Local Adaptation along Smooth Ecological Gradients Causes Phylogeographic Breaks and Phenotypic Clustering

Darren E. Irwin*

Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

Submitted August 16, 2011; Accepted March 9, 2012; Electronically published May 24, 2012

Dryad data: http://dx.doi.org/10.5061/dryad.71627sm2.

abstract: Coalescent theory has provided a basis for evolutionary biologists to build sophisticated methods for inferring population history from variation in genetic markers, but these methods leave out a major conceptual cornerstone of modern evolutionary theory: natural selection. I provide the first quantitative analysis of the effects of selection on genealogical patterns in a continuously distributed population in which the selective optimum for a trait linked to the marker varies gradually and continuously across the landscape. Simulations show that relatively weak selection for local adaptation can lead to strong phylogeographic structure, in which highly divergent genealogical groups (i.e., clades) are geographically localized and differentially adapted, and dramatically increased standing variation (e.g., coalescence time) compared to neutral expectations. This pattern becomes more likely with increasing population size and with decreasing dispersal distances, mutation rates, and mutation sizes. Under some conditions, the system alternates between a nearly neutral behavior and a behavior in which highly divergent clades are locally adapted. Natural selection on markers commonly used in phylogeographic studies (such as mitochondrial DNA) presents a major challenge to the inference of biogeographic history but also provides exciting opportunities to study how selection affects both between- and within-species biodiversity.

Keywords: coalescent theory, genealogy, local adaptation, mitochondrial DNA, natural selection, phylogeography.

Introduction

In the more than three decades since the publication of the first phylogeographic network (Avise et al. 1979), molecular ecologists have produced many thousands of intraspecific gene genealogies from a myriad of species throughout the tree of life (Avise 2000, 2004; Hickerson et al. 2010). The results are used to infer patterns of current and past gene flow as well as current and past changes in effective population size. One especially common finding is a pattern of spatially localized clades (Avise 2000) that are often highly divergent in terms of the number of mutational steps that separate them compared to the variation within each. For example, in the pocket gophers studied by Avise et al. (1979), there were distinct western and eastern mtDNA clades, with no geographic overlap and relatively little within-clade variation. The typical interpretation of such a pattern is nicely summarized by Avise (2004, p. 288): "Presumably, the localization of ... clades in most species reflects contemporary restraints on gene flow ..., and many of the deeper genetic breaks ... register much longer-term historical population separations." A variety of widely used and sophisticated analytical methods to analyze phylogeographic data are based on this general interpretive framework; these include, for example, the IM ("Isolation with Migration") model (Hey 2005), nested clade analysis (Templeton 1998), and Bayesian phylogeography (e.g., BEAST; Drummond et al. 2002; Drummond and Rambaut 2007; Lemey et al. 2009).

A remarkable and often unstated assumption of these standard approaches to phylogeographic analysis is that the molecular markers under study are selectively neutral: all genetic variants are assumed equal in terms of their effects on fitness (Hein et al. 2005; Dowling et al. 2008; Wakeley 2008). Natural selection is a major conceptual cornerstone of modern evolutionary theory (Schluter 2000); hence, it might seem surprising that a major subfield would rely on methods that assume selection does not exist. This assumption is made because it enables the use of coalescent theory (Kingman 1982*a*, 1982*b*) to make inferences from data regarding historical population sizes, migration rates, and the timing of various events (Hein et al. 2005; Wakeley 2008). By assuming selective neutrality, one can derive relatively straightforward equations regarding probabilities that two current individuals will share common ancestors at various points in the past, given a specified demographic model.

Making an assumption because it makes theory tractable is appealing, but it does not mean the assumption is valid. In fact, there are strong reasons to think natural selection

* E-mail: irwin@zoology.ubc.ca.

Am. Nat. 2012. Vol. 180, pp. 35–49. 0003-0147/2012/18001-53261$15.00.
DOI: 10.1086/666002

could commonly apply to the markers often used in phylogeographic studies. These typically include DNA fragments from unusual but essential parts of the genome such as mitochondrial DNA (mtDNA), chloroplast DNA (cpDNA), the Y chromosome (e.g., in mammals), and the W chromosome (e.g., in birds) but occasionally include parts of autosomal nuclear DNA (e.g., introns). The most commonly used type of DNA in phylogeographic studies has been mitochondrial DNA, which is maternally inherited in most taxa, meaning that the whole mitochondrial genome (typically about 13 protein-coding genes and 17,000 bp in vertebrates; Desjardins and Morais 1990) is passed down as a single unit through the matrilineal genealogy. Such uniparentally inherited markers (most mtDNA, some cpDNA, Y chromosome, and W chromosome) have been particularly appealing for phylogeographic analysis because they are simpler to sequence than biparentally inherited nuclear DNA and usually do not undergo recombination, which would otherwise complicate genealogical inference. These uniparentally inherited markers are far from unimportant to the organism. The genes in mtDNA, for example, code for proteins that play a central role in metabolism (Das 2006; Wallace 2007), and there are strong theoretical arguments for why such proteins could be under selection (Ballard and Kreitman 1995; Hudson and Turelli 2003; Ballard and Whitlock 2004; Wallace 2007; Dowling et al. 2008); these arguments are captured most succinctly by Ballard and Whitlock's (2004, p. 737) "Freshman Law of Selective Neutrality: If it's important enough to learn about in first-year biology, it's probably under strong selection." Indeed, the characteristics of the mitochondrial genome might lead us to suspect that it is far from selectively neutral: it is a perfectly linked group of essential metabolic genes that undergo a high mutation rate (Howell et al. 2003; Santos et al. 2008), and it is easy to conceive of scenarios by which a changing external environment could change the optimal characteristics of at least one of the enzymes encoded by those genes. Because of perfect linkage, any selection applied to one of the genes then affects variation throughout the entire mitochondrial genome. It is also likely that the genes in the mitochondrial genome coevolve with some nuclear genes, especially those that actually encode mitochondrial proteins (Dowling et al. 2008). These plausibility arguments are supported by growing empirical evidence from the fields of phylogeography (Ballard and Kreitman 1995; Bazin et al. 2006; Cheviron and Brumfield 2009; Irwin et al. 2009; Brelsford et al. 2011; Ribeiro et al. 2011), laboratory genetics (Stewart et al. 2008), protein biochemistry (Garvin et al. 2011), and mitochondrial physiology (Dhillon and Schulte 2011). Clearly, there is a need to assess the effects that selection may have on phylogeographic patterns.

Here I provide the first quantitative analysis of genealogical patterns that are expected under a simple model of gradually varying natural selection across continuous space. I envision a locus that specifies a phenotypic trait that is subject to natural selection; the optimal value for this phenotypic trait varies gradually across the range. I conduct individual-based computer simulations of a species evolving on this range and keep track of the genealogy of individuals with respect to this locus as well as the distribution of phenotypic values across the range. The analysis is intended as a first step in a process of incorporating selection into phylogeographic theory and as such is intended to be as simple as possible while incorporating such biologically realistic factors as mutation, local competition, variation in reproduction, and variation in dispersal distances. Genealogies within a continuously distributed species with limited dispersal have not been modeled analytically because of the mathematical difficulties involved (Felsenstein 1975; Barton and Wilson 1995; Wilkins 2004), necessitating the simulation approach used here and in several studies that have modeled neutral genealogies in such species (Irwin 2002; Rauch and Bar-Yam 2004; Kuo and Avise 2005).

