphi-Beringian small mammals. *J Biogeogr* 40(3):415–429.

Hormiga G, Arnedo M, Gillespie RG (2003) Speciation on a conveyor belt: Sequential colonization of the hawaiian islands by *Orsonwelles* spiders (Araneae, Linyphiidae). *Syst Biol* 52(1):70–88.

Hotaling S, Foley ME, Lawrence NM, Bocanegra J, Blanco MB, Rasoloarison R, Kappeler PM, Barrett MA, Yoder AD, Weisrock DW (2016) Species discovery and validation in a cryptic radiation of endangered primates: Coalescent-based species delimitation in Madagascar's mouse lemurs. *Mol Ecol* 25(9):2029–2045.

Hu A, Meehl GA, Otto-Bliesner BL, Waelbroeck C, Han W, Loutre M-F, Lambeck K, Mitrovica JX Rosenbloom N (2010) Influence of Bering Strait flow and North Atlantic circulation on glacial sea-level changes. *Nat Geosci* 3(2):118–121.

Hu FS, Hampe A, Petit RJ (2009) Paleoecology meets genetics: Deciphering past vegetational dynamics. *Front Ecol Environ* 7(7):371–379.

Hua X, Wiens JJ (2010) Latitudinal variation in speciation mechanisms in frogs. *Evolution* 64(2):429–443.

Hua X, Wiens JJ (2013) How does climate influence speciation? *Am Nat* 182(1):1–12.

Huang W, Takebayashi N, Qi Y, Hickerson MJ (2011) MTML-msBayes: Approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. *BMC Bioinformatics* 12(1):1.

Hudjashov G, Kivisild T, Underhill PA, Endicott P, Sanchez JJ, Lin AA, Shen P, Oefner P, Renfrew C, Villems R, Forster P (2007) Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc Natl Acad Sci USA* 104(21):8726–8730.

Hudson MJ (2013) Japan: Archaeology. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 224–229.

Hudson RR (1990) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology,* eds Futuyma DJ, Antonovics J (Oxford Univ Press, Oxford, UK), Vol 7, pp 1–44.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18(2):337–338.

Hudson RR, Coyne JA (2002) Mathematical consequences of the genealogical species concept. *Evolution* 56(8):1557–1565.

Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, Lai J, Morrell PL, Shannon LM, Song C, Springer NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J (2012) Comparative population genomics of maize domestication and improvement. *Nat Genet* 44(7):808–811.

Hugall AF, O'Hara TD, Hunjan S, Nilsen R, Moussalli A (2016) An exon-capture system for the entire class Ophiuroidea. *Mol Biol Evol* 33(1):281–294.

Hughes SS, Millard AR, Lucy SJ, Chenery CA, Evans JA, Nowelle G, Pearson DG (2014) Anglo-Saxon origins investigated by isotopic analysis of burials from Berinsfield, Oxfordshire, UK. *J Archaeol Sci* 42:81–92.

Humbert H (1955) Les territoires phytogéographiques de Madagascar. *Année Biol 3e* 31: 439–448.

Hurles ME, Sykes BC, Jobling MA, Forster P (2005) The dual origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from maternal and paternal lineages. *Am J Hum Genet* 76(5):894–901.

Hutchings L, van der Lingen CD, Shannon LJ, Crawford RJM, Verheye HMS, Bartholomae CH, van der Plas AK, Louw D, Kreiner A, Ostrowski M, Fidel Q, Barlow RG, Lamont T, Coetzee J, Shillington F, Veitch J, Currie JC, Monteiro PMS (2009) The Benguela Current: An ecosystem of four components. *Prog Oceanogr* 83(1):15–32.

Iacchei M, Gaither MR, Bowen BW, Toonen RJ (2016) Testing dispersal limits in the sea: Range-wide phylogeography of the pronghorn spiny lobster *Panulirus penicillatus. J Biogeogr* 43(5):1032–1044.

Igea J, Bogarrn D, Papadopulos AST, Savolainen V (2015) A comparative analysis of island floras challenges taxonomy-based biogeographical models of speciation. *Evolution* 69(2):482–491.

Innan H, Kim Y (2004) Pattern of polymorphism after strong artificial selection in a domestication event. *Proc Natl Acad Sci USA* 101(29):10667–10672.

IPCC (Intergovernmental Panel on Climate Change) (2015) Climate Change 2014: Mitigation of Climate Change (Cambridge Univ Press, New York), Vol 3.

Ives JW (2015) Kinship, demography and Paleoindian modes of colonization: Some western Canadian perspectives. *Mobility and Ancient Society in Asia and the Americas,* eds Frachetti MD, Spengler RN III (Springer Press, New York), pp 127–156.

Ives JW, Froese D, Supernant K, Yanicki G (2013) Vectors, vestiges and Valhallas: Rethinking the corridor. *Paleoamerican Odyssey,* eds Graf KE, Ketron CV, Waters MR (Texas A&M Univ Press, College Station), pp 149–169.

Jackson JA, Steel DJ, Beerli P, Congdon BC, Olavarría C, Leslie MS, Pomilla C, Rosenbaum H, Baker CS (2014) Global diversity and oceanic divergence of humpback whales (*Megaptera novaeangliae*). *Proc Biol Sci* 281(1786):20133222.

Jackson LE Jr, Phillips FM, Shimamura K, Little EC (1997) Cosmogenic ^{36}Cl dating of the Foothills erratics train, Alberta, Canada. *Geology* 25(3):195–198.

Jamamillo-Correa JP, Gerardi S, Beaulieu J, Ledig FT, Bousquet J (2015) Inferring and outlining past population declines with linked microsatellites: A case study in two spruce species. *Tree Genet Genomes* 11:1–12.

James PMA, Coltman DW, Murray BW, Hamelin RC, Sperling FAH (2011) Spatial genetic structure of a symbiotic beetle-fungal system: Toward multi-taxa integrated landscape genetics. *PLoS One* 6(10):e25359.

Janko K, Lecointre G, DeVries A, Couloux A, Cruaud C, Marshall C (2007) Did glacial advances during the Pleistocene influence differently the demographic histories of benthic and pelagic Antarctic shelf fishes? Inferences from intraspecific mitochondrial and nuclear DNA sequence diversity. *BMC Evol Biol* 7(1):220.

Janzen DH (1967) Why mountain passes are higher in the tropics. *Am Nat* 101(919):233–249.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SY, Faircloth BC, Nabholz B, Howard JT, Suh A, Weber CC, da Fonseca RR, Li J, Zhang F, Li H, Zhou L, Narula N, Liu L, Ganapathy G, Boussau B, Bayzid MS, Zavidovych V, Subramanian S, Gabaldón T, Capella-Gutiérrez S, Huerta-Cepas J, Rekepalli B, Munch K, Schierup M, Lindow B, Warren WC, Ray D, Green RE, Bruford MW, Zhan X, Dixon A, Li S, Li N, Huang Y, Derryberry EP, Bertelsen MF, Sheldon FH, Brumfield RT, Mello CV, Lovell PV, Wirthlin M, Schneider MP, Prosdocimi F, Samaniego JA, Vargas Velazquez AM, Alfaro-Núñez A, Campos PF, Petersen B, Sicheritz-Ponten T, Pas A, Bailey T, Scofield P, Bunce M, Lambert DM, Zhou Q, Perelman P, Driskell AC, Shapiro B, Xiong Z, Zeng Y, Liu S, Li Z, Liu B, Wu K, Xiao J, Yinqi X, Zheng Q, Zhang Y, Yang H, Wang J, Smeds L, Rheindt FE, Braun M, Fjeldsa J, Orlando L, Barker FK, Jønsson KA, Johnson W, Koepfli KP, O'Brien S, Haussler D, Ryder OA, Rahbek C, Willerslev E, Graves GR, Glenn TC, McCormack J, Burt D, Ellegren H, Alström P, Edwards SV, Stamatakis A, Mindell DP, Cracraft J, Braun EL, Warnow T, Jun W, Gilbert MT, Zhang G (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346(6215):1320–1331.

Jass CN, Burns JA, Milot PJ (2011) Description of fossil muskoxen and relative abundance of Pleistocene megafauna in central Alberta. *Can J Earth Sci* 48(5):793–800.

Jennings WB, Edwards SV (2005) Speciational history of Australian grass finches (*Poephila*) inferred from thirty gene trees. *Evolution* 59(9):2033–2047.

Jezkova T, Riddle BR, Card DC, Schield DR, Eckstut ME, Castoe TA (2015) Genetic consequences of postglacial range expansion in two codistributed rodents (genus *Dipodomys*) depend on ecology and genetic locus. *Mol Ecol* 24(1):83–97.

Jezkova T, Jaeger JR, Oláh-Hemmings V, Jones KB, Lara-Resendiz RA, Mulcahy DG, Riddle BR (2016) Range and niche shifts in response to past climate change in the desert horned lizard *Phrynosoma platyrhinos*. *Ecography* 39(5):437–448.

Johannesson K, André C (2006) Life on the margin: Genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Mol Ecol* 15(8):2013–2029.

Johnson C (1994–2013) MayflowerHistory.com. Available at www.MayflowerHistory.com. Accessed November 5, 2015.

Johnson MS, Murray J, Clarke B (2000) Parallel evolution in Marquesan partulid land snails. *Biol J Linn Soc Lond* 69(4):577–598.

Johnston WA (1933) Quaternary geology of North America in relation to the migration of man. *The American Aborigines,* ed Jenness D (Univ of Toronto Press, Toronto), pp 11–45.

Jonasson S, Michelsen A, Schmidt IK (1999) Coupling of nutrient cycling and carbon dynamics in the Arctic, integration of soil microbial and plant processes. *Appl Soil Ecol* 11(2-3):135–146.

Jones FC, Chan YF, Schmutz J, Grimwood J, Brady SD, Southwick AM, Absher DM, Myers RM, Reimchen TE, Deagle BE, Schluter D, Kingsley DM (2012) A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr Biol* 22(1):83–90.

Jones MR, Good JM (2016) Targeted capture in evolutionary and ecological genomics. *Mol Ecol* 25(1):185–202.

Jones MR, Forester BR, Teufel AI, Adams RV, Anstett DN, Goodrich BA, Landguth EL, Joost S, Manel S (2013) Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution* 67(12):3455–3468.

Jones R (1995) Tasmanian archaeology—establishing the sequences. *Annu Rev Anthropol* 24:423–446.

Joost S, Bonin A, Bruford MW, Després L, Conord C, Erhardt G, Taberlet P (2007) A spatial analysis method (SAM) to detect candidate loci for selection: Towards a landscape genomics approach to adaptation. *Mol Ecol* 16(18):3955–3969.

Jordan DS (1905) The origin of species through isolation. *Science* 22(566):545–562.

Jordan S, Simon C, Polhemus D (2003) Molecular systematics and adaptive radiation of Hawaii's endemic damselfly genus *Megalagrion* (Odonata: Coenagrionidae). *Syst Biol* 52(1):89–109.

Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, Whibley A, Becuwe M, Baxter SW, Ferguson L, Wilkinson PA, Salazar C, Davidson C, Clark R, Quail MA, Beasley H, Glithero R, Lloyd C, Sims S, Jones MC, Rogers J, Jiggins CD, Ffrench-Constant RH (2011) Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* 477(7363):203–206.

Joy DA, Conn JE (2001) Molecular and morphological phylogenetic analysis of an insular radiation in Pacific black flies (*Simulium*). *Syst Biol* 50(1):18–38.

Juan I, Emerson BC, Orom I, Hewitt GM (2000) Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. *Trends Ecol Evol* 15(3):104–109.

Kaplan NL, Hudson RR, Langley CH (1989) The "hitchhiking effect" revisited. *Genetics* 123(4):887–899.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecol Lett* 7(12):1225–1241.

Kay EA (1980) *Little Worlds of the Pacific: An Essay on Pacific Basin Biogeography (Harold L. Lyon Arboretum Lecture)* (Lyon Arboretum, Honolulu), Lecture No. 9.

Kayser M (2010) The human genetic history of Oceania: Near and remote views of dispersal. *Curr Biol* 20(4):R194–R201.

Kearns AM, Joseph L, Omland KE, Cook LG (2011) Testing the effect of transient Plio-Pleistocene barriers in monsoonal Australo-Papua: Did mangrove habitats maintain genetic connectivity in the black butcherbird? *Mol Ecol* 20(23):5042–5059.

Kearns AM, Joseph L, Toon A, Cook LG (2014) Australia's arid-adapted butcherbirds experienced range expansions during Pleistocene glacial maxima. *Nat Commun* 5:3994.

Keast A (1961) Bird speciation on the Australian continent. *Bull Mus Comp Zool* 123:303–495.

Keegan W (2013) Caribbean Islands: Archeology. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 376–383.

Keeley SC, Funk VA (2011) *The Biology of Island Floras*, eds Bramwell D, Caujapé-Castells J (Cambridge Univ Press, Cambridge, UK), pp 57–88.

Keinan A, Clark AG (2012) Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* 336(6082):740–743.

Kelly RP, Palumbi SR (2010) Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PLoS One* 5(1):e8594.

Kennedy TA, Naeem S, Howe KM, Knops JM, Tilman D, Reich P (2002) Biodiversity as a barrier to ecological invasion. *Nature* 417(6889):636–638.

Kennett JP, Keller G, Srinivasan M (1985) Miocene planktonic foraminiferal biogeography and paleoceanographic development of the Indo-Pacific region. *Geol Soc Am* 163: 197–236.

Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160(2):765–777.

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49(4):561–576.

Kingman JFC (1982) The coalescent. *Stoch Proc Applic* 13(3):235–248.

Kinoshita A, Skinner AR, Guidon N, Ignacio E, Felice GD, de Andrade Buço C, Tatumi S, Yee M, Figueiredo AM, Baffa O (2014) Dating human occupation at Toca do Serrote das Moendas, São Raimundo Nonato, Piauí-Brasil by electron spin resonance and optically stimulated luminescence. *J Hum Evol* 77:187–195.

Kirch PV, Kahn JG (2007) Advances in Polynesian prehistory: A review and assessment of the past decade (1993–2004). *J Archaeol Res* 15(3):191–238.

Kirkpatrick M (2010) How and why chromosome inversions evolve. *PLoS Biol* 8(9):e1000501.

Kisel Y, Barraclough TG (2010) Speciation has a spatial scale that depends on levels of gene flow. *Am Nat* 175(3):316–334.

Kistler L, Montenegro A, Smith BD, Gifford JA, Green RE, Newsom LA, Shapiro B (2014) Transoceanic drift and the domestication of African bottle gourds in the Americas. *Proc Natl Acad Sci USA* 111(8):2937–2941.

Kizirian D, Trager A, Donnelly MA, Wright JW (2004) Evolution of Galapagos Island lava lizards (Iguania: Tropiduridae: *Microlophus*). *Mol Phylogenet Evol* 32(3):761–769.

Klein RG (2009) *The Human Career: Human Biological and Cultural Origins* (Univ of Chicago, Chicago), 3rd Ed.

Knapp M, Horsburgh KA, Prost S, Stanton JA, Buckley HR, Walter RK, Matisoo-Smith EA (2012) Complete mitochondrial DNA genome sequences from the first New Zealanders. *Proc Natl Acad Sci USA* 109(45):18350–18354.

Knowles LL (2001) Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Mol Ecol* 10(3):691–701.

Knowles LL (2004) The burgeoning field of statistical phylogeography. *J Evol Biol* 17(1):1–10.

Knowles LL (2009a) Estimating species trees: Methods of phylogenetic analysis when there is incongruence across genes. *Syst Biol* 58(5):463–467.

Knowles LL (2009b) Statistical phylogeography. *Annu Rev Ecol Evol Syst* 40:593–612.

Knowles LL, Alvarado-Serrano DF (2010) Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: Insights from coupled ecological, demographic and genetic models in montane grasshoppers. *Mol Ecol* 19(17):3727–3745.

Knowles LL, Carstens BC (2007) Delimiting species without monophyletic gene trees. *Syst Biol* 56(6):887–895.

Knowles LL, Kubatko LS (2010) Estimating species trees: An introduction to concepts and models. *Estimating Species Trees: Practical and Theoretical Aspects,* eds Knowles LL, Kubatko LS (Wiley–Blackwell, New York), pp 1–14.

Knowles LL, Maddison WP (2002) Statistical phylogeography. *Mol Ecol* 11(12):2623–2635.

Knowles LL, Massatti R, He QX, Olson LE, Lanier HC (2016) Quantifying the similarity between genes and geography across Alaska's alpine small mammals. *J Biogeogr* 43(7):1464–1476.

Koch PL, Barnosky AD (2006) Late Quaternary extinctions: State of the debate. *Annu Rev Ecol Evol Syst* 37:215–250.

Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437(7058):543–546.

Koopman MM, Carstens BC (2010) Conservation genetic inferences in the carnivorous plant *Sarracenia alata* (Sarraceniaceae). *Conserv Genet* 11(5):2027–2038.

Koscinski D, Yates AG, Handford P, Lougheed SC (2009) Effects of landscape and history on diversification of a montane, stream-breeding amphibian. *J Biogeogr* 36(2):255–265.

Kozak KH, Wiens JJ (2007) Climatic zonation drives latitudinal variation in speciation mechanisms. *Proc Biol Sci* 274(1628):2995–3003.

Kozak KH, Graham CH, Wiens JJ (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends Ecol Evol* 23(3):141–148.

Kraft B, Strous M, Tegetmeyer HE (2011) Microbial nitrate respiration-genes, enzymes and environmental distribution. *J Biotechnol* 155(1):104–117.

Kreft H, Jetz W (2010) A framework for delineating biogeographical regions based on species distributions. *J Biogeogr* 37(11):2029–2053.

Kropf M, Kadereit JW, Comes HP (2003) Differential cycles of range contraction and expansion in European high mountain plants during the Late Quaternary: Insights from *Pritzelago alpina* (L.) O. Kuntze (Brassicaceae). *Mol Ecol* 12(4):931–949.

Kubatko LS, Gibbs HL, Bloomquist EW (2011) Inferring species-level phylogenies and taxonomic distinctiveness using multilocus data in *Sistrurus* rattlesnakes. *Syst Biol* 60(4):393–409.

Kuhner MK (2006) LAMARC 2.0: Maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22(6):768–770.

Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* 22:79–86.

Kuo C-H, Avise JC (2008) Does organismal pedigree impact the magnitude of topological congruence among gene trees for unlinked loci? *Genetica* 132(3):219–225.