While full genealogies have not previously been modeled under such a scenario of a gradually changing ecological optimum, a number of authors have presented models of how a trait subject to such selection varies across a species' range. These fit into two general categories. First, quantitative genetic models result in trait values varying gradually across the range, at each location corresponding to the ecological optimum or being gradually displaced from that optimum as a result of gene flow alone (Slatkin 1973; Kirkpatrick and Barton 1997) or gene flow in combination with a gradually changing environment (Pease et al. 1989) or interspecific competition (Case and Taper 2000). These models were based on deterministic analyses that assumed no stochastic effects due to small local population sizes (i.e., genetic drift), and as quantitative genetic models they effectively assumed that the phenotypes under study were influenced by many loci. Second, Doebeli and Dieckmann (2003) presented an individual-based model in which an asexual population evolves on an ecological gradient; this model is conceptually related to the one considered here, in which uniparentally inherited loci are evolving in a sexual species, but only the present model tracks genealogical relationships as the system evolves. In the Doebeli and Dieckmann asexual model, "evolutionary branching" in the phenotypic trait occurs when the ecological gradient is of moderate steepness compared to the strength of selection. As I show here, a phenotypic trait encoded by a uniparentally inherited marker (such as those usually used for phylogeographic analysis) will under many conditions show the very clustered distribution of phenotypes across

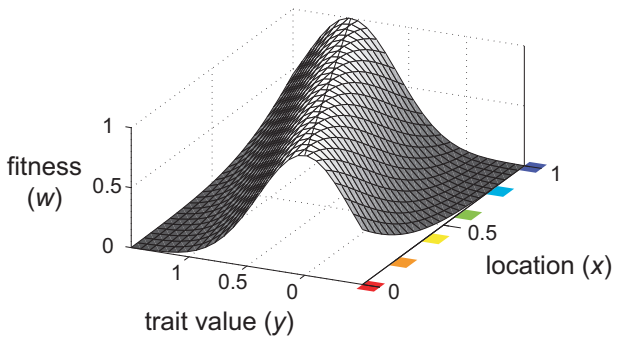


Figure 1: Illustration of how potential fitness of an individual in the model depends on its trait value (y) and its location (x), with the optimal trait value changing linearly across the range. Location (x) and the optimal trait value (y_{opt}) always have the same numerical value ($y_{opt} = x$), with fitness at a given location declining from the optimum trait value according to a Gaussian function with standard deviation σ_w. For purposes of illustration, this figure shows very strong selection ($\sigma_w = 0.4$; table 1); most of the simulations used much weaker selection, under which conditions the selection curve is more difficult to perceive visually. Colors along the location axis indicate the six sampling sites used to generate genealogies (e.g., fig. 3), with 10 individuals sampled from each site. Note that fitness shown in this figure refers only to that component of fitness due to the ecological trait, whereas total fitness of individuals also depended on local population density (see "Methods").

a species range displayed by Doebeli and Dieckmann's (2003) model rather than the gradual phenotypic variation across the range produced by quantitative genetic models. These phenotypic clusters correspond to the underlying genetic structure in the locus that encodes the phenotypic trait.

While the above models have examined spatial variation in selection without tracking genealogies, a variety of theoretical analyses have examined the effects of very simple forms of selection on genealogies in populations without spatial variation (e.g., Takahata 1990; Neuhauser and Krone 1997; Maia et al. 2004; O'Fallon et al. 2010). The emerging consensus is that even weak forms of natural selection can have strong effects on depths and shapes of the genealogies (O'Fallon et al. 2010). Here my intent is to provide the first analysis of how spatially varying selection affects genealogies in continuous space.

I focus on a question centrally important to the fields of phylogeography and historical demography: can deep phylogeographic breaks (which are traditionally interpreted to be the result of long-term barriers to gene flow) be produced as a result of selection along an environmental gradient? If so, how strong does that selection need to be? If relatively weak selection can dramatically change phylogeographic patterns from those expected under neutrality, we may need to dramatically rethink our approaches to inferring process from phylogeographic pattern.

Methods

To model the effects of a simple form of selection on uniparentally inherited gene genealogies, I assumed an ecological gradient in some environmental parameter (e.g., temperature, salinity) and an associated phenotypic characteristic (the "trait") of individuals (e.g., temperature that maximizes physiological performance) that is determined by their genotype. Fitness (i.e., reproduction) of each individual could then be determined in part by the similarity of the trait value to the optimum trait value at the individual's breeding location. The ecological optimum varied linearly across a one-dimensional range. At each location, I specified a Gaussian individual fitness function centered on the ecological optimum for the trait, with standard deviation σ_w (fig. 1) describing the width of the fitness peak at each location (smaller σ_w indicates stronger selection toward the ecological optimum; see table 1). I then modeled an evolving population on this landscape, with steps for reproduction, mutation, and dispersal each generation, while keeping track of genealogical relationships. After each simulation ran for a sufficient period of time, the genealogy of the resulting population could be analyzed in terms of its phylogeographic structure and its coalescence time (the number of generations back to the most recent common ancestor of all current individuals).

The model was designed to apply most directly to uni-

Table 1: Relationship between the width of the selection curve (σ_w) and the fitness ratio between individuals perfectly adapted to locations $x = 0.25$ and $x = 0.75$, when compared at one of those locations

σ_w	Fitness ratio
.4	.4578
.8	.8226
1.6	.9523
3.2	.9879
6.4	.9970
12.8	.9992
25.6	.9998

Note: Fitness ratio is the fitness of an individual with trait value $y = 0.75$ divided by the fitness of an individual with $y = 0.25$, when compared at location $x = 0.25$. This is intended as a rough estimate of the maximum strength of selection that would occur in a simulation run with that σ_w; realized selection would usually be much less strong, since individuals rarely disperse that far across the range.

parentally inherited markers (because those are used most commonly in phylogeographic inference); hence it is a model of matrilineal genealogies when considering mitochondrial DNA (in most species) and the W chromosome (e.g., in birds and butterflies) and a model of patrilineal genealogies when considering the Y chromosome (e.g., in mammals). For markers with more complex inheritance patterns, such as cpDNA and nuclear introns, the model here still applies as long as recombination has not affected inferred genealogical relationships. In all of these cases, sexual reproduction can be ignored; hence there is no mating in the model. Note, however, that the effective population size differs between strictly uniparentally inherited markers and markers with more complex inheritance patterns (i.e., a matrilineal or patrilineal marker generally has one-fourth the effective population size of an autosomal marker; Birky et al. 1989; Hudson and Turelli 2003); results will be interpreted primarily with respect to matrilineal or patrilineal markers.

Simulations were conducted in the Matlab programming environment (Mathworks, Natick, MA), using scripts based on those used in a model of genealogies evolving under selective neutrality and limited dispersal (Irwin 2002) and expanded by incorporating an ecological trait and a gradually varying ecological optimum. Under each set of parameters, a population was initiated and then allowed to evolve for thousands (or millions, in some cases) of generations, while the computer kept track of genealogical relationships. Each generation consisted of reproduction, mutation, and dispersal of offspring, and generations were nonoverlapping. Natural selection was incorporated into the reproduction phase, with the expected number of offspring determined in part by an individual's fitness in the local environment. I describe these steps in detail below.

Simulation Details

Each simulation started with a population of N individuals with random locations (along a one-dimensional range of length 1) and ecological traits: individual i was assigned location x_i and trait y_i, each drawn from a uniform distribution between 0 and 1.

The number of offspring for each individual was determined by drawing from a Poisson distribution with a mean adjusted to account for both local density dependence and selection on the trait. First, the reproductive factor due to local density dependence was approximated by dividing the total range into 50 segments, each of which had a carrying capacity of $k_d = N/50$ and then calculating the ratio of that carrying capacity and the number of individuals in segment d: $R_d = k_d/n_d$. Second, selection on each individual due to the ecological trait was determined by a Gaussian function:

$$w_i = \exp\left[\frac{-(y_i - x_i)^2}{2\sigma_w^2}\right],$$

where σ_w is the width of the selection curve. According to this function, the ecologically optimum trait (y_{opt}) at each location (x) has the same numerical value as the location ($y_{opt} = x$; both range from 0 to 1 across the linear range). The number of offspring of individual i in segment d was then determined by drawing from a Poisson distribution with mean equal to $R_d \times w_i$.

Offspring inherited their parent's trait values, except in the case of mutation, which occurred with frequency μ per individual. When a mutation occurred, the offspring's trait value was determined by drawing from a Gaussian distribution with standard deviation σ_{mut} centered on the parent's trait value y.