Lahann P, Schmid J, Ganzhorn JU (2006) Geographic variation in populations of *Microcebus murinus* in Madagascar: Resource seasonality or Bergmann's rule? *Int J Primatol* 27(4):983–999.

Laity T, Laffan SW, González-Orozco CE, Faith DP, Rosauer DF, Byrne M, Miller JT, Crayn D, Costion C, Moritz CC, Newport K (2015) Phylodiversity to inform conservation policy: An Australian example. *Sci Total Environ* 534:131–143.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin CJ, Wang C, Zamani N, Grant BR, Grant PR, Webster MT, Andersson L (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518(7539):371–375.

Lancaster LT, Kay KM (2013) Origin and diversification of the California flora: Reexamining classic hypotheses with molecular phylogenies. *Evolution* 67(4):1041–1054.

Lande R (1980) Genetic variation and phenotypic evolution during allopatric speciation. *Am Nat* 116(4):463–479.

Lane CS, Chorn BT, Johnson TC (2013) Ash from the Toba supereruption in Lake Malawi shows no volcanic winter in East Africa at 75 ka. *Proc Natl Acad Sci USA* 110(20):8025–8029.

Langlet O (1971) Two hundred years genecology. *Taxon* 20(5-6):653–721.

Lanier HC, Knowles LL (2012) Is recombination a problem for species-tree analyses? *Syst Biol* 61(4):691–701.

Lapointe FJ, Rissler LJ (2005) Congruence, consensus, and the comparative phylogeography of codistributed species in California. *Am Nat* 166(2):290–299.

LaPolla RJ (2013) Eastern Asia: Sino-Tibetan linguistic history. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 204–208.

Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, Tresset A, Vigne JD, Edwards CJ, Schlumbaum A, Dinu A, Balaçsescu A, Dolman G, Tagliacozzo A, Manaseryan N, Miracle P, Van Wijngaarden-Bakker L, Masseti M, Bradley DG, Cooper A (2007a) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc Natl Acad Sci USA* 104(39):15276–15281.

Larson G, Cucchi T, Fujita M, Matisoo-Smith E, Robins J, Anderson A, Rolett B, Spriggs M, Dolman G, Kim TH, Thuy NT, Randi E, Doherty M, Due RA, Bollt R, Djubiantono T, Griffin B, Intoh M, Keane E, Kirch P, Li KT, Morwood M, Pedriña LM, Piper PJ, Rabett RJ, Shooter P, Van den Bergh G, West E, Wickler S, Yuan J, Cooper A, Dobney K (2007b) Phylogeny and ancient DNA of Sus provides insights into Neolithic expansion in Island Southeast Asia and Oceania. *Proc Natl Acad Sci USA* 104(12):4834–4839.

Latta RG, Mitton JB (1997) A comparison of population differentiation across four classes of gene marker in limber pine (*Pinus flexilis* James). *Genetics* 146(3):1153–1163.

Lauber CL, Hamady M, Knight R, Fierer N (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol* 75(15):5111–5120.

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, Berger B, Economou C, Bollongino R, Fu Q, Bos KI, Nordenfelt S, Li H, de Filippo C, Prüfer K, Sawyer S, Posth C, Haak W, Hallgren F, Fornander E, Rohland N, Delsate D, Francken M, Guinet JM, Wahl J, Ayodo G, Babiker HA, Bailliet G, Balanovska E, Balanovsky O, Barrantes R, Bedoya G, Ben-Ami H, Bene J, Berrada F, Bravi CM, Brisighelli F, Busby GB, Cali F, Churnosov M, Cole DE, Corach D, Damba L, van Driem G, Dryomov S, Dugoujon JM, Fedorova SA, Gallego Romero I, Gubina M, Hammer M, Henn BM, Hervig T, Hodoglugil U, Jha AR, Karachanak-Yankova S, Khusainova R, Khusnutdinova E, Kittles R, Kivisild T, Klitz W, Kučinskas V, Kushniarevich A, Laredj L, Litvinov S, Loukidis T, Mahley RW, Melegh B, Metspalu E, Molina J, Mountain J, Näkkäläjärvi K, Nesheva D, Nyambo T, Osipova L, Parik J, Platonov F, Posukh O, Romano V, Rothhammer F, Rudan I, Ruizbakiev R, Sahakyan H, Sajantila A, Salas A, Starikovskaya EB, Tarekegn A, Toncheva D, Turdikulova S, Uktveryte I, Utevska O, Vasquez R, Villena M, Voevoda M, Winkler CA, Yepiskoposyan L, Zalloua P, Zemunik T, Cooper A, Capelli C, Thomas MG, Ruiz-Linares A, Tishkoff SA, Singh L, Thangaraj K, Villems R, Comas D, Sukernik R, Metspalu M, Meyer M, Eichler EE, Burger J, Slatkin M, Pääbo S, Kelso J, Reich D, Krause J (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513(7518):409–413.

Leaché AD, Crews SC, Hickerson MJ (2007) Two waves of diversification in mammals and reptiles of Baja California revealed by hierarchical Bayesian analysis. *Biol Lett* 3(6):646–650.

Leaché AD, Koo MS, Spencer CL, Papenfuss TJ, Fisher RN, McGuire JA (2009) Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*). *Proc Natl Acad Sci USA* 106(30):12418–12423.

Leaché AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species delimitation using genome-wide SNP data. *Syst Biol* 63(4):534–542.

Ledru M, Montade V, Blanchard V, Hely C (2015) Long-term spatial changes in the distribution of the Brazilian Atlantic Forest. *Biotropica* 48:159–169.

Lee JY, Edwards SV (2008) Divergence across Australia's Carpentarian barrier: Statistical phylogeography of the red-backed fairy wren (*Malurus melanocephalus*). *Evolution* 62(12):3117–3134.

Lee SW, Ledig FT, Johnson DR (2002) Genetic variation at allozyme and RAPD markers in *Pinus longaeva* (Pinaceae) of the White Mountains, California. *Am J Bot* 89(4):566–577.

Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. *PLOS Comput Biol* 5(9):e1000520.

Lemmon AR, Lemmon EM (2012) High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. *Syst Biol* 61(5):745–761.

Lemmon EM, Lemmon AR (2013) High-throughput genomic data in systematics and phylogenetics. *Annu Rev Ecol Evol Syst* 44:99–121.

Lemmon AR, Emme SA, Lemmon EM (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst Biol* 61(5):727–744.

Lenormand T (2002) Gene flow and the limits to natural selection. *Trends Ecol Evol* 17(4): 183–189.

Leonard JA, Vilà C, Fox-Dobbs K, Koch PL, Wayne RK, Van Valkenburgh B (2007) Megafaunal extinctions and the disappearance of a specialized wolf ecomorph. *Curr Biol* 17(13):1146–1150.

Lessios HA (2008) The great American schism: Divergence of marine organims after the rise of the Central American Isthmus. *Annu Rev Ecol Syst* 39:63–91.

Lessios HA, Robertson DR (2006) Crossing the impassable: Genetic connections in 20 reef fishes across the Eastern Pacific Barrier. *Proc Biol Sci* 273(1598):2201–2208.

Lessios HA, Kessing BD, Robertson DR (1998) Massive gene flow across the world's most potent marine biogeographic barrier. *Proc Biol Sci* 265(1396):583–588.

Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* 23(1):127–128.

Leuenberger C, Wegmann D (2010) Bayesian computation and model selection without likelihoods. *Genetics* 184(1):243–252.

Levin DA (1981) Dispersal versus gene flow in plants. *Ann Mo Bot Gard* 68(2):233–253.

Lewis MP, Simons GF, Fennig CD (2016) *Ethnologue: Languages of the World* (SIL International, Dallas), 19th Ed.

Lewontin R, Birch L (1966) Hybridization as a source of variation for adaptation to new environments. *Evolution* 3:315–336.

Lewontin RC, Moore JA, Provine WB, Wallace B (2003) *Dobzhansky's Genetics of Natural Populations I-XLIII* (Columbia Univ Press, New York).

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.

Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–496.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319(5866):1100–1104.

Liang M, Nielsen R (2014) The lengths of admixture tracts. *Genetics* 197(3):953–967.

Liepelt S, Bialozyt R, Ziegenhagen B (2002) Wind-dispersed pollen mediates postglacial gene flow among refugia. *Proc Natl Acad Sci USA* 99(22):14590–14594.

Liggins L, Treml EA, Possingham HP, Riginos C (2016) Seascape features, rather than dispersal traits, predict spatial genetic patterns in co-distributed reef fishes. *J Biogeogr* 43(2):256–267.

Lindstrom S (2001) The Bering Strait connection: Dispersal and speciation in boreal macroalgae. *J Biogeogr* 28(2):243–251.

Linhart YB, Grant MC (1996) Evolutionary significance of local genetic differentiation in plants. *Annu Rev Ecol Syst* 27:237–277.

Liu L, Yu L, Kubatko L, Pearl DK, Edwards SV (2009a) Coalescent methods for estimating phylogenetic trees. *Mol Phylogenet Evol* 53(1):320–328.

Liu L, Yu L, Pearl DK, Edwards SV (2009b) Estimating species phylogenies using coalescence times among sequences. *Syst Biol* 58(5):468–477.

Liu L, Xi Z, Wu S, Davis CC, Edwards SV (2015) Estimating phylogenetic trees from genome-scale data. *Ann N Y Acad Sci* 1360:36–53.

Liu M, Lin L, Gao T, Yanagimoto T, Sakurai Y, Grant WS (2012) What maintains the central North Pacific genetic discontinuity in Pacific herring? *PLoS One* 7(12):e50340.

Liu SYV, Chang FT, Borsa P, Chen WJ, Dai CF (2014) Phylogeography of the humbug damselfish, *Dascyllus aruanus* (Linnaeus, 1758): Evidence of Indo-Pacific vicariance and genetic differentiation of peripheral populations. *Biol J Linn Soc Lond* 113(4):931–942.

Liu W, Martinón-Torres M, Cai YJ, Xing S, Tong HW, Pei SW, Sier MJ, Wu XH, Edwards RL, Cheng H, Li YY, Yang XX, de Castro JM, Wu XJ (2015) The earliest unequivocally modern humans in southern China. *Nature* 526(7575):696–699.

Loarie SR, Carter BE, Hayhoe K, McMahon S, Moe R, Knight CA, Ackerly DD (2008) Climate change and the future of California's endemic flora. *PLoS One* 3(6):e2502.

Lohse K, Frantz LAF (2014) Neandertal admixture in Eurasia confirmed by maximum-likelihood analysis of three genomes. *Genetics* 196(4):1241–1251.

Lohse K, Barton NH, Melika G, Stone GN (2012) A likelihood-based comparison of population histories in a parasitoid guild. *Mol Ecol* 21(18):4605–4617.

Lomolino MV (2004) Conservation biogeography. *Frontiers of Biogeography: New Directions in the Geography of Nature,* eds Lomolino MV, Heaney LR (Sinauer, Sunderland, MA), pp 293–296.

Lomolino MV, Riddle BR, Whittaker RJ, Brown JH (2010) *Biogeography* (Sinauer Associates, Sunderland, MA), 4th Ed.

Longhurst A (1995) Seasonal cycles of pelagic production and consumption. *Prog Oceanogr* 36(2):77–167.

Longhurst A (2007) *Ecological Geography of the Sea* (Academic Press, San Diego), 2nd Ed.

Lopez TJ, Hauselman ED, Maxson LR, Wright JW (1992) Preliminary analysis of phylogenetic relationships among Galapagos Island lizards of the genus *Tropidurus*. *Amphibia-Reptilia* 13(4):327–339.

López-Mercader N (2005) Evolutionary processes of the genus *Spermophorides* (Araneae, Pholcidae) in the Canary Islands. PhD Thesis, Universitat de Barcelona, Barcelona, Spain.

Loreau M (2010) Linking biodiversity and ecosystems: Towards a unifying ecological theory. *Philos Trans R Soc Lond B Biol Sci* 365(1537):49–60.

Loreau M, Hector A (2001) Partitioning selection and complementarity in biodiversity experiments. *Nature* 412(6842):72–76.

Loreau M, Naeem S, Inchausti P (2002) *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives* (Oxford Univ Press, Oxford, UK).

Lorenzen ED, Heller R, Siegismund HR (2012) Comparative phylogeography of African savannah ungulates. *Mol Ecol* 21(15):3656–3670.

Losos JB, Ricklefs RE (2009) Adaptation and diversification on islands. *Nature* 457(7231): 830–836.

Losos JB, Schluter D (2000) Analysis of an evolutionary species-area relationship. *Nature* 408(6814):847–850.

Losos JB, Arnold SJ, Bejerano G, Brodie ED 3rd, Hibbett D, Hoekstra HE, Mindell DP, Monteiro A, Moritz C, Orr HA, Petrov DA, Renner SS, Ricklefs RE, Soltis PS, Turner TL (2013) Evolutionary biology for the 21st century. *PLoS Biol* 11(1):e1001466.

Louis EE, Coles MS, Andriantompohavana R, Sommer JA, Engberg SE, Zaonarivelo JR, Mayor MI, Brenneman RA (2006) Revision of the mouse lemurs (*Microcebus*) of eastern Madagascar. *Int J Primatol* 27(2):347–389.

Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. *Annu Rev Ecol Syst* 15:65–95.

Lowrey TK (1995) Phylogeny, adaptive radiation, and biogeography of Hawaiian *Tetramolopium* (Asteraceae: Astereae). *Hawaiian Biogeography*, eds Wagner WL, Funk VA (Smithsonian Institution Press, Washington, DC), pp 195–220.

Lynch M (2007) *The Origins of Genome Architectures* (Sinauer, Sunderland, MA).

Ma B, Heylen K, van Spanning RJM, Frostegård Å (2007) Host range and molecular phylogenies of the soft rot enterobacterial genera *Pectobacterium* and *Dickeya. Phytopathology* 97(9):1150–1163.

Mac Nally R (2002) Multiple regression and inference in ecology and conservation biology: Further comments on identifying important predictor variables. *Biodivers Conserv* 11(8):1397–1401.

MacArthur RH, Wilson EO (1967) *The Theory of Island Biogeography* (Princeton Univ Press, Princeton, NJ).

Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z, Goodwin W, Bulbeck D, Bandelt HJ, Oppenheimer S, Torroni A, Richards M (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308(5724):1034–1036.

MacDonald GM, McLeod TK (1996) The Holocene closing of the "Ice-Free" Corridor: A biogeographical perspective. *Quat Int* 32:87–95.

MacDonald GM, Beukens RP, Kieser WE, Vitt DH (1987) Comparative radiocarbon dating of terrestrial plant macrofossils and aquatic moss from the "Ice-Free Corridor" of western Canada. *Geology* 15(9):837–840.

Mace R, Pagel M (1995) A latitudinal gradient in the density of human languages in North America. *Proc Biol Sci* 261(1360):117–121.

Macías-Hernández N, Oromi P, Arnedo MA (2008) Patterns of diversification on old volcanic islands as revealed by the woodlouse-hunter spider genus *Dysdera* (Araneae, Dysderidae) in the eastern Canary Islands. *Biol J Linn Soc Lond* 94(3):589–615.

Maddison WP (1995) Phylogenetic histories within and among species. *Experimental and Molecular Approaches to Plant Biosystematics*, eds Hoch PC, Stevenson AG (Missouri Botanical Garden Press, St. Louis), pp 273–287.

Madsen DB (2004) *Entering America: Northeast Asia and Beringia After the Last Glacial Maximum* (Univ of Utah Press, Salt Lake City).

Maffi L (2005) Linguistic, cultural and biological diversity. *Annu Rev Anthropol* 34:599–617.

Magalhaes IS, Ornelas-Garcia CP, Leal-Cardin M, Ramírez T, Barluenga M (2015) Untangling the evolutionary history of a highly polymorphic species: Introgressive hybridization and high genetic structure in the desert cichlid fish *Herichtys minckleyi*. *Mol Ecol* 24(17):4505–4520.

Magnacca KN, Price DK (2015) Rapid adaptive radiation and host plant conservation in the Hawaiian picture wing *Drosophila* (Diptera: Drosophilidae). *Mol Phylogenet Evol* 92:226–242.

Malekian M, Cooper SJB, Carthew SM (2010) Phylogeography of the Australian sugar glider (*Petaurus breviceps*): Evidence for a new divergent lineage in eastern Australia. *Aust J Zool* 58(3):165–181.

Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, Miska EA, Durbin R, Genner MJ, Turner GF (2015) Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* 350(6267):1493–1498.

Mallet J (2008) Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philos Trans R Soc Lond B Biol Sci* 363(1506):2971–2986.

Mallet J, Besansky N, Hahn MW (2016) How reticulated are species? *BioEssays* 38(2):140–149.

Mandryk CAS (1996) Late Wisconsinan deglaciation of Alberta: Processes and paleogeography. *Quat Int* 32:79–85.

Mandryk CAS, Josenhans H, Fedje DW, Mathewes RW (2001) Late Quaternary paleoenvironments of northwestern North America: Implications for inland versus coastal migration routes. *Quat Sci Rev* 20(1-3):301–314.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: Combining landscape ecology and population genetics. *Trends Ecol Evol* 18(4):189–197.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Mol Ecol* 19(17): 3824–3835.

Mania D, Heylen K, van Spanning RJM, Frostegård A (2014) The nitrate-ammonifying and *nosZ*-carrying bacterium *Bacillus vireti* is a potent source and sink for nitric and nitrous oxide under high nitrate conditions. *Environ Microbiol* 16(10):3196–3210.

Manne LL (2003) Nothing has lasted forever: Current and threatened levels of biological and cultural diversity. *Evol Ecol Res* 5:517–527.

Manor O, Borenstein E (2015) MUSiCC: A marker genes based framework for metagenomic normalization and accurate profiling of gene abundances in the microbiome. *Genome Biol* 16:53.

Marangoni A, Caramelli D, Manzi G (2014) *Homo sapiens* in the Americas. Overview of the earliest human expansion in the New World. *J Anthropol Sci* 92:79–97.

Marincovich L, Gladenkov AY (1999) Evidence for an early opening of the Bering Strait. *Nature* 397(6715):149–151.

Marjoram P, Wall JD (2006) Fast "coalescent" simulation. *BMC Genet* 7:16.

Marko PB, Hart MW (2011) The complex analytical landscape of gene flow inference. *Trends Ecol Evol* 26(9):448–456.

Marko PB, Hoffman JM, Emme SA, Mcgovern TM, Keever CC, Cox LN (2010) The "Expansion-Contraction" model of Pleistocene biogeography: Rocky shores suffer a sea change? *Mol Ecol* 19(1):146–169.

Marko PB, Eytan RI, Knowlton N (2015) Do large molecular sequence divergences imply an early closure of the Isthmus of Panama? *Proc Natl Acad Sci USA* 112(43):E5766.

Marske KA, Leschen RA, Buckley TR (2012) Concerted versus independent evolution and the search for multiple refugia: Comparative phylogeography of four forest beetles. *Evolution* 66(6):1862–1877.

Marske KA, Rahbek C, Nogues-Bravo D (2013) Phylogeography: Spanning the ecology—Evolution continuum. *Ecography* 36:1169–1181.

Martin RD (1972) A preliminary field-study of the lesser mouse lemur (*Microcebus murinus* J.F. Miller, 1777). *Z Tierpsychol* (Suppl. 9):43–89.

Martiny AC, Treseder K, Pusch G (2013) Phylogenetic conservatism of functional traits in microorganisms. *ISME J* 7(4):830–838.

Martiny JBH, Bohannan BJ, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreås L, Reysenbach AL, Smith VH, Staley JT (2006) Microbial biogeography: Putting microorganisms on the map. *Nat Rev Microbiol* 4(2):102–112.

Martiny JBH, Jones SE, Lennon JT, Martiny AC (2015) Microbiomes in light of traits: A phylogenetic perspective. *Science* 350(6261):aac9323.

Mason NA, Taylor SA (2015) Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. *Mol Ecol* 24(12):3009–3025.

Massatti R, Knowles LL (2014) Microhabitat differences impact phylogeographic concordance of codistributed species: Genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evolution* 68(10):2833–2846.

Massatti R, Knowles LL (2016) Contrasting support for alternative models of genomic variation based on microhabitat preference: Species-specific effects of climate change in alpine sedges. *Mol Ecol* 25(16):3974–3986.

Matisoo-Smith E, Robins JH (2004) Origins and dispersals of Pacific peoples: Evidence from mtDNA phylogenies of the Pacific rat. *Proc Natl Acad Sci USA* 101(24):9167–9172.

Matsen FA, Evans SN (2008) To what extent does genealogical ancestry imply genetic ancestry? *Theor Popul Biol* 74(2):182–190.

Matsumura H, Oxenham M (2013) Eastern Asia and Japan: Human biology. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 217–223.

Mátyás C (1996) Climatic adaptation of trees: Rediscovering provenance tests. *Euphytica* 92(1-2):45–54.

Mavárez J, Salazar CA, Bermingham E, Salcedo C, Jiggins CD, Linares M (2006) Speciation by hybridization in *Heliconius* butterflies. *Nature* 441(7095):868–871.

Maxson LR, Roberts JD (1984) Albumin and Australian frogs: Molecular data a challenge to speciation model. *Science* 225(4665):957–958.

Mayr E (1963) *Animal Species and Evolution* (Oxford Univ Press, London).

McConnell WJ, Kull CA (2014) Deforestation in Madagascar: Debates over the island's forest cover and challenges of measuring forest change. *Conservation and Environmental Management in Madagascar,* ed Scales IR (Earthscan/Routledge, Oxen, NY), pp 67–103.

McConvell P (2013) Australia: Linguistic history. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 327–332.

McCormack JE, Maley JM, Hird SM, Derryberry EP, Graves GR, Brumfield RT (2012) Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Mol Phylogenet Evol* 62(1):397–406.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol Phylogenet Evol* 66(2):526–538.

McCusker MR, Bentzen P (2010) Historical influences dominate the population genetic structure of a sedentary marine fish, Atlantic wolffish (*Anarhichas lupus*), across the North Atlantic Ocean. *Mol Ecol* 19(19):4228–4241.

McDougall I, Brown FH, Fleagle JG (2005) Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* 433(7027):733–736.

McGill BJ, Enquist BJ, Weiher E, Westoby M (2006) Rebuilding community ecology from functional traits. *Trends Ecol Evol* 21(4):178–185.

McLachlan JS, Clark JS, Manos PS (2005) Molecular indicators of tree migration capacity under rapid climate change. *Ecology* 86(8):2088–2098.

McNeill JR (2010) *Mosquito Empires. Ecology and War in the Greater Caribbean, 1620-1914* (Cambridge Univ Press, Cambridge, UK).

McNeill WH (1998) *Plagues and Peoples* (Anchor Press, New York).

McVean G (2009) A genealogical interpretation of principal components analysis. *PLoS Genet* 5(10):e1000686.

Mellars P (2006) Why did modern human populations disperse from Africa *ca*. 60,000 years ago? A new model. *Proc Natl Acad Sci USA* 103(25):9381–9386.

Mellars P, Gori KC, Carr M, Soares PA, Richards MB (2013) Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. *Proc Natl Acad Sci USA* 110(26):10699–10704.

Melville J, Ritchie EG, Chapple SN, Glor RE, Schulte JA 2nd (2011) Evolutionary origins and diversification of dragon lizards in Australia's tropical savannas. *Mol Phylogenet Evol* 58(2):257–270.

Mendelson TC, Shaw KL (2005) Sexual behaviour: Rapid speciation in an arthropod. *Nature* 433(7024):375–376.

Merilä J, Hendry AP (2014) Climate change, adaptation, and phenotypic plasticity: The problem and the evidence. *Evol Appl* 7(1):1–14.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA (2008) The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010:pdb.prot5448.

Mijares AS, Détroit F, Piper P, Grün R, Bellwood P, Aubert M, Champion G, Cuevas N, De Leon A, Dizon E (2010) New evidence for a 67,000-year-old human presence at Callao Cave, Luzon, Philippines. *J Hum Evol* 59(1):123–132.

Miller KG, Kominz MA, Browning JV, Wright JD, Mountain GS, Katz ME, Sugarman PJ, Cramer BS, Christie-Blick N, Pekar SF (2005) The Phanerozoic record of global sea-level change. *Science* 310(5752):1293–1298.

Milton K (1991) Comparative aspects of diet in Amazonian forest-dwellers. *Philos Trans R* Soc *Lond B Biol Sci* 334(1270):253–263.

Minin VN, Suchard MA (2008) Fast, accurate and simulation-free stochastic mapping. *Philos Trans R Soc Lond B Biol Sci* 363(1512):3985–3995.

Mittermeier RA, Gil PR, Hoffman M, Pilgrim J, Brooks T, Mittermeier CG, Lamoreux J, da Fonseca GAB (2005) *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions* (Conservation International, Washington, DC).

Miya M, Nishida M (1997) Speciation in the open ocean. *Nature* 389(6653):803–804.

Miyamoto H, Machida RJ, Nishida S (2012) Global phylogeography of the deep-sea pelagic chaetognath *Eukrohnia hamata. Prog Oceanogr* 104:99–109.

Montes I, Iriondo M, Manzano C, Arrizabalaga H, Jiménez E, Pardo MA, Goñi N, Davies CA, Estonba A (2012) Worldwide genetic structure of albacore *Thunnus alalunga* revealed by microsatellite DNA markers. *Mar Ecol Prog Ser* 471:183–191.

Moodley Y, Linz B, Yamaoka Y, Windsor HM, Breurec S, Wu JY, Maady A, Bernhöft S, Thiberge JM, Phuanukoonnon S, Jobb G, Siba P, Graham DY, Marshall BJ, Achtman M (2009) The peopling of the Pacific from a bacterial perspective. *Science* 323(5913):527–530.

Moore JL, Manne L, Brooks T, Burgess ND, Davies R, Rahbek C, Williams P, Balmford A (2002) The distribution of cultural and biological diversity in Africa. *Proc Biol Sci* 269(1501):1645–1653.

Morales A, Villalobos F, Velazco PM, Simmons NB, Piñero D (2016) Environmental niche drives genetic and morphometric structure in a widespread bat. *J Biogeogr* 43(5): 1057–1068.

Morand S (2012) Phylogeography helps with investigating the building of human parasite communities. *Parasitology* 139(14):1966–1974.

Moreau C, Bhérer C, Vézina H, Jomphe M, Labuda D, Excoffier L (2011) Deep human genealogies reveal a selective advantage to be on an expanding wave front. *Science* 334(6059):1148–1150.

Moreau CS, Hugall AF, McDonald KR, Jamieson BGM, Moritz C (2015) An ancient divide in a contiguous rainforest: Endemic earthworms in the Australian Wet Tropics. *PLoS One* 10(9):e0136943.

Moritz C (1994) Defining "evolutionary significant units" for conservation. *Trends Ecol Evol* 9(10):373–375.

Moritz C, Faith DP (1998) Comparative phylogeography and the identification of genetically divergent areas for conservation. *Mol Ecol* 7(4):419–429.

Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial-DNA—relevance for population biology and systematics. *Annu Rev Ecol Syst* 18:269–292.

Moritz C, Patton JL, Schneider CJ, Smith TB (2000) Diversification of rainforest faunas: An integrated molecular approach. *Annu Rev Ecol Evol Syst* 31(1):533–563.

Moritz C, Hoskin CJ, MacKenzie JB, Phillips BL, Tonione M, Silva N, VanDerWal J, Williams SE, Graham CH (2009) Identification and dynamics of a cryptic suture zone in tropical rainforest. *Proc Biol Sci* 276(1660):1235–1244.

Moritz C, Langham G, Kearney M, Krockenberger A, VanDerWal J, Williams S (2012) Integrating phylogeography and physiology reveals divergence of thermal traits between central and peripheral lineages of tropical rainforest lizards. *Philos Trans R Soc Lond B Biol Sci* 367(1596):1680–1687.

Moritz C, Ens E-J, Potter S, Catullo R (2013) The Australian monsoonal tropics: An opportunity to protect unique biodiversity and secure benefits for Aboriginal communities. *Pac Conserv Biol* 19(4):343–355.

Moritz C, Fujita MK, Rosauer D, Agudo R, Bourke G, Doughty P, Palmer R, Pepper M, Potter S, Pratt R, Scott M, Tonione M, Donnellan S (2016) Multilocus phylogeography reveals nested endemism in a gecko across the monsoonal tropics of Australia. *Mol Ecol* 25(6):1354–1366.

Morrone JJ (2009) *Evolutionary Biogeography: An Integrative Approach with Case Studies* (Columbia Univ Press, New York).

Morrone JJ (2015) Biogeographical regionalisation of the world: A reappraisal. *Aust Syst Bot* 28(2-3):81–90.

Mott BM, Gadau J, Anderson KE (2015) Phylogeography of *Pogonomyrmex barbatus* and *P. rugosus* harvester ants with genetic and environmental caste determination. *Ecol Evol* 5(14):2798–2826.

Moussalli A, Moritz C, Williams SE, Carnaval AC (2009) Variable responses of skinks to a common history of rainforest fluctuation: Concordance between phylogeography and palaeo-distribution models. *Mol Ecol* 18(3):483–499.

Muñoz MM, Crawford NG, McGreevy TJ Jr, Messana NJ, Tarvin RD, Revell LJ, Zandvliet RM, Hopwood JM, Mock E, Schneider AL, Schneider CJ (2013) Divergence in coloration and ecological speciation in the *Anolis marmoratus* species complex. *Mol Ecol* 22(10):2668–2682.

Murray-McIntosh RP, Scrimshaw BJ, Hatfield PJ, Penny D (1998) Testing migration patterns and estimating founding population size in Polynesia by using human mtDNA sequences. *Proc Natl Acad Sci USA* 95(15):9047–9052.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403(6772):853–858.

Myrold DD, Zeglin LH, Jansson JK (2014) The potential of metagenomic approaches for understanding soil microbial processes. *Soil Sci Soc Am J* 78(1):3–10.

Nachman MW, Payseur BA (2012) Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philos Trans R Soc Lond B Biol Sci* 367(1587):409–421.

Nachman MW, Hoekstra HE, D'Agostino SL (2003) The genetic basis of adaptive melanism in pocket mice. *Proc Natl Acad Sci USA* 100(9):5268–5273.

Naduvilezhath L, Rose LE, Metzler D (2011) Jaatha: A fast composite-likelihood approach to estimate demographic parameters. *Mol Ecol* 20(13):2709–2723.

Naeem S, Wright JP (2003) Disentangling biodiversity effects on ecosystem functioning: Deriving solutions to a seemingly insurmountable problem. *Ecol Lett* 6(6):567–579.

Nakhleh L (2013) Computational approaches to species phylogeny inference and gene tree reconciliation. *Trends Ecol Evol* 28(12):719–728.

Nater A, Burri R, Kawakami T, Smeds L, Ellegren H (2015) Resolving evolutionary relationships in closely related species with whole-genome sequencing data. *Syst Biol* 64(6):1000–1017.

Nayfach S, Pollard KS (2015) Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol* 16:51.

Neigel JE, Avise JC (1986) Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. *Evolutionary Processes and Theory*, eds Karlin S, Nevo E (Academic Press, New York), pp 515–534.

Neilson ME, Stepien CA (2009) Evolution and phylogeography of the tubenose goby genus *Proterorhinus* (Gobiidae: Teleostei): Evidence for new cryptic species. *Biol J Linn Soc Lond* 96(3):664–684.

Nelson MB, Berlemont R, Martiny AC, Martiny JBH (2015) Nitrogen cycling potential of a grassland litter microbial community. *Appl Environ Microbiol* 81(20):7012–7022.

Nepokroeff M, Sytsma KJ, Wagner WL, Zimmer EA (2003) Reconstructing ancestral patterns of colonization and dispersal in the Hawaiian understory tree genus *Psychotria* (Rubiaceae): A comparison of parsimony and likelihood approaches. *Syst Biol* 52(6):820–838.

Nettle D (1998) Explaining global patterns of language diversity. *J Anthropol Archaeol* 17(4): 354–374.

Nettle D, Romaine S (2000) *Vanishing Voices* (Oxford Univ Press, Oxford, UK).

Neuenschwander S, Hospital F, Guillaume F, Goudet J (2008) quantiNemo: An individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics* 24(13):1552–1553.

Neyman J, Pearson ES (1933) On the problem of the most efficient tests of statistical hypotheses. *Phil Trans R Soc A* 231:694–706.

Ng J, Landeen EL, Logsdon RM, Glor RE (2013) Correlation between *Anolis* lizard dewlap phenotype and environmental variation indicates adaptive divergence of a signal important to sexual selection and species recognition. *Evolution* 67(2):573–582.

Nickerson RS (1998) Confirmation bias: A ubiquitous phenomenon in many guises. *Rev Gen Psychol* 2(2):175–220.

Nicolas V, Bryja J, Akpatou B, Konecny A, Lecompte E, Colyn M, Lalis A, Couloux A, Denys C, Granjon L (2008) Comparative phylogeography of two sibling species of forest-dwelling rodent (*Praomys rostratus* and *P. tullbergi*) in West Africa: Different reactions to past forest fragmentation. *Mol Ecol* 17(23):5118–5134.

Nielsen R (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* 154(2):931–942.

Nielsen R (2005) Molecular signatures of natural selection. *Annu Rev Genet* 39:197–218.

Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Mol Ecol* 18(6): 1034–1047.

Nishikawa H, Iijima T, Kajitani R, Yamaguchi J, Ando T, Suzuki Y, Sugano S, Fujiyama A, Kosugi S, Hirakawa H, Tabata S, Ozaki K, Morimoto H, Ihara K, Obara M, Hori H, Itoh T, Fujiwara H (2015) A genetic mechanism for female-limited Batesian mimicry in *Papilio* butterfly. *Nat Genet* 47(4):405–409.

Norris RD (2000) Pelagic species diversity, biogeography, and evolution. *Paleobiology* 26(sp4): 236–258.

Norton EL, Goetze E (2013) Equatorial dispersal barriers and limited population connectivity among oceans in a planktonic copepod. *Limnol Oceanogr* 58(5):1581–1596.

Nosil P (2008) Ernst Mayr and the integration of geographic and ecological factors in speciation. *Biol J Linn Soc Lond* 95(1):24–46.

Nosil P (2012) *Ecological Speciation* (Oxford Univ Press, Oxford, UK).

Nosil P, Crespi BJ (2004) Does gene flow constrain adaptive divergence or vice versa? A test using ecomorphology and sexual isolation in *Timema cristinae* walking-sticks. *Evolution* 58(1):102–112.

Nosil P, Feder JL (2012) Genomic divergence during speciation: Causes and consequences. *Philos Trans R Soc Lond B Biol Sci* 367(1587):332–342.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: “Isolation by adaptation” and multiple roles for divergent selection. *Evolution* 62(2):316–336.

Novembre J, Di Rienzo A (2009) Spatial patterns of variation due to natural selection in humans. *Nat Rev Genet* 10(11):745–755.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD (2008) Genes mirror geography within Europe. *Nature* 456(7218):98–101.

Novick PA, Basta H, Floumanhaft M, McClure MA, Boissinot S (2009) The evolutionary dynamics of autonomous non-LTR retrotransposons in the lizard *Anolis carolinensis* shows more similarity to fish than mammals. *Mol Biol Evol* 26(8):1811–1822.

O'Connell JF, Allen J (2004) Dating the colonization of Sahul (Pleistocene Australia-New Guinea): A review of recent research. *J Archaeol Sci* 31(6):835–853.

O'Meara BC, Jackson ND, Morales-Garcia AE, Carstens BC (2015) Phylogeographic inference using approximate likelihoods. *Cold Spring Harbor Lab bioRχiv* 10.1101/025353.

Oaks JR (2014) An improved approximate-Bayesian model-choice method for estimating shared evolutionary history. *BMC Evol Biol* 14(1):150.

Oaks JR, Sukumaran J, Esselstyn JA, Linkem CW, Siler CD, Holder MT, Brown RM (2013) Evidence for climate-driven diversification? A caution for interpreting ABC inferences of simultaneous historical events. *Evolution* 67(4):991–1010.

Oaks JR, Linkem CW, Sukumaran J (2014) Implications of uniformly distributed, empirically informed priors for phylogeographical model selection: A reply to Hickerson et al. *Evolution* 68(12):3607–3617.

Oddou-Muratorio S, Petit RJ, Le Guerroue B, Guesnet D, Demesure B (2001) Pollen- versus seed-mediated gene flow in a scattered forest tree species. *Evolution* 55(6):1123–1135.

Oliver PM, Couper PJ, Pepper M (2014) Independent transitions between monsoonal and arid biomes revealed by systematic revison of a complex of Australian geckos (*Diplodactylus*; Diplodactylidae). *PLoS One* 9(12):e111895.