The breeding location of each offspring (after dispersal) was determined by drawing from a Gaussian distribution centered on the birth location, with standard deviation σ_{disp}. Locations outside of the geographic range (i.e., lower than 0 or greater than 1) were not allowed; draws were repeated until a breeding location within the range was determined.

Length and Number of Simulations

The goal of the simulations was to sample genealogies that would be expected in a population evolving under the parameters and rules applied for an indefinitely long period of time, such that genealogies are independent of starting conditions. To ensure that the simulations were run long enough to reach this state, I did the following. In each simulation, the generation at which all individuals became descended from a single initial individual was recorded (fig. 2), and the stopping time of the simulation was then determined as 2–3 times (drawn from a uniform distribution) this number of generations. Under a given set of parameters, 10 simulations were run according to the above, and then the maximum coalescence time out of the final 10 populations was multiplied by 2; this became the new minimum number of generations that each simulation should be run. Any simulation using that parameter set that had not yet run for that amount of time was then continued until that number of generations was reached. The goal of this process was to run each simulation a number of generations sufficient to ensure that the resulting properties of the genealogy (e.g., coalescence time) were independent of starting conditions. This seemingly cumbersome process was necessary because it was impossible to determine in advance how long simulations

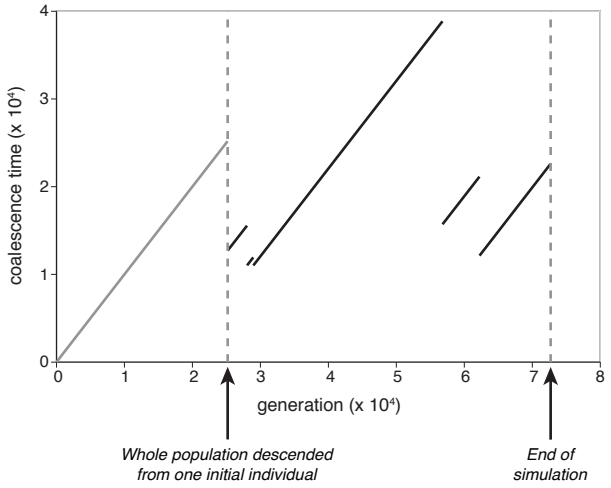


Figure 2: An example of how coalescence time of a population changes from generation to generation throughout a single simulation. This simulation begins with $N = 1{,}600$ individuals on the fitness landscape similar to that shown in figure 1 but under conditions of near neutrality ($\sigma_w = 25.6$; table 1) and high dispersal ($\sigma_{disp} = 0.16$). At the start of the simulation, coalescence time cannot be determined because there is no history to the population (indicated by the gray line to the left of the dashed gray line). After roughly 2,500 generations (in this particular run), the population becomes descended from only a single starting individual (i.e., lineages descending from all other starting individuals have become extinct), such that coalescence time can then be determined (indicated by the black lines after this point). Thereafter, coalescence time (measured in generations) of the population increases by one each generation, unless one of the two most divergent clades leaves no offspring; in the latter case, coalescence time collapses to the coalescence time of the two deepest clades remaining. After an arbitrary number of generations (in this case, just beyond 7,000 generations) sufficient to ensure that results are independent of starting conditions (see "Methods"), the simulation stops and a genealogy can be sampled from the population (fig. 3). Other parameter settings were $\mu = 0.0025$ and $\sigma_{mut} = 0.05$.

should be run to reach independence from starting conditions.

Fifty simulations were run under each set of parameters, except in cases in which one of the first 10 simulations took longer than 100,000 generations for the population to become descended from a single starting individual, in which case only 10 simulations were run under that set of parameters. Under some sets of parameters, locally adapted clades are so stable that the simulated population takes an extremely high (under some conditions an effectively infinite) number of generations to become descended from a single initial individual, making it impossible to obtain results. Hence, exact quantitative results were not obtained under parameter sets that commonly took longer than 300,000 generations to become descended from a single individual.

Parameters

Given that the model has five parameters, I used an exploratory approach to search parameter space to understand how genealogical patterns depended on conditions. I explored combinations of the following parameter settings: $N = (200; 400; 800; 1{,}600)$, $\sigma_w = (0.4; 0.8; 1.6; 3.2; 6.4; 12.8; 25.6)$, $\sigma_{disp} = (0.01; 0.02; 0.04; 0.08; 0.16)$, $\mu = (0.00125; 0.0025; 0.005; 0.01; 0.02; 0.04)$, and $\sigma_{mut} = (0.00625; 0.0125; 0.025; 0.05; 0.1; 0.2; 0.4; 0.8)$. These parameter values were chosen to cover as much of the realistic parameter space as possible, given the constraints of computation time. Note that the mutation rate (μ) refers to the per-individual rate of trait-changing mutation over the entire uniparentally inherited genome. While this has not been measured empirically, the per-individual rate of mutation for large regions of mtDNA has. In a metanalysis of studies of human mtDNA, Howell et al. (2003) documented a total per-transmission (i.e., mother to daughter) mutation rate of 0.011 across 1,122 bp of the control region. Santos et al. (2008) measured a mutation rate of 0.006 across 1,102 bp of coding mtDNA. The per-transmission mutation rate across the whole mtDNA genome (roughly 17,000 bp) would be much larger (e.g., extrapolating from the above, roughly 0.1). To convert that estimate to the mutation rate as used in the present model, we would need to know what fraction of mutations affect the trait under selection; assuming this fraction is 0.01 or larger, the range of values for μ used in the model includes realistic values.

A total of 18,890 simulations were run at 413 combinations of the above parameters, encompassing a total of 4.2×10^8 generations and 3.5×10^{11} individuals.

Quantification of Phylogeographic Structure

Results were summarized in two ways. First, the coalescence time of each genealogy was recorded as the number of generations to the most recent common ancestor of all current individuals. This can be compared to the expected matrilineal (or patrilineal) coalescence time in a single nonspatially structured population of N females (or males) with no selection, which is given by $2N$ (Avise 2000; Hein et al. 2005; Wakeley 2008). Second, following Irwin (2002), to summarize phylogeographic structure within each completed simulation, 60 individuals were sampled across the range: 10 from each of six locations that were evenly distributed across the linear range (at locations 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0; hereafter referred to as sites 1 through 6). Average coalescence time between individuals of adjacent locations, D, was then calculated for each pair of adjacent locations. The ratio of the maximum to the min-

imum D (i.e., D_{max}/D_{min}) was then used as an index of phylogeographic structure (Irwin 2002).

All 18,890 genealogies and the trait values of individuals in those genealogies (60 per genealogy) have been deposited in the Dryad repository: doi:10.5061/dryad.71627sm2.

Results

Under conditions of extremely weak selection (σ_w = 25.6; table 1) and high dispersal (σ_{disp} = 0.16), the model produces results that are consistent with the standard nonspatial neutral model (figs. 3, 4). There is little if any phylogeographic structure (mean D_{max}/D_{min} = 1.4; e.g., fig. 3), and the mean coalescence time is approximately $2N$ (fig. 4). Even under these conditions, there is high variation in the coalescence time because of the stochastic nature of the coalescent process (Hein et al. 2005; Wakeley 2008). Genealogies produced under such neutral conditions often consist of two relatively divergent clades (e.g., fig. 3), but they generally are not localized on the landscape.