Olivieri GL, Sousa V, Chikhi L, Radespiel U (2008) From genetic diversity and structure to conservation: Genetic signature of recent population declines in three mouse lemur species (*Microcebus* spp.). *Biol Conserv* 141(5):1257–1271.

Olsen JL, Stam WT, Coyer JA, Reusch TB, Billingham M, Boström C, Calvert E, Christie H, Granger S, la Lumière R, Milchakova N, Oudot-Le Secq MP, Procaccini G, Sanjabi B, Serrao E, Veldsink J, Widdicombe S, Wyllie-Echeverria S (2004) North Atlantic phylogeography and large-scale population differentiation of the seagrass *Zostera marina* L. *Mol Ecol* 13(7):1923–1941.

Oppenheimer S (2003) *Out of Eden. The Peopling of the World* (Constable, London).

Oppenheimer S (2006) *The Origins of the British* (Constable & Robinson Ltd., London).

Oppenheimer S (2009) The great arc of dispersal of modern humans: Africa to Australia. *Quat Int* 202(1–2):2–13.

Oppenheimer S (2012) Out-of-Africa, the peopling of continents and islands: Tracing uniparental gene trees across the map. *Philos Trans R Soc Lond B Biol Sci* 367(1590):770–784.

Ornelas JF, Sosa V, Soltis DE, Daza JM, González C, Soltis PS, Gutiérrez-Rodríguez C, de los Monteros AE, Castoe TA, Bell C, Ruiz-Sanchez E (2013) Comparative phylogeographic analyses illustrate the complex evolutionary history of threatened cloud forests of northern Mesoamerica. *PLoS One* 8(2):e56283.

Ortego J, Gugger PF, Sork VL (2015a) Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. *J Biogeogr* 42(2):328–338.

Ortego J, Noguerales V, Gugger PF, Sork VL (2015b) Evolutionary and demographic history of the Californian scrub white oak species complex: An integrative approach. *Mol Ecol* 24(24):6188–6208.

Osborne AH, Vance D, Rohling EJ, Barton N, Rogerson M, Fello N (2008) A humid corridor across the Sahara for the migration of early modern humans out of Africa 120,000 years ago. *Proc Natl Acad Sci USA* 105(43):16444–16447.

Ottaviano GIP, Peri G (2005) Cities and cultures. *J Urban Econ* 58(2):304–337.

Otte D (1994) *The Crickets of Hawaii: Origin, Systematics and Evolution* (Orthopterists' Society, Philadelphia), p 338.

Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A (2006) Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science* 311(5768):1751–1753.

Ożgo M, Schilthuizen M (2012) Evolutionary change in *Cepaea nemoralis* shell colour over 43 years. *Glob Change Biol* 18(1):74–81.

Pagel M, Mace R (2004) The cultural wealth of nations. *Nature* 428(6980):275–278.

Palamara PF, Lencz T, Darvasi A, Pe'er I (2012) Length distributions of identity by descent reveal fine-scale demographic history. *Am J Hum Genet* 91(5):809–822.

Palamara PF, Francioli LC, Wilton PR, Genovese G, Gusev A, Finucane HK, Sankararaman S; Genome of the Netherlands Consortium, Sunyaev SR, de Bakker PI, Wakeley J, Pe'er I, Price AL (2015) Leveraging distant relatedness to quantify human mutation and gene-conversion rates. *Am J Hum Genet* 97(6):775–789.

Papadopoulou A, Knowles LL (2015a) Genomic tests of the species-pump hypothesis: Recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. *Evolution* 69(6):1501–1517.

Papadopoulou A, Knowles LL (2015b) Species-specific responses to island connectivity cycles: Refined models for testing phylogeographic concordance across a Mediterranean Pleistocene Aggregate Island Complex. *Mol Ecol* 24(16):4252–4268.

Papadopoulou A, Anastasiou I, Keskin B, Vogler AP (2009) Comparative phylogeography of tenebrionid beetles in the Aegean archipelago: The effect of dispersal ability and habitat preference. *Mol Ecol* 18(11):2503–2517.

Papadopulos AST, Baker WJ, Crayn D, Butlin RK, Kynast RG, Hutton I, Savolainen V (2011) Speciation with gene flow on Lord Howe Island. *Proc Natl Acad Sci USA* 108(32):13188–13193.

Papadopulos AST, Kaye M, Devaux C, Hipperson H, Lighten J, Dunning LT, Hutton I, Baker WJ, Butlin RK, Savolainen V (2014) Evaluation of genetic isolation within an island flora reveals unusually widespread local adaptation and supports sympatric speciation. *Philos Trans R Soc Lond B Biol Sci* 369(1648):20130342.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2):289–290.

Parchman TL, Benkman CW, Britch SC (2006) Patterns of genetic variation in the adaptive radiation of New World crossbills (Aves: *Loxia*). *Mol Ecol* 15(7):1873–1887.

Parent CE, Crespi BJ (2009) Ecological opportunity in adaptive radiation of Galápagos endemic land snails. *Am Nat* 174(6):898–905.

Parent CE, Caccone A, Petren K (2008) Colonization and diversification of Galápagos terrestrial fauna: A phylogenetic and biogeographical synthesis. *Philos Trans R Soc Lond B Biol Sci* 363(1508):3347–3361.

Park HJ, Nakhleh L (2012) Inference of reticulate evolutionary histories by maximum likelihood: The performance of information criteria. *BMC Bioinformatics* 13(Suppl 19):S12.

Parmakelis A, Rigal F, Mourikis T, Balanika K, Terzopoulou S, Rego C, Amorim IR, Crespo L, Pereira F, Triantis KA, Whittaker RJ, Borges PA (2015) Comparative phylogeography of endemic Azorean arthropods. *BMC Evol Biol* 15:250.

Pastorini J, Thalmann U, Martin RD (2003) A molecular approach to comparative phylogeography of extant Malagasy lemurs. *Proc Natl Acad Sci USA* 100(10):5879–5884.

Patarnello T, Volckaert FA, Castilho R (2007) Pillars of Hercules: Is the Atlantic-Mediterranean transition a phylogeographical break? *Mol Ecol* 16(21):4426–4444.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D (2012) Ancient admixture in human history. *Genetics* 192(3):1065–1093.

Paupério J, Herman JS, Melo-Ferreira J, Jaarola M, Alves PC, Searle JB (2012) Cryptic speciation in the field vole: A multilocus approach confirms three highly divergent lineages in Eurasia. *Mol Ecol* 21(24):6015–6032.

Payseur BA, Nachman MW (2005) The genomics of speciation: Investigating the molecular correlates of X chromosome introgression across the hybrid zone between *Mus domesticus* and *Mus musculus. Biol J Linn Soc Lond* 84(3):523–534.

Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and speciation. *Mol Ecol* 25(11):2337–2360.

Paz A, Ibáñez R, Lips KR, Crawford AJ (2015) Testing the role of ecology and life history in structuring genetic variation across a landscape: A trait-based phylogeographic approach. *Mol Ecol* 24(14):3723–3737.

Pearse DE, Miller MR, Abadía-Cardoso A, Garza JC (2014) Rapid parallel evolution of standing variation in a single, complex, genomic region is associated with life history in steelhead/rainbow trout. *Proc Biol Sci* 281(1783):20140012.

Peel AJ, Sargan DR, Baker KS, Hayman DT, Barr JA, Crameri G, Suu-Ire R, Broder CC, Lembo T, Wang LF, Fooks AR, Rossiter SJ, Wood JL, Cunningham AA (2013) Continent-wide panmixia of an African fruit bat facilitates transmission of potentially zoonotic viruses. *Nat Commun* 4:2770.

Peijnenburg KT, Goetze E (2013) High evolutionary potential of marine zooplankton. *Ecol Evol* 3(8):2765–2781.

Peijnenburg KT, Fauvelot C, Breeuwer JA, Menken SB (2006) Spatial and temporal genetic structure of the planktonic *Sagitta setosa* (Chaetognatha) in European seas as revealed by mitochondrial and nuclear DNA markers. *Mol Ecol* 15(11):3319–3338.

Pelletier TA, Carstens BC (2014) Model choice for phylogeographic inference using a large set of models. *Mol Ecol* 23(12):3028–3043.

Pelletier TA, Crisafulli C, Wagner S, Zellmer AJ, Carstens BC (2015) Historical species distribution models predict species limits in western Plethodon salamanders. *Syst Biol* 64(6):909–925.

Pellissier L, Leprieur F, Parravicini V, Cowman PF, Kulbicki M, Litsios G, Olsen SM, Wisz MS, Bellwood DR, Mouillot D (2014) Quaternary coral reef refugia preserved fish diversity. *Science* 344(6187):1016–1019.

Pemberton JM (2008) Wild pedigrees: The way forward. *Proc Biol Sci* 275(1635):613–621.

Pennings PS, Hermisson J (2006) Soft sweeps III: The signature of positive selection from recurrent mutation. *PLoS Genet* 2(12):e186.

Pennington RT, Prado DE, Pendry CA (2000) Neotropical seasonally dry forests and Quaternary vegetation changes. *J Biogeogr* 27:261–273.

Peres EA, Sobral-Souza T, Perez MF, Bonatelli IA, Silva DP, Silva MJ, Solferini VN (2015) Pleistocene niche stability and lineage diversification in the subtropical spider *Araneus omnicolor* (Araneidae). *PLoS One* 10(4):e0121543.

Perez MF, Carstens BC, Rodrigues GL, Moraes EM (2016) Anonymous nuclear markers reveal taxonomic incongruence and long-term disjunction in a cactus species complex with continental-island distribution in South America. *Mol Phylogenet Evol* 95:11–19.

Perrier de la Bâthie H (1921) La v*é*g*é*tation malgache. *Ann Mus Colonial, Marseille* 9:1–266.

Peter BM (2016) Admixture, population structure, and *F*-statistics. *Genetics* 202(4):1485–1501.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra E (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7(5):e37135.

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annu Rev Ecol Evol Syst* 37:187–214.

Petit R, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S, Vendramin GG (2003) Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science* 300(5625):1563–1565.

Petraglia M, Korisettar R, Boivin N, Clarkson C, Ditchfield P, Jones S, Koshy J, Lahr MM, Oppenheimer C, Pyle D, Roberts R, Schwenninger JL, Arnold L, White K (2007) Middle Paleolithic assemblages from the Indian subcontinent before and after the Toba super-eruption. *Science* 317(5834):114–116.

Petren K (2013) The evolution of landscape genetics. *Evolution* 67(12):3383–3385.

Pfennig DW, Wund MA, Snell-Rood EC, Cruickshank T, Schlichting CD, Moczek AP (2010) Phenotypic plasticity's impacts on diversification and speciation. *Trends Ecol Evol* 25(8): 459–467.

Pfenninger M, Patel S, Arias-Rodriguez L, Feldmeyer B, Riesch R, Plath M (2015) Unique evolutionary trajectories in repeated adaptation to hydrogen sulphide-toxic habitats of a neotropical fish (*Poecilia mexicana*). *Mol Ecol* 24(21):5446–5459.

Philippot L, Bru D, Saby NP, Cuhel J, Arrouays D, Simek M, Hallin S (2009) Spatial patterns of bacterial taxa in nature reflect ecological traits of deep branches of the 16S rRNA bacterial tree. *Environ Microbiol* 11(12):3096–3104.

Phillimore AB, Freckleton RP, Orme CDL, Owens IPF (2006) Ecology predicts large-scale patterns of phylogenetic diversification in birds. *Am Nat* 168(2):220–229.

Phillips SJ, Dudík M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. *Proceedings of the Twenty-First International Conference on Machine Learning* (Association for Computing Machinery, New York), pp 655–662.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Modell* 190(3-4):231–259.

Pinto MA, Henriques D, Chávez-Galarza J, Kryger P, Garnery L, van der Zee R, Dahle B, Soland-Reckeweg G, de la Rúa P, Dall'Olio R, Carreck NL, Johnston JS (2014) Genetic integrity of the dark European honey bee (*Apis mellifera mellifera*) from protected populations: A genome-wide assessment using SNPs and mtDNA sequence data. *J Apic Res* 53(2):269–278.

Pitulko VV, Nikolsky PA, Girya EY, Basilyan AE, Tumskoy VE, Koulakov SA, Astakhov SN, Pavlova EY, Anisimov MA (2004) The Yana RHS site: Humans in the Arctic before the Last Glacial Maximum. *Science* 303(5654):52–56.

Pitulko VV, Tikhonov AN, Pavlova EY, Nikolskiy PA, Kuper KE, Polozov RN (2016) Paleoanthropology. Early human presence in the Arctic: Evidence from 45,000-year-old mammoth remains. *Science* 351(6270):260–263.

Planas E, Ribera C (2014) Uncovering overlooked island diversity: Colonization and diversification of the medically important spider genus *Loxosceles* (Arachnida: Sicariidae) on the Canary Islands. *J Biogeogr* 41(7):1255–1266.

Pool JE, Nielsen R (2009) Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics* 181(2):711–719.

Pope KO, Terrell JE (2008) Environmental setting of human migrations in the circum-Pacific region. *J Biogeogr* 35(1):1–21.

Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14(9):817–818.

Potter S, Cooper SJ, Metcalfe CJ, Taggart DA, Eldridge MD (2012a) Phylogenetic relationships of rock-wallabies, *Petrogale* (Marsupialia: Macropodidae) and their biogeographic history within Australia. *Mol Phylogenet Evol* 62(2):640–652.

Potter S, Eldridge MD, Taggart DA, Cooper SJ (2012b) Multiple biogeographical barriers identified across the monsoon tropics of northern Australia: Phylogeographic analysis of the *brachyotis* group of rock-wallabies. *Mol Ecol* 21(9):2254–2269.

Potter S, Bragg JG, Peter BM, Bi K, Moritz C (2016) Phylogenomics at the tips: Inferring lineages and their demographic history in a tropical lizard, *Carlia amax. Mol Ecol* 25(6):1367–1380.

Poulakakis N, Glaberman S, Russello M, Beheregaray LB, Ciofi C, Powell JR, Caccone A (2008) Historical DNA analysis reveals living descendants of an extinct species of Galápagos tortoise. *Proc Natl Acad Sci USA* 105(40):15464–15469.

Prado DE, Gibbs PE (1993) Patterns of species distributions in the dry seasonal forests of South America. *Ann Miss Bot Gard* 80(4):902–927.

Prates I, Rodrigues MT, Melo-Sampaio PR, Carnaval AC (2015) Phylogenetic relationships of Amazonian anole lizards (*Dactyloa*): Taxonomic implications, new insights about phenotypic evolution and the timing of diversification. *Mol Phylogenet Evol* 82(Pt A): 258–268.

Prober SM, Leff JW, Bates ST, Borer ET, Firn J, Harpole WS, Lind EM, Seabloom EW, Adler PB, Bakker JD, Cleland EE, DeCrappeo NM, DeLorenze E, Hagenah N, Hautier Y, Hofmockel KS, Kirkman KP, Knops JM, La Pierre KJ, MacDougall AS, McCulley RL, Mitchell CE, Risch AC, Schuetz M, Stevens CJ, Williams RJ, Fierer N (2015) Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. *Ecol Lett* 18(1):85–95.

Prosser JI (2015) Dispersing misconceptions and identifying opportunities for the use of "omics" in soil microbial ecology. *Nat Rev Microbiol* 13(7):439–446.

Pruesse E, Peplies J, Glöckner FO (2012) SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28(14):1823–1829.

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, Li H, Mallick S, Dannemann M, Fu Q, Kircher M, Kuhlwilm M, Lachmann M, Meyer M, Ongyerth M, Siebauer M, Theunert C, Tandon A, Moorjani P, Pickrell J, Mullikin JC, Vohr SH, Green RE, Hellmann I, Johnson PL, Blanche H, Cann H, Kitzman JO, Shendure J, Eichler EE, Lein ES, Bakken TE, Golovanova LV, Doronichev VB, Shunkov MV, Derevianko AP, Viola B, Slatkin M, Reich D, Kelso J, Pääbo S (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43–49.

Pudlo P, Marin J-M, Estoup A, Cornuet J-M, Gautier M, Robert CP (2016) Reliable ABC model choice via random forests. *Bioinformatics* 32(6):859–866.

Pugach I, Delfin F, Gunnarsdóttir E, Kayser M, Stoneking M (2013) Genome-wide data substantiate Holocene gene flow from India to Australia. *Proc Natl Acad Sci USA* 110(5):1803–1808.

Pyron RA, Burbrink FT (2007) Hard and soft allopatry: Physically and ecologically mediated modes of geographic speciation. *J Biogeogr* 37(10):2005–2015.

Pyron RA, Wiens JJ (2013) Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. *Proc Roy Soc B* 280(1770):20131622.

Qi X, Cui C, Peng Y, Zhang X, Yang Z, Zhong H, Zhang H, Xiang K, Cao X, Wang Y, Ouzhuluobu, Basang, Ciwangsangbu, Bianba, Gonggalanzi, Wu T, Chen H, Shi H, Su B (2013) Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the Tibetan plateau. *Mol Biol Evol* 30(8):1761–1778.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* 41(Database issue, D1):D590–D596.

Quéméré E, Amelot X, Pierson J, Crouau-Roy B, Chikhi L (2012) Genetic data suggest a natural prehuman origin of open habitats in northern Madagascar and question the deforestation narrative in this region. *Proc Natl Acad Sci USA* 109(32):13028–13033.

Quinn RA, Lim YW, Maughan H, Conrad D, Rohwer F, Whiteson KL (2014) Biogeochemical forces shape the composition and physiology of polymicrobial communities in the cystic fibrosis lung. *MBio* 5(2):e00956–e13.

R Core Team (2011) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).

Rabbie JM (1992) The effects of intragroup cooperation and intergroup competition on in-group cohesion and out-group hostility. *Coalitions and Alliances in Humans and Other Animals*, eds Harcourt AH, de Waal FBM (Oxford Univ Press, Oxford, UK), pp 175–205.

Rabosky DL, Donnellan SC, Talaba AL, Lovette IJ (2007) Exceptional among-lineage variation in diversification rates during the radiation of Australia's most diverse vertebrate clade. *Proc Biol Sci* 274(1628):2915–2923.

Rademaker K, Hodgins G, Moore K, Zarrillo S, Miller C, Bromley GR, Leach P, Reid DA, Álvarez WY, Sandweiss DH (2014) Paleoindian settlement of the high-altitude Peruvian Andes. *Science* 346(6208):466–469.

Radespiel U, Ratsimbazafy JH, Rasoloharijaona S, Raveloson H, Andriaholinirina N, Rakotondravony R, Randrianarison RM, Randrianambinina B (2012) First indications of a highland specialist among mouse lemurs (*Microcebus* spp.) and evidence for a new mouse lemur species from eastern Madagascar. *Primates* 53(2):157–170.

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW Jr, Orlando L, Metspalu E, Karmin M, Tambets K, Rootsi S, Mägi R, Campos PF, Balanovska E, Balanovsky O, Khusnutdinova E, Litvinov S, Osipova LP, Fedorova SA, Voevoda MI, DeGiorgio M, Sicheritz-Ponten T, Brunak S, Demeshchenko S, Kivisild T, Villems R, Nielsen R, Jakobsson M, Willerslev E (2014) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505(7481):87–91.

Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, Albrechtsen A, Valdiosera C, Ávila-Arcos MC, Malaspinas AS, Eriksson A, Moltke I, Metspalu M, Homburger JR, Wall J, Cornejo OE, Moreno-Mayar JV, Korneliussen TS, Pierre T, Rasmussen M, Campos PF, de Barros Damgaard P, Allentoft ME, Lindo J, Metspalu E, Rodríguez-Varela R, Mansilla J, Henrickson C, Seguin-Orlando A, Malmström H, Stafford T Jr, Shringarpure SS, Moreno-Estrada A, Karmin M, Tambets K, Bergström A, Xue Y, Warmuth V, Friend AD, Singarayer J, Valdes P, Balloux F, Leboreiro I, Vera JL, Rangel-Villalobos H, Pettener D, Luiselli D, Davis LG, Heyer E, Zollikofer CP, Ponce de León MS, Smith CI, Grimes V, Pike KA, Deal M, Fuller BT, Arriaza B, Standen V, Luz MF, Ricaut F, Guidon N, Osipova L, Voevoda MI, Posukh OL, Balanovsky O, Lavryashina M, Bogunov Y, Khusnutdinova E, Gubina M, Balanovska E, Fedorova S, Litvinov S, Malyarchuk B, Derenko M, Mosher MJ, Archer D, Cybulski J, Petzelt B, Mitchell J, Worl R, Norman PJ, Parham P, Kemp BM, Kivisild T, Tyler-Smith C, Sandhu MS, Crawford M, Villems R, Smith DG, Waters MR, Goebel T, Johnson JR, Malhi RS, Jakobsson M, Meltzer DJ, Manica A, Durbin R, Bustamante CD, Song YS, Nielsen R, Willerslev E (2015) Population genetics. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349(6250):aab3884.

Rakotondranary SJ, Ganzhorn JU (2011) Habitat separation of sympatric *Microcebus* spp. in the dry spiny forest of south-eastern Madagascar. *Folia Primatol* (*Basel*) 82(4-5):212–223.

Ralston J, Kirchman JJ (2012) Continent-scale genetic structure in a boreal forest migrant, the blackpoll warbler (*Setophaga striata*). *Auk* 129(3):467–478.

Ramakrishnan U, Hadly EA (2009) Using phylochronology to reveal cryptic population histories: Review and synthesis of 29 ancient DNA studies. *Mol Ecol* 18(7):1310–1330.

Ramette A (2007) Multivariate analyses in microbial ecology. *FEMS Microbiol Ecol* 62(2): 142–160.

Ramirez KS, Craine JM, Fierer N (2012) Consistent effects of nitrogen amendments on soil microbial communities and processes across biomes. *Glob Change Biol* 18(6):1918–1927.

Randall JE (2007) *Reef and Shore Fishes of the Hawaiian Islands* (Univ of Hawaii Sea Grant College Program, Honolulu).

Randall JE, Cea A (2011) *Shore Fishes of Easter Island* (Univ of Hawaii Press, Honolulu).

Rangel TF, Diniz-Filho JAF, Bini LM (2010) SAM: A comprehensive application for spatial analysis in macroecology. *Ecography* 33(1):46–50.

Rannala B, Yang Z (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 164(4):1645–1656.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert MT, Wang Y, Raghavan M, Campos PF, Kamp HM, Wilson AS, Gledhill A, Tridico S, Bunce M, Lorenzen ED, Binladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre TL, Grønnow B, Meldgaard M, Andreasen C, Fedorova SA, Osipova LP, Higham TF, Ramsey CB, Hansen TV, Nielsen FC, Crawford MH, Brunak S, Sicheritz-Pontén T, Villems R, Nielsen R, Krogh A, Wang J, Willerslev E (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463(7282):757–762.

Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW Jr, Rasmussen S, Moltke I, Albrechtsen A, Doyle SM, Poznik GD, Gudmundsdottir V, Yadav R, Malaspinas AS, White SS 5th, Allentoft ME, Cornejo OE, Tambets K, Eriksson A, Heintzman PD, Karmin M, Korneliussen TS, Meltzer DJ, Pierre TL, Stenderup J, Saag L, Warmuth VM, Lopes MC, Malhi RS, Brunak S, Sicheritz-Ponten T, Barnes I, Collins M, Orlando L, Balloux F, Manica A, Gupta R, Metspalu M, Bustamante CD, Jakobsson M, Nielsen R, Willerslev E (2014) The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506(7487):225–229.

Rasoloarison RM, Goodman SM, Ganzhorn JU (2000) Taxonomic revision of mouse lemurs (*Microcebus*) in the western portions of Madagascar. *Int J Primatol* 21:963–1019.

Rasoloarison RM, Weisrock DW, Yoder AD, Rakotondravony D, Kappeler PM (2013) Two new species of mouse lemurs (Cheirogaleidae: *Microcebus*) from eastern Madagascar. *Int J Primatol* 34(3):455–469.

Raupach MJ, Thatje Sven, Dambach J, Rehm P, Misof B, Leese F (2010) Genetic homogeneity and circum-Antarctic distribution of two benthic shrimp species of the Southern Ocean, *Chorismus antarcticus* and *Nematocarcinus lanceopes. Mar Biol* 157(8):1783–1797.

Raven PH, Axelrod DI (1978) *Origin and Relationships of the California Flora* (Univ of California Press, Berkeley).

Raxworthy CJ, Ingram CM, Rabibisoa N, Pearson RG (2007) Applications of ecological niche modeling for species delimitation: A review and empirical evaluation using day geckos (*Phelsuma*) from Madagascar. *Syst Biol* 56(6):907–923.

Ray N, Excoffier L (2009) Inferring past demography using spatially explicit population genetic models. *Hum Biol* 81(2-3):141–157.

Ray N, Currat M, Foll M, Excoffier L (2010) SPLATCHE2: A spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics* 26(23):2993–2994.

Razafindrazaka H, Ricaut FX, Cox MP, Mormina M, Dugoujon JM, Randriamarolaza LP, Guitard E, Tonasso L, Ludes B, Crubézy E (2010) Complete mitochondrial DNA sequences provide new insights into the Polynesian motif and the peopling of Madagascar. *Eur J Hum Genet* 18(5):575–581.

Ree RH, Smith SA (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst Biol* 57(1):4–14.

Reece JS, Bowen BW, Smith DG, Larson A (2010) Molecular phylogenetics of moray eels (Muraenidae) demonstrates multiple origins of a shell-crushing jaw (*Gymnomuraena, Echidna*) and multiple colonizations of the Atlantic Ocean. *Mol Phylogenet Evol* 57(2):829–835.

Reed FA, Tishkoff SA (2006) African human diversity, origins and migrations. *Curr Opin Genet Dev* 16(6):597–605.

Reed HE, Martiny JBH (2013) Microbial composition affects the functioning of estuarine sediments. *ISME J* 7(4):868–879.

Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, Richter DJ, Lander ES, Altshuler D (2002) Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Genet* 32(1):135–142.

Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature* 461(7263):489–494.

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, García LF, Triana O, Blair S, Maestre A, Dib JC, Bravi CM, Bailliet G, Corach D, Hünemeier T, Bortolini MC, Salzano FM, Petzl-Erler ML, Acuña-Alonzo V, Aguilar-Salinas C, Canizales-Quinteros S, Tusié-Luna T, Riba L, Rodríguez-Cruz M, Lopez-Alarcón M, Coral-Vazquez R, Canto-Cetina T, Silva-Zolezzi I, Fernandez-Lopez JC, Contreras AV, Jimenez-Sanchez G, Gómez-Vázquez MJ, Molina J, Carracedo A, Salas A, Gallo C, Poletti G, Witonsky DB, Alkorta-Aranburu G, Sukernik RI, Osipova L, Fedorova SA, Vasquez R, Villena M, Moreau C, Barrantes R, Pauls D, Excoffier L, Bedoya G, Rothhammer F, Dugoujon JM, Larrouy G, Klitz W, Labuda D, Kidd J, Kidd K, Di Rienzo A, Freimer NB, Price AL, Ruiz-Linares A (2012) Reconstructing Native American population history. *Nature* 488(7411):370–374.

Reichstein M, Bahn M, Mahecha MD, Kattge J, Baldocchi DD (2014) Linking plant and ecosystem functional biogeography. *Proc Natl Acad Sci USA* 111(38):13697–13702.

Reid N, Demboski JR, Sullivan J (2012) Phylogeny estimation of the radiation of western North American chipmunks (*Tamias*) in the face of introgression using reproductive protein genes. *Syst Biol* 61(1):44–62.

Reid NM, Hird SM, Brown JM, Pelletier TA, McVay JD, Satler JD, Carstens BC (2014) Poor fit to the multispecies coalescent is widely detectable in empirical data. *Syst Biol* 63(3):322–333.

Reilly SB, Wake DB (2015) Cryptic diversity and biogeographical patterns within the black salamander (*Aneides flavipunctatus*) complex. *J Biogeogr* 42(2):280–291.

Reilly SB, Corl A, Wake DB (2015) An integrative approach to phylogeography: Investigating the effects of ancient seaways, climate, and historical geology on multilocus phylogeographic boundaries of the arboreal salamander (*Aneides lugubris*). *BMC Evol Biol* 15:241.

Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Ramsey CB, Buck CE, Cheng H, Edwards RL, Friedrich M, Grootes PM, Guilderson TP, Haflidason H, Hajdas I, Hatté C, Heaton TJ, Hoffmann DL, Hogg AG, Hughen KA, Kaiser KF, Kromer B, Manning SW, Niu M, Reimer RW, Richards DA, Scott EM, Southon JR, Staff RA, Turney CSM, van der Plicht J (2013) IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* 55(4):1869–1887.

Relethford JH (2004) Boas and beyond: Migration and craniometric variation. *Am J Hum Biol* 16(4):379–386.

Remington CL (1968) Suture-zones of hybrid interaction between recently joined biotas. *Evolutionary Biology*, eds Dobzhansky T, Hecht MK, Steere WC (Plenum Press, New York), pp 321–428.

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES 4th (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci USA* 98(20):11479–11484.

Revell LJ, Graham Reynolds R (2012) A new Bayesian method for fitting evolutionary models to comparative data with intraspecific variation. *Evolution* 66(9):2697–2707.

Reyes-Centeno H, Ghirotto S, Détroit F, Grimaud-Hervé D, Barbujani G, Harvati K (2014) Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia. *Proc Natl Acad Sci USA* 111(20):7248–7253.

Reygondeau G, Maury O, Beaugrand G, Fromentin JM, Fonteneau A, Cury, P (2012) Biogeography of tuna and billfish communities. *J Biogeogr* 39(1):114–129.

Reygondeau G, Longhurst A, Martinez E, Beaugrand G, Antoine D, Maury O (2013) Dynamic biogeochemical provinces in the global ocean. *Global Biogeochem Cycles* 27(4):1046–1058.

Rheindt FE, Edwards SV (2011) Genetic introgression: An integral but neglected component of speciation in birds. *Auk* 128(4):620–632.

Rice AM, Pfennig DW (2010) Does character displacement initiate speciation? Evidence of reduced gene flow between populations experiencing divergent selection. *J Evol Biol* 23(4):854–865.

Richards CL, Carstens BC, Knowles LL (2007) Distribution modelling and statistical phylogeography: An integrative framework for generating and testing alternative biogeographical hypotheses. *J Biogeogr* 34(11):1833–1845.

Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014) Microgeographic adaptation and the spatial scale of evolution. *Trends Ecol Evol* 29(3):165–176.

Richmond JQ, Reeder TW (2002) Evidence for parallel ecological speciation in scincid lizards of the *Eumeces skiltonianus* species group (Squamata: Scincidae). *Evolution* 56(7): 1498–1513.

Ricklefs RE (1987) Community diversity: Relative roles of local and regional processes. *Science* 235(4785):167–171.

Ricklefs RE (2004) A comprehensive framework for global patterns in biodiversity. *Ecol Lett* 7(1):1–15.

Riddle BR, Hafner DJ (2006) A step-wise approach to integrating phylogeographic and phylogenetic biogeographic perspectives on the history of a core North American warm deserts biota. *J Arid Environ* 65(3):435–461.

Riddle BR, Hafner DJ, Alexander LF, Jaeger JR (2000) Cryptic vicariance in the historical assembly of a Baja California peninsular desert biota. *Proc Natl Acad Sci USA* 97(26):14438–14443.

Riddle BR, Dawson MN, Hadly EA, Hafner DJ, Hickerson MJ, Mantooth SJ (2008) The role of molecular genetics in sculpting the future of integrative biogeography. *Prog Phys Geogr* 32(2):173–202.

Ridenhour BJ, Brodie ED Jr, Brodie ED III (2007) Patterns of genetic differentiation in *Thamnophis* and *Taricha* from the Pacific Northwest. *J Biogeogr* 34(4):724–735.

Riginos C, Buckley YM, Blomberg SP, Treml EA (2014) Dispersal capacity predicts both population genetic structure and species richness in reef fishes. *Am Nat* 184(1):52–64.

Riordan EC, Gugger PF, Ortego J, Smith C, Gaddis K, Thompson P, Sork VL (2016) Association of genetic and phenotypic variability with geography and climate in three southern California oaks. *Am J Bot* 103(1):73–85.

Rissler LJ, Apodaca JJ (2007) Adding more ecology into species delimitation: Ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Syst Biol* 56(6):924–942.

Rissler LJ, Smith WH (2010) Mapping amphibian contact zones and phylogeographical break hotspots across the United States. *Mol Ecol* 19(24):5404–5416.

Rissler LJ, Hijmans RJ, Graham CH, Moritz C, Wake DB (2006) Phylogeographic lineages and species comparisons in conservation analyses: A case study of California herpetofauna. *Am Nat* 167(5):655–666.

Rittmeyer EN, Austin CC (2015) Combined next-generation sequencing and morphology reveal fine-scale speciation in crocodile skinks (Squamata: Scincidae: *Tribolonotus*). *Mol Ecol* 24(2):466–483.

Robertshaw P (2013) Sub-Saharan Africa: Human genetics. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 115–124.

Robertson A (1961) Inbreeding in artificial selection programmes. *Genet Res* 2:189–194.

Rocca JD, Hall EK, Lennon JT, Evans SE, Waldrop MP, Cotner JB, Nemergut DR, Graham EB, Wallenstein MD (2015) Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. *ISME J* 9(8):1693–1699.

Rocha LA, Bass AL, Robertson DR, Bowen BW (2002) Adult habitat preferences, larval dispersal, and the comparative phylogeography of three Atlantic surgeonfishes (Teleostei: Acanthuridae). *Mol Ecol* 11(2):243–252.

Rocha LA, Bernal MA, Gaither MR, Alfaro ME (2013) Massively parallel DNA sequencing: The new frontier in biogeography. *Frontiers Biogeogr* 5(1):67–77.

Roderick GK, Gillespie RG (1998) Speciation and phylogeography of Hawaiian terrestrial arthropods. *Mol Ecol* 7(4):519–531.

Roderick GK, Croucher PJP, Vandergast AG, Gillespie RG (2012) Species differentiation on a dynamic landscape: Shifts in metapopulation and genetic structure using the chronology of the Hawaiian archipelago. *Evol Biol* 39(2):192–206.

Rodríguez A, Börner M, Pabijan M, Gehara M, Haddad C, Vences M (2015) Genetic divergence in tropical anurans: Deeper phylogeographic structure in forest specialists and in topographically complex regions. *Evol Ecol* 29(5):765–785.

Roe AD, Rice AV, Coltman DW, Cooke JEK, Sperling FAH (2011) Comparative phylogeography, genetic differentiation and contrasting reproductive modes in three fungal symbionts of a multipartite bark beetle symbiosis. *Mol Ecol* 20(3):584–600.

Rohde DL, Olson S, Chang JT (2004) Modelling the recent common ancestry of all living humans. *Nature* 431(7008):562–566.

Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res* 22(5):939–946.

Rohland N, Siedel H, Hofreiter M (2010) A rapid column-based ancient DNA extraction method for increased sample throughput. *Mol Ecol Resour* 10(4):677–683.

Rollins LA, Svedin N, Pryke SR, Griffith SC (2012) The role of the Ord Arid Intrusion in the historical and contemporary genetic division of long-tailed finch subspecies in northern Australia. *Ecol Evol* 2(6):1208–1219.

Rominger AJ, Goodman KR, Lim JY, Armstrong EE, Becking LE, Bennett GM, Brewer MS, Cotoras DD, Ewing CP, Harte J, Martinez ND, O'Grady PM, Percy DM, Price DK, Roderick GK, Shaw KL, Valdovinos FS, Gruner DS, Gillespie RG (2016) Community assembly on isolated islands: Macroecology meets evolution. *Glob Ecol Biogeogr* 25(7):769–780.

Rosen DE (1978) Vicariant patterns and historical explanation in biogeography. *Syst Zool* 27(2):159–188.

Rosenberg NA (2002) The probability of topological concordance of gene trees and species trees. *Theor Popul Biol* 61(2):225–247.

Rosenberg NA (2003) The shapes of neutral gene genealogies in two species: Probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution* 57(7):1465–1477.

Rosenberg NA (2007) Statistical tests for taxonomic distinctiveness from observations of monophyly. *Evolution* 61(2):317–323.

Rosenberg NA (2011) A population-genetic perspective on the similarities and differences among worldwide human populations. *Hum Biol* 83(6):659–684.

Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat Rev Genet* 3(5):380–390.

Rosenblum EB, Sarver BA, Brown JW, Des Roches S, Hardwick KM, Hether TD, Eastman JM, Pennell MW, Harmon LJ (2012) Goldilocks meets Santa Rosalia: An ephemeral speciation model explains patterns of diversification across time scales. *Evol Biol* 39(2):255–261.

Rosindell J, Phillimore AB (2011) A unified model of island biogeography sheds light on the zone of radiation. *Ecol Lett* 14(6):552–560.