Selection can have dramatic effects on these genealogical patterns. Under fairly weak selection (σ_w = 3.2, corresponding to a rough maximal fitness ratio of 0.988 between different individuals; see table 1) and moderate dispersal (σ_{disp} = 0.02), with all other parameters identical to the nearly neutral case described above, genealogies tend to become highly geographically structured (out of 50 runs: mean D_{max}/D_{min} = 39.5; e.g., fig. 5), and coalescence times increase dramatically from the neutral expectation (mean of 26 times N, for N = 1,600). This is because different mitochondrial types, corresponding to distinct clades, tend to become adapted to different parts of the range (fig. 6). Once established, this situation is stable for long periods. Within each clade, haplotypes are nearly equivalent selectively, and hence coalescence time within each clade tends to be short, close to the neutral expectation of twice the number of females within the clade. In contrast, the coalescence time between the two clades can become very long, because each clade is well adapted to one portion of the range, giving it a selective advantage in that area. The two clades persist indefinitely, until a rare event leads to the extinction of one of them. This extinction can occur after a mutation causes an individual of one clade to have a trait value similar to the other clade; in that case, the mutant can eventually drift to higher frequency and replace the clade that was previously common in that part

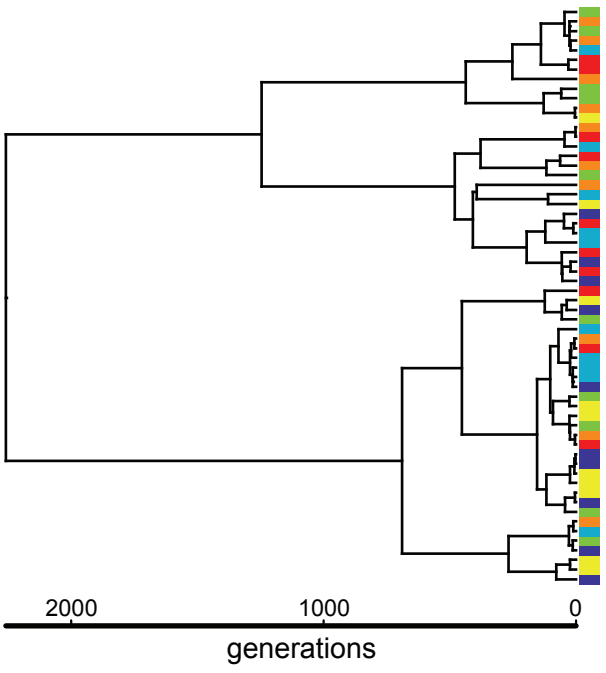


Figure 3: An example genealogy produced by the model under conditions of near neutrality (σ_w = 25.6; table 1) and high dispersal (σ_{disp} = 0.16), showing little if any geographic structure. This genealogy was sampled from the population at the end of the simulation in figure 2. Colors at the tips refer to sampling locations (see fig. 1), and the scale bar represents generations back in time. Other parameter settings were N = 1,600, μ = 0.0025, and σ_{mut} = 0.05.

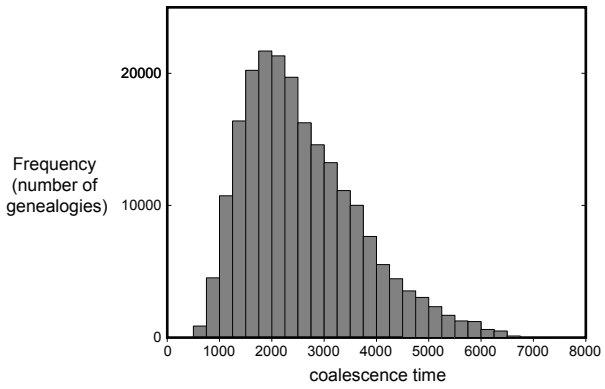


Figure 4: Histogram of coalescence times of populations (e.g., fig. 3) under conditions of near neutrality (σ_w = 25.6) and high dispersal (σ_{disp} = 0.16). Other parameter settings were N = 1,600, μ = 0.0025, and σ_{mut} = 0.05. The mean coalescence time (2,589) is relatively close to the theoretical expectation under a neutral nonspatial model of twice the population size (N = 1,600; actually, the mean is a bit less than $2N$ because the simulations tend to have slightly lower average population sizes than N and slightly higher variance in reproduction among individuals than would be expected in the nonspatial neutral model; both of these are a consequence of spatial variation in density during the simulations). To generate this histogram, coalescence times were compiled from the latter halves (in terms of generations) of 50 simulations run under these settings. Each of the 50 simulations is independent, but data from within each are not, leading this histogram to show slight nonindependence.

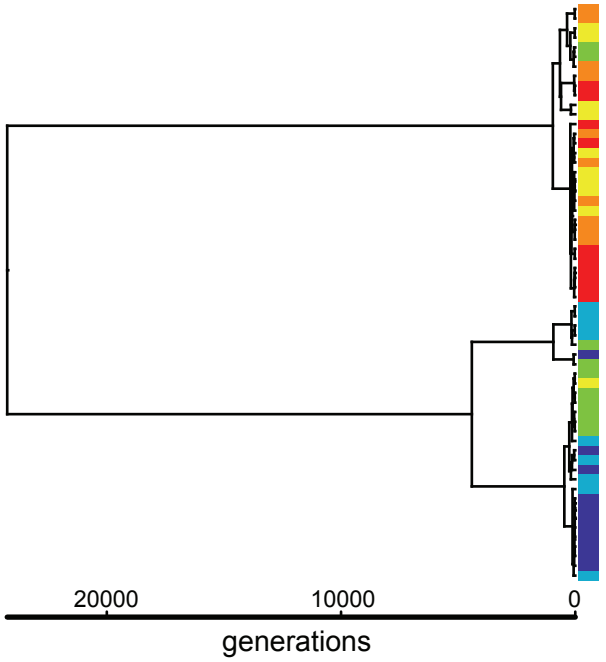


Figure 5: An example genealogy produced under conditions of fairly weak selection ($\sigma_w = 3.2$; table 1) and moderate dispersal ($\sigma_{disp} = 0.02$) showing two highly divergent clades that are geographically localized. Colors at the tips refer to sampling locations (see fig. 1), and the scale bar represents generations back in time. Other parameter settings were $N = 1{,}600$, $\mu = 0.0025$, and $\sigma_{mut} = 0.05$.

of the range. Under some conditions, such replacements are so rare that the two clades are essentially permanently stable.

Generally, phylogeographic structure tends to increase with increasing selection strength (i.e., narrowing width of the selection curve; table 1) and with decreasing dispersal distances (figs. 7, 8). The effects of population size, however, are interestingly dependent on other parameters. Under conditions of low dispersal distances (low σ_{disp}) and weak selection (high σ_w), increasing population size tends to decrease phylogeographic structure (figs. 7, 8; Irwin 2002; Kuo and Avise 2005). This is because the phylogeographic structure is due to the coalescent process being restricted by low dispersal distances (Irwin 2002); when population size is large, the coalescent process under the neutral model takes long enough (e.g., $2N$) that it is not limited by dispersal (see also Wilkins 2004). In stark contrast, when phylogeographic structure is driven by selection (low to intermediate σ_w), larger population sizes tend to stabilize geographic patterns: stochastic effects have a smaller influence at larger population sizes; hence selection has a more dominant role. While the simulations had limited population sizes due to constraints imposed by computation time, the influence of selection on phylogeographic structure is expected to have even larger effects at larger population sizes than those modeled here. It should also be noted that there is a noticeable threshold effect in the production of phylogeographic structure: under many parameter values, a slight change in dispersal, selection strength, or population size can cause resulting genealogies to switch from highly geographically structured to highly mixed across the landscape (figs. 7, 8).

Near this threshold the system switches back and forth between two behaviors: a nearly neutral behavior, in which coalescence times are short (e.g., roughly $2N$) and genealogical structure is weak, and a second behavior in which local adaptation leads to much structure and deep coalescence times. To demonstrate, the coalescence time of a population subject to a particular set of parameter values ($N = 800$, $\sigma_w = 1.6$, $\sigma_{disp} = 0.08$, $\mu = 0.0025$, and $\sigma_{mut} = 0.05$) was recorded at each generation for one million generations. Results (fig. 9*A*) show that for certain time periods, the population has low coalescence times, and there is frequent loss of one of the deepest clades, thereby keeping coalescence times low. During these time periods, there is little local adaptation, with the entire population having little variation in the selected trait, and little phylogeographic structure (fig. 9*B*, 9*D*). At other times, there is much local adaptation, leading to highly stable and geographically structured clades that persist for a long time (fig. 9*C*, 9*E*). Under such conditions, the overall distribution of coalescence times is highly skewed, with a

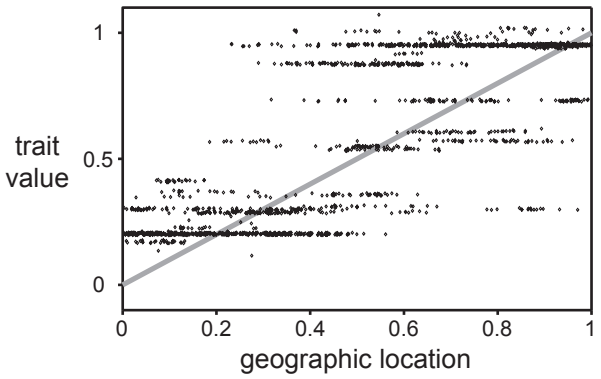


Figure 6: An example of trait values and geographic locations of all individuals at the end of a simulation under fairly weak selection and moderate dispersal. See figure 5 for parameter values, as well as a genealogy sampled from this population. In contrast with the ecological optimum, which changes gradually across the range (the gray line), traits of individuals tend to fall into discrete clusters, corresponding to distinct clades (fig. 5) adapted to different ecological conditions. A small amount of random jitter was applied to avoid overlap of individuals.

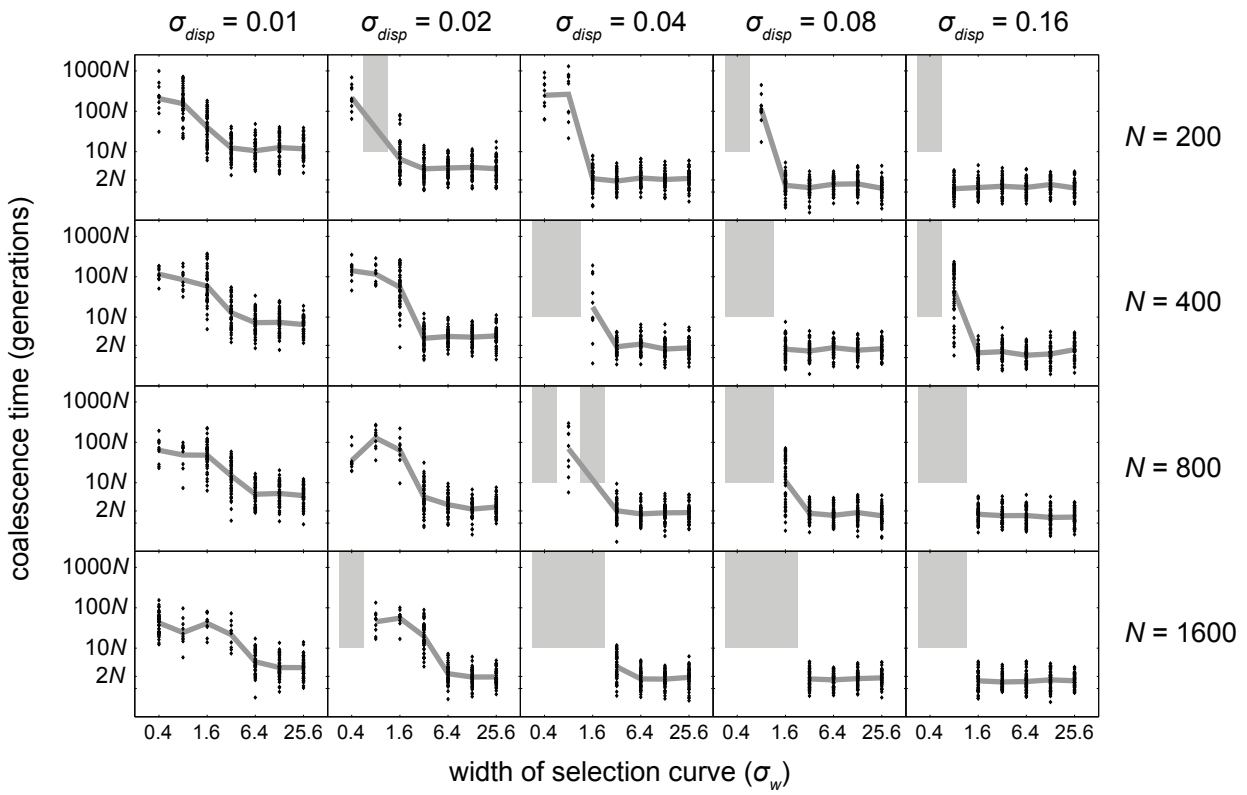


Figure 7: A subset of simulation results under fixed mutation rate ($\mu = 0.0025$) and mutation size ($\sigma_{mut} = 0.05$) shows that genealogical coalescence time (on a logarithmic scale) depends on the strength of selection (inversely correlated with width of the selection curve, σ_w; see table 1), individual dispersal distance (σ_{disp}), and population size (N). This figure summarizes coalescence times under 140 combinations of these parameters, each with 50 simulated genealogies, or 10 in cases that took long to simulate. Gray lines connect mean coalescence times across selection strengths under a given set of other parameters. Gray boxes indicate parameter combinations that produce very high coalescence times—so high that they could not be estimated accurately (likely bordering on infinite under some parameter combinations). Under high dispersal (high σ_{disp}) and low selection (high σ_w), coalescence times tend to be near the neutral expectation ($2N$), but they rise dramatically as selection increases and dispersal decreases. Generally, high coalescence times are associated with high phylogeographic structure; see figure 8.

range extending from well below $2N$ to well over $100N$ (fig. 10).

An unexpected result is that phylogeographic structure depends critically on both the mutation rate and the distribution of mutation sizes (fig. 11). Under most conditions, phylogeographic structure tends to decrease with increasing mutation rate and mutation size. This is because mutation can cause an individual in a clade adapted to one part of the range to mutate to a trait value well adapted to another part of the range, eventually leading to replacement (through drift) of the clade adapted to that other part of the range; these large mutations thus increase the potential for clades to replace each other. It should also be noted that under conditions of low population size, very low mutation rates and sizes tend to prevent phylogeographic structure, apparently because they prevent the establishment of locally adapted clades in the first place (fig. 11).

Discussion

These results provide the first quantitative demonstration that strong phylogeographic structure can arise under rather weak selection in a continuously distributed species with gradual environmental variation and that this effect increases with increasing population size. The great majority of phylogeographic studies do not consider this possibility, instead assuming that the molecular markers under study are selectively neutral and that phylogeographic breaks are the result of long-term barriers to gene flow (e.g., Avise 2000). There is increasing awareness of the fact that phylogeographic breaks in neutral markers can in