Ross P, Hogg ID, Pilditch CA, Lundquist C (2009) Phylogeography of New Zealand's coastal benthos. *N Z J Mar Freshw Res* 43(5):1009–1027.

Ross-Ibarra J, Tenaillon M, Gaut BS (2009) Historical divergence and gene flow in the genus *Zea*. *Genetics* 181(4):1399–1413.

Roullier C, Benoit L, McKey DB, Lebot V (2013) Historical collections reveal patterns of diffusion of sweet potato in Oceania obscured by modern plant movements and recombination. *Proc Natl Acad Sci USA* 110(6):2205–2210.

Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, Knight R, Fierer N (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J* 4(10):1340–1351.

Rowe KC, Singhal S, Macmanes MD, Ayroles JF, Morelli TL, Rubidge EM, Bi K, Moritz CC (2011) Museum genomics: Low-cost and high-accuracy genetic data from historical specimens. *Mol Ecol Resour* 11(6):1082–1092.

RoyChoudhury A, Thompson EA (2012) Ascertainment correction for a population tree via a pruning algorithm for likelihood computation. *Theor Popul Biol* 82(1):59–65.

RoyChoudhury A, Felsenstein J, Thompson EA (2008) A two-stage pruning algorithm for likelihood computation for a population tree. *Genetics* 180(2):1095–1105.

Rubin BER, Ree RH, Moreau CS (2012) Inferring phylogenies from RAD sequence data. *PLoS One* 7(4):e33394.

Rubinoff D, Haines W-P (2005) Web-spinning caterpillar stalks snails. *Science* 309(5734):575.

Ruegg KC, Smith TB (2002) Not as the crow flies: A historical explanation for circuitous migration in Swainson's thrush (*Catharus ustulatus*). *Proc Biol Sci* 269(1498):1375–1381.

Rumeu B, Caujapé-Castells J, Blanco-Pastor JL, Jaén-Molina R, Nogales M, Elias RB, Vargas P (2011) The colonization history of *Juniperus brevifolia* (Cupressaceae) in the Azores Islands. *PLoS One* 6(11):e27697.

Rundell RJ, Holland BS, Cowie RH (2004) Molecular phylogeny and biogeography of the endemic Hawaiian Succineidae (Gastropoda: Pulmonata). *Mol Phylogenet Evol* 31(1): 246–255.

Runemark A, Hey J, Hansson B, Svensson EI (2012) Vicariance divergence and gene flow among islet populations of an endemic lizard. *Mol Ecol* 21(1):117–129.

Rutting T, Boeckx P, Muller C, Klemedtsson L (2011) Assessment of the importance of dissimilatory nitrate reduction to ammonium for the terrestrial nitrogen cycle. *Biogeosciences* 8(7):1779–1791.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D; International SNP Map Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409(6822):928–933.

Samonds KE, Godfrey LR, Ali JR, Goodman SM, Vences M, Sutherland MR, Irwin MT, Krause DW (2013) Imperfect isolation: Factors and filters shaping Madagascar's extant vertebrate fauna. *PLoS One* 8(4):e62086.

Sanford RA, Wagner DD, Wu Q, Chee-Sanford JC, Thomas SH, Cruz-García C, Rodríguez G, Massol-Deyá A, Krishnani KK, Ritalahti KM, Nissen S, Konstantinidis KT, Löffler FE (2012) Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. *Proc Natl Acad Sci USA* 109(48):19709–19714.

Sanmartín I, Van Der Mark P, Ronquist F (2008) Inferring dispersal: A Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *J Biogeogr* 35(3):428–449.

Sargsyan O, Wakeley J (2008) A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms. *Theor Popul Biol* 74(1):104–114.

Satler JD, Zellmer AJ, Carstens BC (2016) Biogeographic barriers drive co-diversification within associated eukaryotes of the *Sarracenia alata* pitcher plant system. *Peer J* 4:e1576.

Savage AE, Zamudio KR (2011) MHC genotypes associate with resistance to a frog-killing fungus. *Proc Natl Acad Sci USA* 108(40):16705–16710.

Savage AE, Zamudio KR (2016) Adaptive tolerance to a pathogenic fungus drives major histocompatibility complex evolution in natural amphibian populations. *Proc Biol Sci* 283(1827):20153115.

Savage AE, Becker CG, Zamudio KR (2015) Linking genetic and environmental factors in amphibian disease risk. *Evol Appl* 8(6):560–572.

Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nat Rev Genet* 14(11):807–820.

Scally A, Durbin R (2012) Revising the human mutation rate: Implications for understanding human evolution. *Nat Rev Genet* 13(10):745–753.

Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, McCarthy S, Montgomery SH, Schwalie PC, Tang YA, Ward MC, Xue Y, Yngvadottir B, Alkan C, Andersen LN, Ayub Q, Ball EV, Beal K, Bradley BJ, Chen Y, Clee CM, Fitzgerald S, Graves TA, Gu Y, Heath P, Heger A, Karakoc E, Kolb-Kokocinski A, Laird GK, Lunter G, Meader S, Mort M, Mullikin JC, Munch K, O'Connor TD, Phillips AD, Prado-Martinez J, Rogers AS, Sajjadian S, Schmidt D, Shaw K, Simpson JT, Stenson PD, Turner DJ, Vigilant L, Vilella AJ, Whitener W, Zhu B, Cooper DN, de Jong P, Dermitzakis ET, Eichler EE, Flicek P, Goldman N, Mundy NI, Ning Z, Odom DT, Ponting CP, Quail MA, Ryder OA, Searle SM, Warren WC, Wilson RK, Schierup MH, Rogers J, Tyler-Smith C, Durbin R (2012) Insights into hominid evolution from the gorilla genome sequence. *Nature* 483(7388):169–175.

Schäffler L, Saborowski J, Kappeler PM (2015) Agent-mediated spatial storage effect in heterogeneous habitat stabilizes competitive mouse lemur coexistence in Menabe Central, western Madagascar. *BMC Ecol* 15:7.

Scheiner SM, Willig MR (2011) *The Theory of Ecology* (Univ of Chicago Press, Chicago).

Schemske DW, Bierzychudek P (2001) Perspective: Evolution of flower color in the desert annual *Linanthus parryae:* Wright revisited. *Evolution* 55(7):1269–1282.

Schemske DW, Bierzychudek P (2007) Spatial differentiation for flower color in the desert annual *Linanthus parryae*: Was Wright right? *Evolution* 61(11):2528–2543.

Schield DR, Card DC, Adams RH, Jezkova T, Reyes-Velasco J, Proctor FN, Spencer CL, Herrmann HW, Mackessy SP, Castoe TA (2015) Incipient speciation with biased gene flow between two lineages of the western diamondback rattlesnake (*Crotalus atrox*). *Mol Phylogenet Evol* 83:213–223.

Schierenbeck KA (2014) *Phylogeography of California: An Introduction* (Univ of California Press, Oakland, CA).

Schiffels S, Durbin R (2014) Inferring human population size and separation history from multiple genome sequences. *Nat Genet* 46(8):919–925.

Schimel JP, Schaeffer SM (2012) Microbial control over carbon cycling in soil. *Front Microbiol* 3:348.

Schluter D (2001) Ecology and the origin of species. *Trends Ecol Evol* 16(7, S1): 372–380.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science* 323(5915): 737–741.

Schluter D (2016) Speciation, ecological opportunity, and latitude (American Society of Naturalists Address). *Am Nat* 187(1):1–18.

Schmid J, Kappeler PM (1994) Sympatric mouse lemurs (*Microcebus* spp.) in western Madagascar. *Folia Primatol* (*Basel*) 63(3):162–170.

Schmitz P, Cibois A, Landry B (2007) Molecular phylogeny and dating of an insular endemic moth radiation inferred from mitochondrial and nuclear genes: The genus *Galagete* (Lepidoptera: Autostichidae) of the Galapagos Islands. *Mol Phylogenet Evol* 45(1):180–192.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: Complexity, diversity, and dynamics. *Science* 326(5956):1112–1115.

Schneider CJ, Cunningham M, Moritz C (1998) Comparative phylogeography and the history of endemic vertebrates in the Wet Tropics rainforests of Australia. *Mol Ecol* 7(4): 487–498.

Schneider N, Chikhi L, Currat M, Radespiel U (2010) Signals of recent spatial expansions in the grey mouse lemur (*Microcebus murinus*). *BMC Evol Biol* 10:105.

Scholz CA, Cohen AS, Johnson TC, King J, Talbot MR, Brown ET (2011) Scientific drilling in the Great Rift Valley: The 2005 Lake Malawi Scientific Drilling Project—An overview of the past 145,000 years of climate variability in Southern Hemisphere East Africa. *Palaeogeogr Palaeoclimatol Palaeoecol* 303(1-4):3–19.

Schoville SD, Bonin A, François O, Lobreaux S, Melodelima C, Manel S (2012) Adaptive genetic variation on the landscape: Methods and cases. *Annu Rev Ecol Evol Syst* 43:23–43.

Schroeder KB, Schurr TG, Long JC, Rosenberg NA, Crawford MH, Tarskaia LA, Osipova LP, Zhadanov SI, Smith DG (2007) A private allele ubiquitous in the Americas. *Biol Lett* 3(2):218–223.

Schultz JK, Feldheim KA, Gruber SH, Ashley MV, McGovern TM, Bowen BW (2008) Global phylogeography and seascape genetics of the lemon sharks (genus *Negaprion*). *Mol Ecol* 17(24):5336–5348.

Schurr TG (2004) The peopling of the New World: Perspectives from molecular anthropology. *Annu Rev Anthropol* 33(1):551–583.

Searle JB, Jones CS, Gündüz I, Scascitelli M, Jones EP, Herman JS, Rambau RV, Noble LR, Berry RJ, Giménez MD, Jóhannesdóttir F (2009) Of mice and (Viking?) men: Phylogeography of British and Irish house mice. *Proc Biol Sci* 276(1655):201–207.

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G-P, Bank C, Brannstrom A, Brelsford A, Clarkson CS, Eroukhmanoff F, Feder JL, Fischer MC, Foote AD, Franchini P, Jiggins CD, Jones FC, Lindholm AK, Lucek K, Maan ME, Marques DA, Martin SH, Matthews B, Meier JI, Most M, Nachman MW, Nonaka E, Rennison DJ, Schwarzer J, Watson ET, Westram AM, Widmer A (2014) Genomics and the origin of species. *Nat Rev Genet* 15(3):176–192.

Selkoe KA, Gaggiotti OE, Treml EA, Wren JLK, Donovan MK, Hawai'i Reef Connectivity Consortium, Toonen RJ (2016) The DNA of coral reef biodiversity: Predicting and protecting genetic diversity of reef assemblages. *Proc R Soc B* 283:20160354.

Sequeira AS, Lanteri AA, Albelo LR, Bhattacharya S, Sijapati M (2008) Colonization history, ecological shifts and diversification in the evolution of endemic Galápagos weevils. *Mol Ecol* 17(4):1089–1107.

Serva M, Petroni F, Volchenkov D, Wichmann S (2012) Malagasy dialects and the peopling of Madagascar. *J R Soc Interface* 9(66):54–67.

Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution* 68(1):1–15.

Shafer ABA, Wolf JBW (2013) Widespread evidence for incipient ecological speciation: A meta-analysis of isolation-by-ecology. *Ecol Lett* 16(7):940–950.

Shafer AB, Cullingham CI, Côté SD, Coltman DW (2010) Of glaciers and refugia: A decade of study sheds new light on the phylogeography of northwestern North America. *Mol Ecol* 19(21):4589–4621.

Shapiro B, Hofreiter M (2014) A paleogenomic perspective on evolution and gene function: New insights from ancient DNA. *Science* 343(6169):1236573.

Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Gilbert MT, Barnes I, Binladen J, Willerslev E, Hansen AJ, Baryshnikov GF, Burns JA, Davydov S, Driver JC, Froese DG, Harington CR, Keddie G, Kosintsev P, Kunz ML, Martin LD, Stephenson RO, Storer J, Tedford R, Zimov S, Cooper A (2004) Rise and fall of the Beringian steppe bison. *Science* 306(5701):1561–1565.

Shapiro LH, Strazanac JS, Roderick GK (2006) Molecular phylogeny of *Banza* (Orthoptera: Tettigoniidae), the endemic katydids of the Hawaiian archipelago. *Mol Phylogenet Evol* 41(1):53–63.

Shaw KL (2001) The genealogical view of speciation. Commentary on C-I. Wu, The genic view of the process of speciation. *J Evol Biol* 14(6):880–882.

Shaw KL (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proc Natl Acad Sci USA* 99(25):16122–16127.

Shcheglovitova M, Anderson RP (2013) Estimating optimal complexity for ecological niche models: A jackknife approach for species with small sample sizes. *Ecol Modell* 269:9–17.

Shi H, Zhong H, Peng Y, Dong YL, Qi XB, Zhang F, Liu LF, Tan SJ, Ma RZ, Xiao CJ, Wells RS, Jin L, Su B (2008) Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol* 6:45.

Silva G, Horne JB, Castilho R (2014) Anchovies go north and west without losing diversity: Post-glacial range expansions in a small pelagic fish. *J Biogeogr* 41(6):1171–1182.

Simpson GG (1940) Mammals and land bridges. *J Wash Acad Sci* 30:137–163.

Simpson GG (1950) History of the fauna of Latin America. *Am Sci* 38(3):361–389.

Sinclair SJ, White MD, Newell GR (2010) How useful are species distribution models for managing biodiversity under future climates. *Ecol Soc* 15(8):1–13.

Sinervo B, Svensson E, Comendant T (2000) Density cycles and an offspring quantity and quality game driven by natural selection. *Nature* 406(6799):985–988.

Singhal S, Moritz C (2013) Reproductive isolation between phylogeographic lineages scales with divergence. *Proc Biol Sci* 280(1772):20132246.

Sites JW Jr, Marshall JC (2004) Operational criteria for delimiting species. *Annu Rev Ecol Evol Syst* 35:199–227.

Sjödin P, Kaj I, Krone S, Lascoux M, Nordborg M (2005) On the meaning and existence of an effective population size. *Genetics* 169(2):1061–1070.

Skoglund P, Malmström H, Omrak A, Raghavan M, Valdiosera C, Günther T, Hall P, Tambets K, Parik J, Sjögren KG, Apel J, Willerslev E, Storå J, Götherström A, Jakobsson M (2014) Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* 344(6185):747–750.

Skyrms B, Avise JC, Ayala FJ (2014) In the light of evolution VIII: Darwinian thinking in the social sciences. *Proc Natl Acad Sci USA* 111(Suppl 3):10781–10896.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* 236(4803):787–792.

Slatkin M (2008) Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* 9(6):477–485.

Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT (2014a) Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Syst Biol* 63(1):83–95.

Smith BT, McCormack JE, Cuervo AM, Hickerson MJ, Aleixo A, Cadena CD, Pérez-Emán J, Burney CW, Xie X, Harvey MG, Faircloth BC, Glenn TC, Derryberry EP, Prejean J, Fields S, Brumfield RT (2014b) The drivers of tropical speciation. *Nature* 515(7527):406–409.

Smith CI, Tank S, Godsoe W, Levenick J, Strand E, Esque T, Pellmyr O (2011) Comparative phylogeography of a coevolved community: Concerted population expansions in Joshua trees and four yucca moths. *PLoS One* 6(10):e25628.

Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genet Res* 23(1):23–35.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity (Edinb)* 82(5):561–573.

Sobel JM, Chen GF, Watt LR, Schemske DW (2010) The biology of speciation. *Evolution* 64(2):295–315.

Solís-Lemus C, Knowles LL, Ané C (2015) Bayesian species delimitation combining multiple genes and traits in a unified framework. *Evolution* 69(2):492–507.

Soltis DE, Gitzendanner MA, Strenge DD, Solits PA (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Syst Evol* 206(1-4):353–373.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Mol Ecol* 15(14):4261–4293.

Song S, Liu L, Edwards SV, Wu S (2012) Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc Natl Acad Sci USA* 109(37):14942–14947.

Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, Buerkle CA, Feder JL, Bast J, Schwander T, Egan SP, Crespi BJ, Nosil P (2014) Stick insect genomes reveal natural selection's role in parallel speciation. *Science* 344(6185):738–742.

Sork VL, Smouse PE (2006) Genetic analysis of landscape connectivity in tree populations. *Landsc Ecol* 21(6):821–836.

Sork VL, Waits L (2010) Contributions of landscape genetics: Approaches, insights, and future potential. *Mol Ecol* 19(17):3489–3495.

Sork VL, Werth S (2014) Phylogeography of *Ramalina menziesii,* a widely distributed lichen-forming fungus in western North America. *Mol Ecol* 23(9):2326–2339.

Sork VL, Davis FW, Westfall R, Flint A, Ikegami M, Wang H, Grivet D (2010) Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. *Mol Ecol* 19(17):3806–3823.

Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB (2013) Putting the landscape into the genomics of trees: Approaches for understanding local adaptation and population responses to changing climate. *Tree Genet Genomes* 9(4):901–911.

Sork VL, Smouse PE, Grivet D, Scofield DG (2015) Impact of asymmetric male and female gamete dispersal on allelic diversity and spatial genetic structure in valley oak (*Quercus lobata* Née). *Evol Ecol* 29(6):927–945.

Sork VL, Squire K, Gugger PF, Steele SE, Levy ED, Eckert AJ (2016) Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata*. *Am J Bot* 103(1):33–46.

Sousa VC, Grelaud A, Hey J (2011) On the nonidentifiability of migration time estimates in isolation with migration models. *Mol Ecol* 20(19):3956–3962.

Souza RC, Hungria M, Cantão ME, Ribeiro Vasconcelos AT, Nogueira MA, Vicente VA (2015) Metagenomic analysis reveals microbial functional redundancies and specificities in a soil under different tillage and crop-management regimes. *Appl Soil Ecol* 86:106–112.

Spriggs M (2013) Oceania: Lapita migration. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 308–313.

Sproul JS, Houston DD, Nelson CR, Evans RP, Crandall KA, and Shiozawa DK (2015) Climate oscillations, glacial refugia, and dispersal ability: Factors influencing the genetic structure of the least salmonfly, *Pteronarcella badia* (Plecoptera), in Western North America. *BMC Evol Biol* 15(1):279.

Stadler T, Degnan JH (2012) A polynomial time algorithm for calculating the probability of a ranked gene tree given a species tree. *Algorithms Mol Biol* 7(1):7.