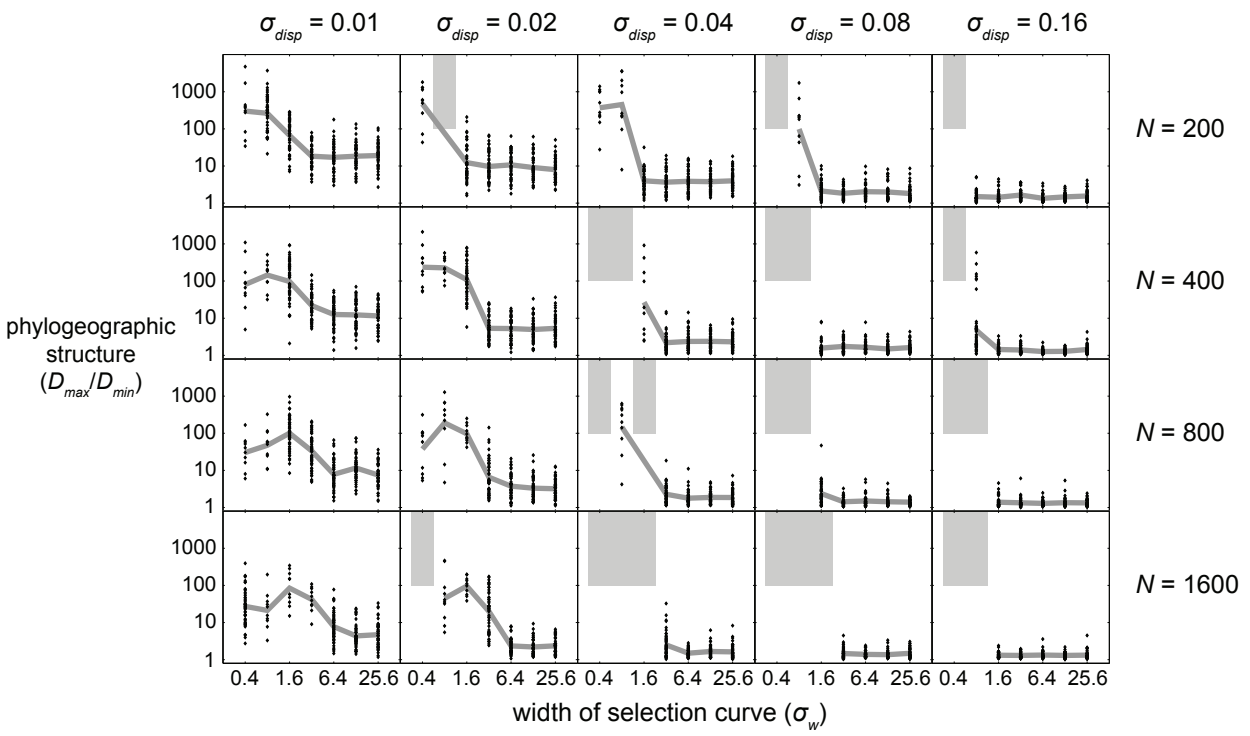
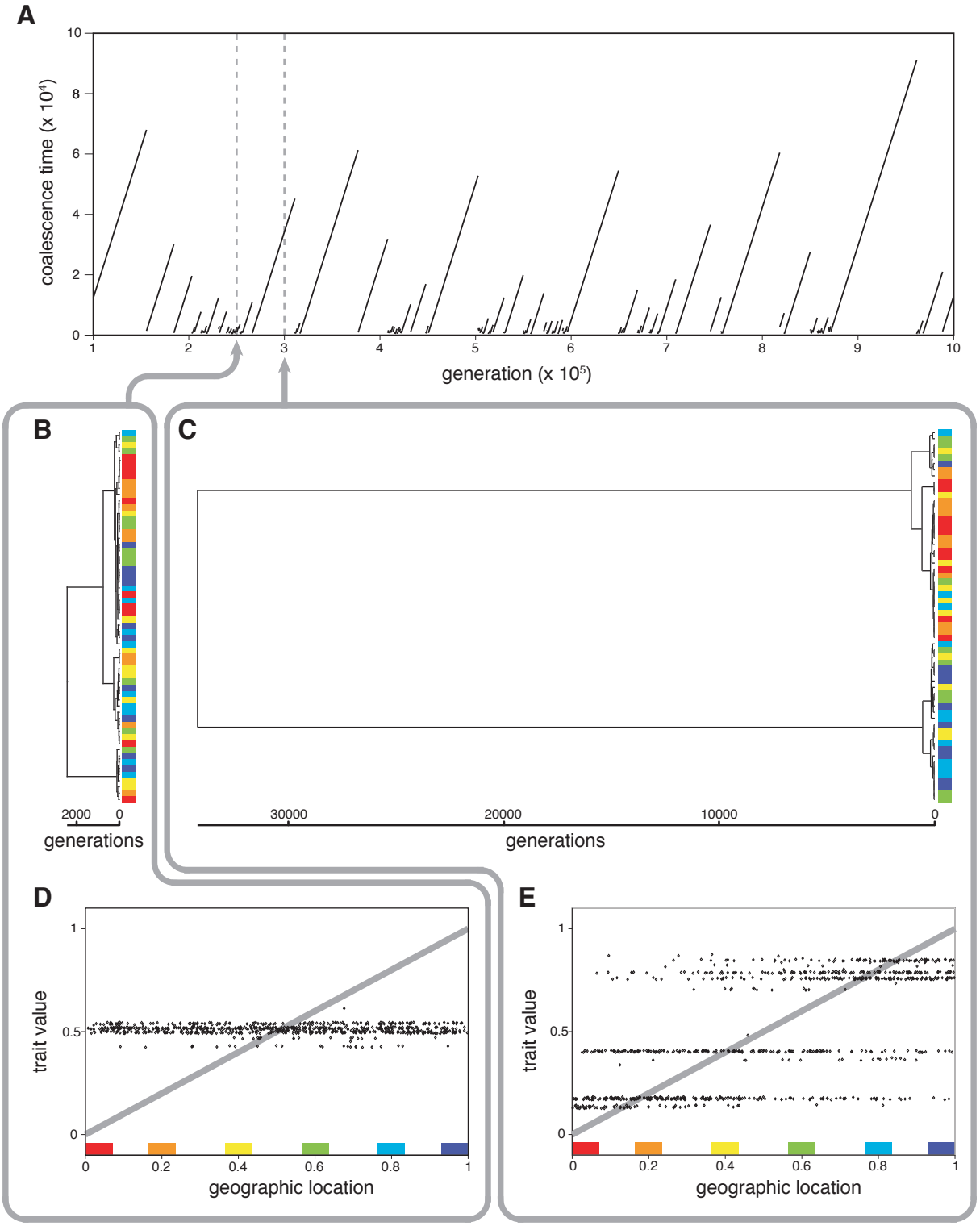


Figure 8: Phylogeographic structure, as measured by D_{max}/D_{min}, depends on the strength of selection (inversely correlated with width of the selection curve, σ_w; see table 1), individual dispersal distance (σ_{disp}), and population size (N). This figure shows the same subset of simulations ($\mu = 0.0025$, $\sigma_{mut} = 0.05$) for which coalescence time is shown in figure 7. Generally, high phylogeographic structure is associated with high coalescence time. See figure 7 for additional details.

theory arise without geographic barriers to gene flow (Irwin 2002; Rauch and Bar-Yam 2004; Kuo and Avise 2005), but that effect requires very low dispersal distances and low population sizes. My results show that a small amount of selection for local adaptation dramatically increases the range of conditions under which phylogeographic breaks can arise. One particularly notable result is that selection-driven phylogeographic structure tends to increase with increasing population size. Many studies have discounted the possibility that phylogeographic breaks arise without geographic barriers because the species under study have large populations (e.g., Chen et al. 2011), but the results here show that species with large populations may be particularly likely to show phylogeographic structure when it is driven in part by selection.

Under conditions that produce strong geographic clustering of genealogical clades, there is also strong clustering of trait values, a result that seemingly contrasts markedly with deterministic analyses that predict that traits should vary gradually across the range when the ecological optimum varies continuously (Slatkin 1973; Pease et al. 1989; Kirkpatrick and Barton 1997; Case and Taper 2000). The difference is explained by three factors. First, these earlier studies used quantitative genetic models that implicitly assume the trait of interest is encoded by many genes of small effect, each of which segregates independently from the others, whereas in the present model a single inherited unit causes the trait. When considering traits encoded by uniparentally inherited markers, the latter assumption is appropriate because the entire uniparentally inherited genome can be thought of as a single unit in terms of how it is inherited, even though it often consists of many protein-coding genes. Second, the role of stochasticity due to finite population sizes was incorporated into the present simulations but was not included in the earlier deterministic models. When population sizes are finite, there are a limited number of mutations, giving rise to discrete trait values, and many of these variants are lost from the population as a result of genetic drift. Third, there is a strong correlation between genealogical relatedness of two individuals and the similarity of their trait values, because as genealogical distance increases there is more opportunity for mutation to change traits. Hence, trait values tend to be clustered into distinct groups that correspond to distantly related clades. The earlier studies did not incorporate this effect because they did not model genealogies.


A
coalescence time (x 10⁴)
generation (x 10⁵)
B
2000
0
generations
C
30000
20000
10000
0
generations
D
trait value
geographic location
E
trait value
geographic location

The present findings show that clustered distributions of trait values need not indicate that there are multiple adaptive peaks; the underlying selective forces can be weak and gradually varying geographically.

The phylogeographic structure and clustering of trait values in these simulations can be seen as a dynamic outcome of two competing forces: natural selection and genetic drift. All else being equal, genetic drift has greater influence on systems with smaller population sizes, whereas natural selection has greater influence on systems with larger population sizes. To understand the outcome of a simulation under a particular set of parameter values, we can consider the question, Is the population size sufficiently large that selection can overwhelm the homogenizing effects of dispersal and the stochastic effects of genetic drift? If the answer is yes, then the population tends to diverge into two genealogical clades that are adapted to different ecological optima, each having on average half the range. We can then ask the same question about each of the two clades; however, each clade differs from the overall population by having about half the population size, half the range of ecological optima, and half the range size in relation to dispersal distance. These differences between the clade and the overall population all tend to reduce the impact of selection in comparison with drift and dispersal. Perhaps these values are still sufficient to enable selection to overcome drift, in which case the clade divides into subclades with different ecological adaptations. But when the population is divided sufficiently that drift within each clade is stronger than selection, each clade will consist of highly related individuals that have similar trait values, and each clade will evolve in a nearly neutral manner. Overall, the presence of each major clade is stabilized over time because each is adapted to different parts of the range.

Clustering of trait values along an ecological gradient was also observed in the individual-based asexual model of Doebeli and Dieckmann (2003), which did not track genealogies. One interesting difference between the two models is that Doebeli and Dieckmann (2003) included phenotype-dependent local competition between individuals, such that individuals with similar phenotypes competed more strongly than individuals with more dissimilar phenotypes; local competition is also an important component of the present model, but in this case competition is based simply on the number of local individuals rather than their phenotypes. Thus, there is no phenotype-dependent local competition in the current model, and phenotypic clustering in this model cannot be understood as being driven by localized disruptive selection on phenotypes. Rather, phenotypic clustering is a result of the overall population being large enough that selection can overwhelm drift on a broad spatial scale, while within each clade drift prevents close adaptation to the ecological gradient. In fact, Doebeli and Dieckmann's results are re-

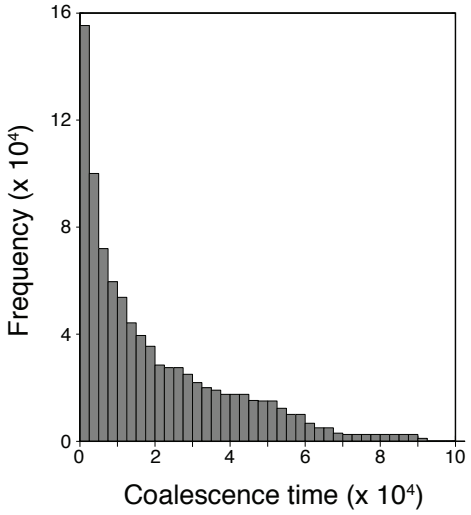


Figure 10: The distribution of coalescence times under conditions of the simulation shown in figure 9*A* is extremely skewed, far from that expected under conditions of neutrality (e.g., fig. 4).

Figure 9: Populations sampled at different times during one simulation can differ dramatically in their phylogeographic structure and trait distributions. *A*, Coalescence times during each of 900,000 generations of a simulated population (the first 100,000 generations of the simulation are not shown so as to avoid effects of initial conditions) under the following parameter values: $N = 800$, $\sigma_w = 1.6$, $\sigma_{disp} = 0.08$, $\mu = 0.0025$, and $\sigma_{mut} = 0.05$. Each generation, the coalescence time increases by one generation, unless there is an extinction of one of the two most divergent clades, in which case the coalescence time collapses to the coalescence time of the remaining two most divergent clades. *B*, *C*, Sampled genealogies at generations 250,000 (*B*) and 300,000 (*C*), with colors indicating the six sampling sites across the range (see fig. 1). The scale in generations is the same in the two genealogies. *D*, *E*, Trait values (y) and locations (x) of all individuals across the range at the time of sampling of genealogies *B* and *C*, respectively. Although both genealogies are sampled from a single population evolving under constant parameter values, they differ highly: the first example (*B* and *D*) shows no local adaptation and a shallow genealogy with little geographic structure, whereas the second example (*C* and *E*) shows deeply divergent clades that are strongly differentiated in the ecological trait and are strongly geographically structured. Examination of the coalescence time versus generation graph (*A*) shows that the population switches between different states, sometimes showing little geographic structure, short coalescence times, and essentially no local adaptation and other times showing strong phylogeographic structure, deep coalescence times, and clades that are adapted to different parts of the range.

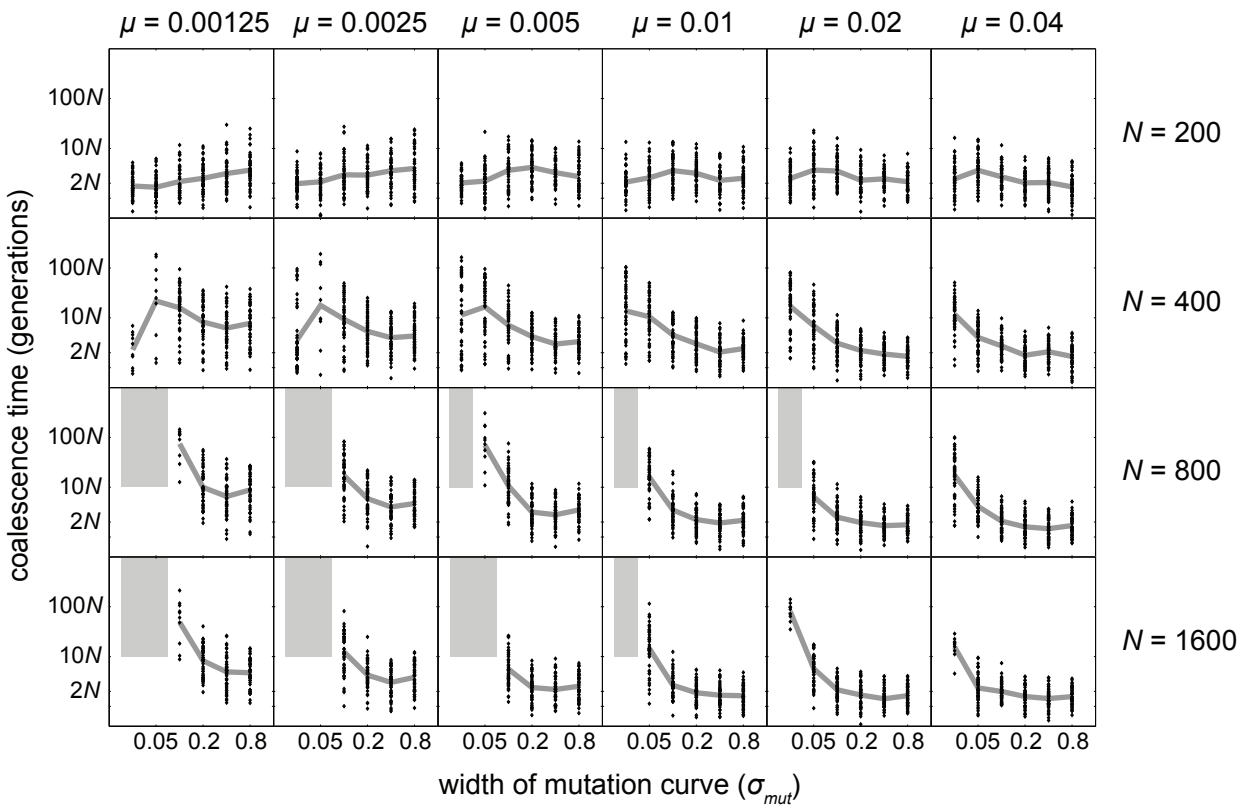


Figure 11: A subset of simulation results under fixed dispersal ($\sigma_{disp} = 0.04$) and selection ($\sigma_w = 1.6$) shows that genealogical coalescence time (shown on a logarithmic scale) depends on mutation rate (μ), mutation size (σ_{mut}), and population size (N). The figure summarizes coalescence times under 144 combinations of these parameters, each with 50 (or in some cases 10; see "Methods") independent simulated genealogies. Gray lines connect mean coalescence times across mutation sizes under a given set of other parameters. Gray boxes indicate parameter combinations that produce very high coalescence times—so high that they could not be estimated accurately (likely effectively infinite under some parameter combinations).

markably consistent with those presented here: they show that phenotypic clustering did not depend on phenotype-dependent local competition (i.e., when their $\sigma_c \gg \sigma_K$) as long as dispersal distances were sufficiently small.