Stalker AM (1977) The probable extent of classical Wisconsin ice in southern and central Alberta. *Can J Earth Sci* 14(11):2614–2619.

Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.

Stanton ML, Galen C, Shore J (1997) Population structure along a steep environmental gradient: Consequences of flowering time and habitat variation in the snow buttercup, *Ranunculus adoneus*. *Evolution* 51(1):79–94.

Stebbins GL, Major J (1965) Endemism and speciation in the California flora. *Ecol Monogr* 35(1):2–35.

Steinfartz S, Glaberman S, Lanterbecq D, Marquez C, Rassmann K, Caccone A (2007) Genetic impact of a severe El Niño event on Galápagos marine iguanas (*Amblyrhynchus cristatus*). *PLoS One* 2(12):e1285.

Stenz NWM, Larget B, Baum DA, Ané C (2015) Exploring tree-like and non-tree-like patterns using genome sequences: An example using the inbreeding plant species *Arabidopsis thaliana* (L.) Heynh. *Syst Biol* 64(5):809–823.

Steppan SJ, Zawadzki C, Heaney L (2003) Molecular phylogeny of the endemic Philippine rodent *Apomys* (Muridae) and the dynamics of diversification in an oceanic archipelago. *Biol J Linn Soc Lond* 80(4):699–715.

Stern DL (2013) The genetic causes of convergent evolution. *Nat Rev Genet* 14(11):751–764.

Stewart JR, Lister AM, Barnes I, Dalen L (2010) Refugia revisited: Individualistic responses of species in space and time. *Proc Biol Sci* 277(1682):661–671.

Stone AC, Griffiths RC, Zegura SL, Hammer MF (2002) High levels of Y-chromosome nucleotide diversity in the genus *Pan. Proc Natl Acad Sci USA* 99(1):43–48.

Stone GN, Lohse K, Nicholls JA, Fuentes-Utrilla P, Sinclair F, Schönrogge K, Csóka G, Melika G, Nieves-Aldrey JL, Pujade-Villar J, Tavakoli M, Askew RR, Hickerson MJ (2012) Reconstructing community assembly in time and space reveals enemy escape in a Western Palearctic insect community. *Curr Biol* 22(6):532–537.

Stoneking M, Harvati K (2013) Early Old World migrations of *Homo sapiens*: Human biology. *The Global Prehistory of Human Migration*, ed Bellwood P (Wiley-Blackwell, Chichester, UK), pp 26–37.

Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP (2007) Putting the “landscape” in landscape genetics. *Heredity (Edinb)* 98(3):128–142.

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: Where are we now? *Mol Ecol* 19(17):3496–3514.

Strassmann JE, Queller DC, Avise JC, Ayala FJ (2011) In the light of evolution V: Cooperation and conflict. *Proc Natl Acad Sci USA* 108(Suppl 2):10787–10926.

Strauss SY, Webb CO, Salamin N (2006) Exotic taxa less related to native species are more invasive. *Proc Natl Acad Sci USA* 103(15):5841–5845.

Streicher JW, Devitt TJ, Goldberg CS, Malone JH, Blackmon H, Fujita MK (2014) Diversification and asymmetrical gene flow across time and space: Lineage sorting and hybridization in polytypic barking frogs. *Mol Ecol* 23(13):3273–3291.

Strickland MS, Lauber C, Fierer N, Bradford MA (2009) Testing the functional significance of microbial community composition. *Ecology* 90(2):441–451.

Striedter GF, Avise JC, Ayala FJ (2012) In the light of evolution VI: Brain and behavior. *Proc Natl Acad Sci USA* 109(Suppl 1):10607–10741.

Strous M, Fuerst JA, Kramer EH, Logemann S, Muyzer G, van de Pas-Schoonen KT, Webb R, Kuenen JG, Jetten MS. (1999) Missing lithotroph identified as new planctomycete. *Nature* 400(6743):446–449.

Sukumaran J, Holder MT (2010) DendroPy: A Python library for phylogenetic computing. *Bioinformatics* 26(12):1569–1571.

Sun W, Cupples AM (2012) Diversity of five anaerobic toluene-degrading microbial communities investigated using stable isotope probing. *Appl Environ Microbiol* 78(4):972–980.

Surovell TA (2000) Early Paleoindian women, children, mobility, and fertility. *Am Antiq* 65(3):493–508.

Swaegers J, Mergeay J, Therry L, Bonte D, Larmuseau MH, Stoks R (2014) Unravelling the effects of contemporary and historical range expansion on the distribution of genetic diversity in the damselfly *Coenagrion scitulum*. *J Evol Biol* 27(4):748–759.

Swaegers J, Mergeay J, Van Geystelen A, Therry L, Larmuseau MH, Stoks R (2015) Neutral and adaptive genomic signatures of rapid poleward range expansion. *Mol Ecol* 24(24):6163–6176.

Swenson NG (2010) Mapping the suturing of a continental biota. *Mol Ecol* 19(24):5324–5327.

Swenson NG, Howard DJ (2004) Do suture zones exist? *Evolution* 58(11):2391–2397.

Swenson NG, Howard DJ (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *Am Nat* 166(5):581–591.

Syring J, Farrell K, Businský R, Cronn R, Liston A (2007) Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus. Syst Biol* 56(2):163–181.

Szabó Z, Snelgrove B, Craig M, Rocha L, Bowen B (2014) Phylogeography of the Manybar Goatfish, *Parupeneus multifasciatus* reveals moderate structure between the Central and North Pacific and a cryptic endemic species in the Marquesas. *Bull Mar Sci* 90(1): 493–512.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Mol Ecol* 7(4):453–464.

Tackney JC, Potter BA, Raff J, Powers M, Watkins WS, Warner D, Reuther JD, Irish JD, O'Rourke DH (2015) Two contemporaneous mitogenomes from terminal Pleistocene burials in eastern Beringia. *Proc Natl Acad Sci USA* 112(45):13833–13838.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105(2):437–460.

Takahata N, Nei M (1985) Gene genealogy and variance of interpopulational nucleotide differences. *Genetics* 110(2):325–344.

Takahata N, Slatkin M (1990) Genealogy of neutral genes in two partially isolated populations. *Theor Popul Biol* 38(3):331–350.

Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, Fedorova SA, Golubenko MV, Stepanov VA, Gubina MA, Zhadanov SI, Ossipova LP, Damba L, Voevoda MI, Dipierri JE, Villems R, Malhi RS (2007) Beringian standstill and spread of Native American founders. *PLoS One* 2(9):e829.

Tanabe K, Mita T, Jombart T, Eriksson A, Horibe S, Palacpac N, Ranford-Cartwright L, Sawai H, Sakihama N, Ohmae H, Nakamura M, Ferreira MU, Escalante AA, Prugnolle F, Björkman A, Färnert A, Kaneko A, Horii T, Manica A, Kishino H, Balloux F (2010) *Plasmodium falciparum* accompanied the human expansion out of Africa. *Curr Biol* 20(14):1283–1289.

Tavaré S (1984) Line-of-descent and genealogical processes, and their applications in population genetics models. *Theor Popul Biol* 26(2):119–164.

Tavares C, Roques A, Courtial B, Branco M, Jactel H, Lopez-Vaamonde C (2015) Phylogeography of the ladybird *Iberorhyzobius rondensis*, a potential biological control agent of the invasive alien pine bast scale *Matsucoccus feytaudi*. *BioControl* 60(1):59–69.

Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, Villarreal Ruiz L, Vasco-Palacios AM, Thu PQ, Suija A, Smith ME, Sharp C, Saluveer E, Saitta A, Rosas M, Riit T, Ratkowsky D, Pritsch K, Põldmaa K, Piepenbring M, Phosri C, Peterson M, Parts K, Pärtel K, Otsing E, Nouhra E, Njouonkou AL, Nilsson RH, Morgado LN, Mayor J, May TW, Majuakim L, Lodge DJ, Lee SS, Larsson KH, Kohout P, Hosaka K, Hiiesalu I, Henkel TW, Harend H, Guo LD, Greslebin A, Grelet G, Geml J, Gates G, Dunstan W, Dunk C, Drenkhan R, Dearnaley J, De Kesel A, Dang T, Chen X, Buegger F, Brearley FQ, Bonito G, Anslan S, Abell S, Abarenkov K (2014) Fungal biogeography. Global diversity and geography of soil fungi. *Science* 346(6213):1256688.

Templeton AR (2008) Nested clade analysis: An extensively validated method for strong phylogeographic inference. *Mol Ecol* 17(8):1877–1880.

Templeton AR (2009) Statistical hypothesis testing in intraspecific phylogeography: Nested clade phylogeographical analysis vs. approximate Bayesian computation. *Mol Ecol* 18(2):319–331.

Templeton AR (2010) Coalescent-based, maximum likelihood inference in phylogeography. *Mol Ecol* 19(3):431–435, discussion 436–446.

Tenggardjaja KA, Bowen BW, Bernardi G (2014) Vertical and horizontal genetic connectivity in *Chromis verater*, an endemic damselfish found on shallow and mesophotic reefs in the Hawaiian archipelago and adjacent Johnston Atoll. *PLoS One* 9(12):e115493.

Tenggardjaja KA, Bowen BW, Bernardi G (2016) Reef fish dispersal in the Hawaiian archipelago: Comparative phylogeography of three endemic damselfishes. *J Mar Biol* 2016:3251814.

Teske PR, Von Der Heyden S, McQuaid CD, Barker NP (2011) A review of marine phylogeography in southern Africa. *S Afr J Sci* 107(5-6):43–53.

Thalmann U, Rakotoarison N (1994) Distribution of lemurs in central western Madagascar, with a regional distribution hypothesis. *Folia Primatol* (*Basel*) 63(3):156–161.

The 1000 Genome Project Consortium (2015) A global reference for human genetic variation. *Nature* 526:68–74.

Theisen TC, Bowen BW, Lanier W, Baldwin JD (2008) High connectivity on a global scale in the pelagic wahoo, *Acanthocybium solandri* (tuna family Scombridae). *Mol Ecol* 17(19): 4233–4247.

Thomas T, Gilbert J, Meyer F (2012) Metagenomics—a guide from sampling to data analysis. *Microb Inform Exp* 2(1):3.

Thomé MCT, Sequeira F, Brusquetti F, Carstens B, Haddad CFB, Rodrigues MT, Alexandrino J (2016) Recurrent connections between Amazon and Atlantic forests shaped diversity in Caatinga four-eyed frogs. *J Biogeogr* 43(5):1045–1056.

Thompson EA (1975) *Human Evolutionary Trees* (Cambridge Univ Press, Cambridge, UK).

Thornhill DJ, Mahon AR, Norenburg JL, Halanych KM (2008) Open-ocean barriers to dispersal: A test case with the Antarctic Polar Front and the ribbon worm *Parborlasia corrugatus* (Nemertea: Lineidae). *Mol Ecol* 17(23):5104–5117.

Tibayrenc M, Avise JC, Ayala FJ (2015) In the light of evolution IX: Clonal reproduction: Alternatives to sex. *Proc Natl Acad Sci USA* 112:8824–8936.

Timm J, Figiel M, Kochzius M (2008) Contrasting patterns in species boundaries and evolution of anemonefishes (Amphiprioninae, Pomacentridae) in the centre of marine biodiversity. *Mol Phylogenet Evol* 49(1):268–276.

Ting CT, Tsaur SC, Wu CI (2000) The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus. Proc Natl Acad Sci USA* 97(10):5313–5316.

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, Ibrahim M, Omar SA, Lema G, Nyambo TB, Ghori J, Bumpstead S, Pritchard JK, Wray GA, Deloukas P (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* 39(1):31–40.

Tollis M, Ausubel G, Ghimire D, Boissinot S (2012) Multi-locus phylogeographic and population genetic analysis of *Anolis carolinensis*: Historical demography of a genomic model species. *PLoS One* 7(6):e38474.

Toon A, Hughes JM, Joseph L (2010) Multilocus analysis of honeyeaters (Aves: Meliphagidae) highlights spatio-temporal heterogeneity in the influence of biogeographic barriers in the Australian monsoonal zone. *Mol Ecol* 19(14):2980–2994.

Toonen RJ, Andrews KR, Baums IB, Bird CE, Concepcion GT, Daly-Engel TS, Eble JA, Faucci A, Gaither MR, Iacchei M, Puritz JB, Schultz JK, Skillings DJ, Timmers MA, Bowen BW (2011) Defining boundaries for ecosystem-based management: A multispecies case study of marine connectivity across the Hawaiian archipelago. *J Mar Biol* 2011:460173.

Toonen RJ, Wilhelm T, Maxwell SM, Wagner D, Bowen BW, Sheppard CR, Taei SM, Teroroko T, Moffitt R, Gaymer CF, Morgan L, Lewis N, Sheppard AL, Parks J, Friedlander AM, Big Ocean Think Tank (2013) One size does not fit all: The emerging frontier in large-scale marine conservation. *Mar Pollut Bull* 77(1-2):7–10.

Tornabene L, Valdez S, Erdmann M, Pezold F (2015) Support for a ‘Center of Origin’ in the Coral Triangle: Cryptic diversity, recent speciation, and local endemism in a diverse lineage of reef fishes (Gobiidae: *Eviota*). *Mol Phylogenet Evol* 82(Pt A):200–210.

Torres-Florez JP, Hucke-Gaete R, LeDuc R, Lang A, Taylor B, Pimper LE, Bedriñana-Romano L, Rosenbaum HC, Figueroa CC (2014) Blue whale population structure along the eastern South Pacific Ocean: Evidence of more than one population. *Mol Ecol* 23(24): 5998–6010.

Toth I, Humphris S, Campbell E, Pritchard L (2015) Why genomics research on *Pectobacterium* and *Dickeya* makes a difference. *Am J Potato Res* 92(2):218–222.

Tribsch A, Schonswetter P (2003) Patterns of endemism and comparative phylogeography confirm palaeoenvironmental evidence for Pleistocene refugia in the Eastern Alps. *Taxon* 52(3):477–497.

Tsai Y-HE, Carstens BC (2013) Assessing model fit in phylogeographic investigations: An example from the North American willow *Salix melanopsis*. *J Biogeogr* 40:131–141.

Tsiknia M, Paranychianakis NV, Varouchakis EA, Nikolaidis NP (2015) Environmental drivers of the distribution of nitrogen functional genes at a watershed scale. *FEMS Microbiol Ecol* 91(6):fiv052.

Tumonggor MK, Karafet TM, Hallmark B, Lansing JS, Sudoyo H, Hammer MF, Cox MP (2013) The Indonesian archipelago: An ancient genetic highway linking Asia and the Pacific. *J Hum Genet* 58(3):165–173.

Uchimura A, Higuchi M, Minakuchi Y, Ohno M, Toyoda A, Fujiyama A, Miura I, Wakana S, Nishino J, Yagi T (2015) Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Res* 25(8):1125–1134.

Usher KM, Bergman B, Raven JA (2007) Exploring cyanobacterial mutualisms. *Annu Rev Ecol Evol Syst* 38:255–273.

Vajda E (2010) A Siberian link with Na-Dene languages. *Anthropol Pap Univ Alaska New Series* 5(1):33–99.

van der Heijden MGA, Bardgett RD, van Straalen NM (2008) The unseen majority: Soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol Lett* 11(3):296–310.

van der Linden P, Mitchell JE (2009) *ENSEMBLES: Climate Change and Its Impacts: Summary of Research and Results from the ENSEMBLES Project* (Met Office Hadley Centre, Exeter, UK).

van Holst Pellekaan S (2013) Genetic evidence for the colonization of Australia. *Quat Int* 285(1):44–56.

van Kessel MA, Speth DR, Albertsen M, Nielsen PH, Op den Camp HJ, Kartal B, Jetten MS, Lücker S (2015) Complete nitrification by a single microorganism. *Nature* 528(7583): 555–559.

Van Riel P, Jordaens K, Van Houtte N, Martins AM, Verhagen R, Backeljau T (2005) Molecular systematics of the endemic *Leptaxini* (Gastropoda: Pulmonata) on the Azores islands. *Mol Phylogenet Evol* 37(1):132–143.

Vandergast AG, Bohonak AJ, Weissman DB, Fisher RN (2007) Understanding the genetic effects of recent habitat fragmentation in the context of evolutionary history: Phylogeography and landscape genetics of a southern California endemic Jerusalem cricket (Orthoptera: Stenopelmatidae: *Stenopelmatus*). *Mol Ecol* 16(5):977–992.

VanderWerf EA, Young LC, Yeung NW, Carlon DB (2010) Stepping stone speciation in Hawaii's flycatchers: Molecular divergence supports new island endemics within the elepaio. *Conserv Genet* 11(4):1283–1298.

Vargas SM, Jensen MP, Ho SY, Mobaraki A, Broderick D, Mortimer JA, Whiting SD, Miller J, Prince RI, Bell IP, Hoenner X, Limpus CJ, Santos FR, FitzSimmons NN (2016) Phylogeography, genetic diversity, and management units of hawksbill turtles in the Indo-Pacific. *J Hered* 107(3):199–213.

Varin T, Lovejoy C, Jungblut AD, Vincent WF, Corbeil J (2010) Metagenomic profiling of Arctic microbial mat communities as nutrient scavenging and recycling systems. *Limnol Oceanogr* 55(5):1901–1911.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol Ecol* 13(4):921–935.

Vellend M (2010) Conceptual synthesis in community ecology. *Q Rev Biol* 85(2):183–206.

Vences M, Wollenberg KC, Vieites DR, Lees DC (2009) Madagascar as a model region of species diversification. *Trends Ecol Evol* 24(8):456–465.

Vera-Escalona I, Habit E, Ruzzante DE (2015) Echoes of a distant time: Effects of historical processes on contemporary genetic patterns in *Galaxias platei* in Patagonia. *Mol Ecol* 24(16):4112–4128.

Vermeij GJ (1991) Anatomy of an invasion: The trans-Arctic interchange. *Paleobiology* 17(3): 281–307.

Veron JEN (2000) *Corals of the World* (Australian Institute of Marine Science, Townsville, Australia), Vols 1–3.

Vickery RK (1992) Pollinator preferences for yellow, orange, and red flowers of *Mimulus verbenaceus* and *M. cardinalis*. *Great Basin Nat* 52(2):145–148.

Vinas J, Alvarado Bremer J, Pla C (2004) Inter-oceanic genetic differentiation among albacore (*Thunnus alalunga*) populations. *Mar Biol* 145(2):225–232.

Vincent B, Dionne M, Kent MP, Lien S, Bernatchez L (2013) Landscape genomics in Atlantic salmon (*Salmo salar*): Searching for gene-environment interactions driving local adaptation. *Evolution* 67(12):3469–3487.

Violle C, Reich PB, Pacala SW, Enquist BJ, Kattge J (2014) The emergence and promise of functional biogeography. *Proc Natl Acad Sci USA* 111(38):13690–13696.

Vitousek PM (1984) Litterfall, nutrient cycling, and nutrient limitation in tropical forests. *Ecology* 65(1):285–298.

Vitousek PM, Aber JD, Howarth RW, Likens GE, Matson PA, Schindler DW, Schlesinger WH, Tilman DG (1997) Human alteration of the global nitrogen cycle: Sources and consequences. *Ecol Appl* 7(3):737–750.

Vorontsova MS, Besnard G, Forest F, Malakasi P, Moat J, Clayton WD, Ficinski P, Savva GM, Nanjarisoa OP, Razanatsoa J, Randriatsara FO, Kimeu JM, Luke WR, Kayombo C, Linder HP (2016) Madagascar's grasses and grasslands: Anthropogenic or natural? *Proc Biol Sci* 283(1823):20152262.

Wagner CE, Keller I, Wittwer S, Selz OM, Mwaiko S, Greuter L, Sivasundar A, Seehausen O (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Mol Ecol* 22(3):787–798.

Wakeley J (2000) The effects of subdivision on the genetic divergence of populations and species. *Evolution* 54(4):1092–1101.

Wakeley J, King L, Low BS, Ramachandran S (2012) Gene genealogies within a fixed pedigree, and the robustness of Kingman's coalescent. *Genetics* 190(4):1433–1445.

Waldrop E, Hobbs J-PA, Randall JE, DiBattista JD, Rocha LA, Kosaki RK, Berumen ML, Bowen BW (2016) Phylogeography, population structure and evolution of coral-eating butterflyfishes (Family Chaetodontidae, genus *Chaetodon*, subgenus *Corallochaetodon*). *J Biogeogr* 43(6):1116–1129.

Wallace AR (1852) On the monkeys of the Amazon. *Proc Zool Soc Lond* 20(2):107–110.

Wallace AR (1876) *The Geographical Distribution of Animals* (Macmillan, London).

Wallace AR (1880) *Island Life, or the Phenomena and Causes of Insular Faunas and Floras* (Macmillan, London).

Wallace BP, DiMatteo AD, Hurley BJ, Finkbeiner EM, Bolten AB, Chaloupka MY, Hutchinson BJ, Abreu-Grobois FA, Amorocho D, Bjorndal KA, Bourjea J, Bowen BW, Dueñas RB, Casale P, Choudhury BC, Costa A, Dutton PH, Fallabrino A, Girard A, Girondot M, Godfrey MH, Hamann M, López-Mendilaharsu M, Marcovaldi MA, Mortimer JA, Musick JA, Nel R, Pilcher NJ, Seminoff JA, Troëng S, Witherington B, Mast RB (2010) Regional management units for marine turtles: A novel framework for prioritizing conservation and research across multiple scales. *PLoS One* 5(12):e15465.

Wallberg A, Han F, Wellhagen G, Dahle B, Kawata M, Haddad N, Simões ZL, Allsopp MH, Kandemir I, De la Rúa P, Pirk CW, Webster MT (2014) A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera. Nat Genet* 46(10):1081–1088.

Waltari E, Hijmans RJ, Peterson AT, Nyári AS, Perkins SL, Guralnick RP (2007) Locating Pleistocene refugia: Comparing phylogeographic and ecological niche model predictions. *PLoS One* 2(6):e563.

Waltari E, Schroeder R, McDonald K, Anderson RP, Carnaval A (2014) Bioclimatic variables derived from remote sensing: Assessment and application for species distribution modelling. *Methods Ecol Evol* 5(10):1033–1042.

Wang IJ (2010) Recognizing the temporal distinctions between landscape genetics and phylogeography. *Mol Ecol* 19(13):2605–2608.

Wang IJ (2011) Inversely related aposematic traits: Reduced conspicuousness evolves with increased toxicity in a polymorphic poison-dart frog. *Evolution* 65(6):1637–1649.

Wang IJ (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution* 67(12):3403–3411.

Wang IJ, Bradburd GS (2014) Isolation by environment. *Mol Ecol* 23(23):5649–5662.

Wang IJ, Summers K (2010) Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Mol Ecol* 19(3):447–458.

Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecol Lett* 16(2):175–182.

Wang RL, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize domestication. *Nature* 398(6724):236–239.

Wang Y, Zhou Y, Li L, Chen X, Liu Y, Ma ZM, Xu S (2014) A new method for modeling coalescent processes with recombination. *BMC Bioinformatics* 15:273.

Wares JP, Cunningham CW (2001) Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* 55(12):2455–2469.

Warren BH, Simberloff D, Ricklefs RE, Aguilée R, Condamine FL, Gravel D, Morlon H, Mouquet N, Rosindell J, Casquet J, Conti E, Cornuault J, Fernández-Palacios JM, Hengl T, Norder SJ, Rijsdijk KF, Sanmartín I, Strasberg D, Triantis KA, Valente LM, Whittaker RJ, Gillespie RG, Emerson BC, Thébaud C (2015) Islands as model systems in ecology and evolution: Prospects fifty years after MacArthur-Wilson. *Ecol Lett* 18(2):200–217.

Warren DL (2012) In defense of "niche modeling." *Trends Ecol Evol* 27(9):497–500.

Warwick AR, Travis J, Lemmon EM (2015) Geographic variation in the Pine Barrens treefrog (*Hyla andersonii*): Concordance of genetic, morphometric and acoustic signal data. *Mol Ecol* 24(13):3281–3298.

Waters JM, Fraser CI, Hewitt GM (2013) Founder takes all: Density-dependent processes structure biodiversity. *Trends Ecol Evol* 28(2):78–85.

Waters MR, Stafford TW Jr (2013) The first Americans: A review of the evidence for the Late-Pleistocene peopling of the Americas. *Paleoamerican Odyssey,* eds Graf KE, Ketron CV, Waters MR (Texas A&M Univ Press, College Station), pp 543–562.

Waters MR, Forman SL, Jennings TA, Nordt LC, Driese SG, Feinberg JM, Keene JL, Halligan J, Lindquist A, Pierson J, Hallmark CT, Collins MB, Wiederhold JE (2011) The Buttermilk Creek complex and the origins of Clovis at the Debra L. Friedkin site, Texas. *Science* 331(6024):1599–1603.

Waters MR, Stafford TW Jr, Kooyman B, Hills LV (2015) Late Pleistocene horse and camel hunting at the southern margin of the ice-free corridor: Reassessing the age of Wally's Beach, Canada. *Proc Natl Acad Sci USA* 112(14):4263–4267.

Watterson GA (1970) On the equivalence of random mating and random union of gametes models in finite, monoecious populations. *Theor Popul Biol* 1(2):233–250.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7(2):256–276.

Wegmann D, Currat M, Excoffier L (2006) Molecular diversity after a range expansion in heterogeneous environments. *Genetics* 174(4):2009–2020.

Wegmann D, Leuenberger C, Excoffier L (2009) Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* 182(4):1207–1218.

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* 11(1):116.

Weisrock DW, Rasoloarison RM, Fiorentino I, Ralison JM, Goodman SM, Kappeler PM, Yoder AD (2010) Delimiting species without nuclear monophyly in Madagascar's mouse lemurs. *PLoS One* 5(3):e9883.

Weisrock DW, Smith SD, Chan LM, Biebouw K, Kappeler PM, Yoder AD (2012) Concatenation and concordance in the reconstruction of mouse lemur phylogeny: An empirical demonstration of the effect of allele sampling in phylogenetics. *Mol Biol Evol* 29(6):1615–1630.

Welsh A, Chee-Sanford JC, Connor LM, Löffler FE, Sanford RA (2014) Refined NrfA phylogeny improves PCR-based *nrfA* gene detection. *Appl Environ Microbiol* 80(7):2110–2119.

Wen D, Yu Y, Hahn MW, Nakhleh L (2016) Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Mol Ecol* 25(11):2361–2372.

Werneck FP, Costa GC, Colli GR, Prado DE, Sites JW (2011) Revisiting the historical distribution of seasonally dry tropical forests: New insights based on palaeodistribution modelling and palynological evidence. *Glob Ecol Biogeogr* 20(2):272–288.

Werth S, Sork VL (2008) Local genetic structure in a North American epiphytic lichen, *Ramalina menziesii* (Ramalinaceae). *Am J Bot* 95(5):568–576.

Werth S, Sork VL (2010) Identity and genetic structure of the photobiont of the epiphytic lichen *Ramalina menziesii* on three oak species in southern California. *Am J Bot* 97(5):821–830.

Werth S, Sork VL (2014) Ecological specialization in *Trebouxia* (Trebouxiophyceae) photobionts of *Ramalina menziesii* (Ramalinaceae) across six range-covering ecoregions of western North America. *Am J Bot* 101(7):1127–1140.

Weyenberg G, Huggins PM, Schardl CL, Howe DK, Yoshida R (2014) kdetrees: Nonparametric estimation of phylogenetic tree distributions. *Bioinformatics* 30(16):2280–2287.

White WM, McBirney AR, Duncan RA (1993) Petrology and geochemistry of the Galapagos Islands: Portrait of a pathological mantle plume. *J Geophys Res Solid Earth* 98(B11): 19533–19563.

Whiteman NK, Kimball RT, Parker PG (2007) Co-phylogeography and comparative population genetics of the threatened Galápagos hawk and three ectoparasite species: Ecology shapes population histories within parasite communities. *Mol Ecol* 16(22):4759–4773.

Whitman WB, Coleman DC, Wiebe WJ (1998) Prokaryotes: The unseen majority. *Proc Natl Acad Sci USA* 95(12):6578–6583.

Whittaker RJ, Araújo MB, Jepson P, Ladle RJ, Watson JEM, Willis KJ (2005) Conservation biogeography: Assessment and prospect. *Divers Distrib* 11(1):3–23.

Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York).

Wickler S, Spriggs M (1988) Pleistocene human occupation of the Solomon-Islands, Melanesia. *Antiquity* 62(237):703–706.

Wiens JJ (2004) What is speciation and how should we study it? *Am Nat* 163(6):914–923.

Wiens JJ, Donoghue MJ (2004) Historical biogeography, ecology and species richness. *Trends Ecol Evol* 19(12):639–644.

Wiens JJ, Sukumaran J, Pyron RA, Brown RM (2009) Evolutionary and biogeographic origins of high tropical diversity in old world frogs (Ranidae). *Evolution* 63(5):1217–1231.

Wiens JJ, Ackerly DD, Allen AP, Anacker BL, Buckley LB, Cornell HV, Damschen EI, Jonathan Davies T, Grytnes JA, Harrison SP, Hawkins BA, Holt RD, McCain CM, Stephens PR (2010) Niche conservatism as an emerging principle in ecology and conservation biology. *Ecol Lett* 13(10):1310–1324.

Wilmé L, Goodman SM, Ganzhorn JU (2006) Biogeographic evolution of Madagascar's microendemic biota. *Science* 312(5776):1063–1065.

Wilmshurst JM, Anderson AJ, Higham TFG, Worthy TH (2008) Dating the late prehistoric dispersal of Polynesians to New Zealand using the commensal Pacific rat. *Proc Natl Acad Sci USA* 105(22):7676–7680.

Wilmshurst JM, Hunt TL, Lipo CP, Anderson AJ (2011) High-precision radiocarbon dating shows recent and rapid initial human colonization of East Polynesia. *Proc Natl Acad Sci USA* 108(5):1815–1820.

Wilson JT (1963) A possible origin of the Hawaiian Islands. *Can J Phys* 41(6):863–870.

Wilson MC (1996) Late Quaternary vertebrates and the opening of the ice-free corridor, with special reference to the genus *Bison. Quat Int* 32:97–105.

Winger BM, Bates JM (2015) The tempo of trait divergence in geographic isolation: Avian speciation across the Marañon Valley of Peru. *Evolution* 69(3):772–787.

Wollenberg K, Avise JC (1998) Sampling properties of genealogical pathways underlying population pedigrees. *Evolution* 52:957–966.

Wollstein A, Lao O, Becker C, Brauer S, Trent RJ, Nürnberg P, Stoneking M, Kayser M (2010) Demographic history of Oceania inferred from genome-wide data. *Curr Biol* 20(22):1983–1992.

Wood DA, Vandergast AG, Barr KR, Inman RD, Esque TC, Nussear KE, Fisher RN (2013) Comparative phylogeography reveals deep lineages and regional evolutionary hotspots in the Mojave and Sonoran Deserts. *Divers Distrib* 19(7):722–737.

Wood HM, Gillespie RG, Griswold CE, Wainwright PC (2015) Why is Madagascar special? The extraordinarily slow evolution of pelican spiders (Araneae, Archaeidae). *Evolution* 69(2):462–481.

Woodward SP (1851) *A Manual of the Mollusca; or, Rudimentary Treatise of Recent and Fossil Shells* (John Weale, London).

Wright JW (1983) The evolution and biogeography of the lizards of the Galapagos archipelago: Evolutionary genetics of *Phyllodactylus* and *Tropidurus* populations. *Patterns of Evolution in Galapagos Organisms*, eds Bowman RI, Berson MI, Leviton AE (AAAS Symposium Volume, San Francisco), pp. 123–155.

Wright JW, Stanton ML, Scherson R (2006) Local adaptation to serpentine and non-serpentine soils in Collinsia sparsiflora. *Evol Ecol Res* 8(1):1–21.

Wright S (1931) Evolution in Mendelian populations. *Genetics* 16(2):97–159.

Wright S (1937) The distribution of gene frequencies in populations. *Proc Natl Acad Sci USA* 23(6):307–320.

Wright S (1943) Isolation by distance. *Genetics* 28(2):114–138.

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005) The effects of artificial selection on the maize genome. *Science* 308(5726):1310–1314.

Wu C-I (1991) Inferences of species phylogeny in relation to segregation of ancient polymorphisms. *Genetics* 127(2):429–435.

Wu C-I (2001) The genic view of the process of speciation. *J Evol Biol* 14(6):851–865.

Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, Hooper SD, Pati A, Lykidis A, Spring S, Anderson IJ, D'haeseleer P, Zemla A, Singer M, Lapidus A, Nolan M, Copeland A, Han C, Chen F, Cheng JF, Lucas S, Kerfeld C, Lang E, Gronow S, Chain P, Bruce D, Rubin EM, Kyrpides NC, Klenk HP, Eisen JA (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 462(7276):1056–1060.

Wu Y (2012) Coalescent-based species tree inference from gene tree topologies under incomplete lineage sorting by maximum likelihood. *Evolution* 66(3):763–775.

Xue AT, Hickerson MJ (2015) The aggregate site frequency spectrum for comparative population genomic inference. *Mol Ecol* 24(24):6223–6240.

Yamagishi H, Allison TD, Ohara M (2005) Effect of snowmelt timing on the genetic structure of an *Erythronium grandiflorum* population in an alpine environment. *Ecol Res* 20(2):199–204.

Yang ZH (2015) The BPP program for species tree estimation and species delimitation. *Curr Zool* 61(5):854–865.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proc Natl Acad Sci USA* 107(20):9264–9269.

Yang Z, Yoder AD (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(5):705–716.

Yeaman S (2013) Hybridization and the porous genome: Patterns of isolation and introgression in manakins. *Mol Ecol* 22(12):3195–3197.

Yergeau E, Kang S, He Z, Zhou J, Kowalchuk GA (2007) Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect. *ISME J* 1(2):163–179.

Yoder AD, Yang Z (2004) Divergence dates for Malagasy lemurs estimated from multiple gene loci: Geological and evolutionary context. *Mol Ecol* 13(4):757–773.

Yoder AD, Rasoloarison RM, Goodman SM, Irwin JA, Atsalis S, Ravosa MJ, Ganzhorn JU (2000) Remarkable species diversity in Malagasy mouse lemurs (primates, *Microcebus*). *Proc Natl Acad Sci USA* 97(21):11325–11330.

Yoon S, Cruz-García C, Sanford R, Ritalahti KM, Löffler FE (2015) Denitrification versus respiratory ammonification: Environmental controls of two competing dissimilatory NO_3^-/NO_2^- reduction pathways in *Shewanella loihica* strain PV-4. *ISME J* 9(5):1093–1104.

Yu Y, Degnan JH, Nakhleh L (2012) The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genet* 8(4):e1002660.

Yue H, Fan Z, Liu S, Liu Y, Song Z, Zhang X (2012) A mitogenome of the Chevrier's field mouse (*Apodemus chevrieri*) and genetic variations inferred from the cytochrome *b* gene. *DNA Cell Biol* 31(4):460–469.

Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292(5517):686–693.

Zazula GD, MacKay G, Andrews TD, Shapiro B, Letts B, Brock F (2009) A late Pleistocene steppe bison (*Bison priscus*) partial carcass from Tsiigehtchic, Northwest Territories, Canada. *Quat Sci Rev* 28(25-26):2734–2742.

Zehr SM, Taylor JP, Roach RG, Haring D, Cameron FH, Dean M, Yoder AD (2014) Prosimian primate life history profiles generated from the new Duke Lemur Center Database (coming soon to a URL near you!). *Am J Phys Anthropol* 153:281.

Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, Li P, Yuldasheva N, Ruzibakiev R, Xu J, Shu Q, Du R, Yang H, Hurles ME, Robinson E, Gerelsaikhan T, Dashnyam B, Mehdi SQ, Tyler-Smith C (2003) The genetic legacy of the Mongols. *Am J Hum Genet* 72(3):717–721.

Zhou L, Bawa R, Holliday JA (2014) Exome resequencing reveals signatures of demographic and adaptive processes across the genome and range of black cotton-wood (*Populus trichocarpa*). *Mol Ecol* 23(10):2486–2499.

Zhu S, Degnan JH, Steel M (2011) Clades, clans, and reciprocal monophyly under neutral evolutionary models. *Theor Popul Biol* 79(4):220–227.

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* (*Basel*) 69(2):106–114.

Zink RM (2002) Methods in comparative phylogeography, and their application to studying evolution in the North American aridlands. *Integr Comp Biol* 42(5):953–959.

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Mol Ecol* 17(9):2107–2121.

Zohdy S, Gerber BD, Tecot S, Blanco MB, Winchester JM, Wright PC, Jernvall J (2014) Teeth, sex, and testosterone: Aging in the world's smallest primate. *PLoS One* 9(10):e109528.

Index