For purposes of conceptual clarity and computational tractability, my model was designed to be simple, but we can consider how a variety of changes might affect the results. In particular, increasing the dimensionality of the species range from one to two is likely to increase phylogeographic structure when it is driven by selection, since increasing population size tends to strengthen phylogeographic structure. Likewise, changing the shape of the environmental gradient from linear across the range to a curved gradient that is steeper in one part of the range will also likely strengthen phylogeographic structure, since the boundary between clades would likely be stabilized in the region where environmental change is steepest. Note that such steep environmental change could also lead to concordance in phylogeographic breaks among multiple unlinked molecular markers, a pattern that is usually interpreted as evidence for a long-term barrier to gene flow (Kuo and Avise 2005).

An important simplification in the present model is that mutations change only the ecological optimum of a trait; all variants in the model have equal fitness when each is in its optimal environment. In reality, mutations might also change the absolute quality of a trait, leading to unequal fitnesses when each variant is in its optimal environment. There is much evidence for the majority of mutations being deleterious, leading to stabilizing selection (Stewart et al. 2008) and possibly compensatory mutations that correct for the accumulation of mildly deleterious mutations ("Muller's ratchet"; Lynch 1996; Rand 2008). It is likely that incorporating these factors into the present model would lead to differing effects depending on mutation rate and the distribution of beneficial and deleterious mutations. Under some conditions, new mutations could often be universally favored across the range, causing

selective sweeps and low phylogeographic structure, while under other conditions, Muller's ratchet could lead to compensatory mutations within each locally adapted type, thereby stabilizing phylogeographic structure.

Genealogies based on uniparentally inherited markers need not be predictive of overall patterns of variation in autosomal DNA, which is inherited through both female and male ancestors (Irwin 2002). Thus the results reported here add further evidence that phylogeographic breaks might not be due to long-term barriers to gene flow (Irwin 2002), potentially explaining situations in which mitochondrial phylogeographic breaks do not predict sharp transitions in phenotype and/or overall nuclear DNA (Irwin et al. 2005; Cheviron and Brumfield 2009; Ribeiro et al. 2011; see also Toews and Brelsford, forthcoming). Ribeiro et al. (2011) demonstrate a striking example: the Karoo scrub-robin (*Cercotrichas coryphaeus*) has divergent western and eastern mitochondrial types that have different amino acid substitutions in the ATPase6 gene, which plays a key role in the oxidative phosphorylation pathway, and the distribution of these types is well explained by a climatic gradient. In contrast, morphological variables and nuclear DNA (four nuclear introns and microsatellites) show little change across the climatic gradient, indicating extensive nuclear gene flow. Ribeiro et al. (2011) attribute the pattern to strong local adaptation of mitochondrial DNA and suggest that "selective pressures on physiology, mediated by the mitochondrial genome, may well be a common mechanism for facilitating local adaptation to new climatic conditions." Another potential example of mitochondrial adaptation to differing environmental conditions is provided by the killifish (*Fundulus heteroclitus*), which is distributed along a broad temperature gradient along the east coast of North America, from Florida to Newfoundland. Fangue et al. (2009) and Dhillon and Schulte (2011) have demonstrated a variety of differences in mitochondrial physiology between fish from divergent northern and southern mitochondrial clades. While it is unclear at the present time whether these differences arise from genes encoded in the nuclear or mitochondrial genomes (or both), the results do show that mitochondria in the northern and southern clades are exposed to different intracellular conditions, making it likely that selective forces on the mitochondrial genome differ substantially between the northern and southern clades. As Dalziel et al. (2009) point out, studies that examine the mechanistic links between genotype, phenotype, and fitness should play an increasingly important role in testing how natural selection influences variation in molecular markers.

The killifish and other examples point to the interesting possibility that local adaptation leading to phylogeographic structure could influence both mitochondrial genes and nuclear genes that interact with mitochondrial genes, since a number of proteins that function in the mitochondria are actually encoded by the nuclear genome. There is much current interest in the possibility of coevolution between these nuclear genes and mitochondrial genes ("cytonuclear coevolution"; Dowling et al. 2007, 2008). Similar coevolutionary dynamics could occur between the chloroplast genomes and nuclear genome because most proteins that function in chloroplasts are encoded in the nuclear genome (Myouga et al. 2010). The local adaptation in uniparentally inherited genomes seen in my models might induce the adaptive divergence of associated nuclear genes, which could likewise promote further divergence in the uniparentally inherited genomes. This coevolutionary process could lead to reduced fitness of individuals that have mismatched nuclear and cytoplasmic genes, a form of reproductive isolation between the diverging groups. If so, divergence in phylogeographic markers, while usually viewed as simply a consequence of reproductive isolation, might actually be an initial cause of speciation (Dowling et al. 2008).

While I have emphasized mitochondrial DNA because of its common use in the field of phylogeography, the conclusions reported here also apply to other forms of DNA that are uniparentally inherited, such as chloroplast DNA in plants (Kapralov and Filatov 2006), the W chromosome in birds and butterflies (Lane 2008), the Y chromosome in mammals, maternally inherited symbionts in arthropods (Hurst and Jiggins 2005), and to autosomal markers that have not experienced internal recombination since their coalescence time. Furthermore, when two such units of DNA are inherited in the same way (e.g., mtDNA and the W chromosome in birds), they together act as a single linked group in terms of how selection affects their joint genealogy (Lane 2008). This fact increases the potential role for selection in influencing phylogeographic patterns. Overall, these simulation results further point to the "profound, yet neglected" (Dowling et al. 2008) role for selection in shaping patterns of variation in uniparentally inherited DNA. Embracing the potential role of selection on uniparentally inherited DNA presents major challenges for traditional methods of phylogeographic inference but also provides exciting opportunities to better understand how selection shapes patterns of within-species biodiversity. The ongoing growth in genomic sequencing technology will provide increasingly detailed estimates of overall genomic relationships between populations and allow sophisticated tests of whether genealogies based on uniparentally inherited markers have been shaped by selection.

Acknowledgments

I thank A. Brelsford, M. Doebeli, G. Irwin, J. Irwin, J. Lee-Yaw, S. Otto, B. Payseur, T. Price, T. Schulte, D. Toews, M. Whitlock, and two anonymous reviewers for providing helpful comments and/or discussion. This work was supported by Discovery Grants (311931-2005 and 311931-2010) to D.E.I. from the Natural Sciences and Engineering Research Council of Canada. I dedicate this article to Ward Watt, Professor of Biology at Stanford University, who first introduced me to the importance of natural selection along ecological gradients in shaping intraspecific genetic variation.

Literature Cited

Avise, J. C. 2000. Phylogeography: the history and formation of species. Harvard University Press, Cambridge, MA.

———. 2004. Molecular markers, natural history and evolution. 2nd ed. Sinauer, Sunderland, MA.

Avise, J. C., C. Giblin-Davidson, J. Laerm, J. C. Patton, and R. A. Lansman. 1979. Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. Proceedings of the National Academy of Sciences of the USA 76:6694–6698.

Ballard, J. W. O., and M. Kreitman. 1995. Is mitochondrial DNA a strictly neutral marker? Trends in Ecology & Evolution 10:485–488.

Ballard, J. W. O., and M. C. Whitlock. 2004. The incomplete natural history of mitochondria. Molecular Ecology 13:729–744.

Barton, N. H., and I. Wilson. 1995. Genealogies and geography. Philosophical Transactions of the Royal Society B: Biological Sciences 349:49–59.

Bazin, E., S. Glémin, and N. Galtier. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science 312: 570–572.

Birky, C. W., Jr., P. Fuerst, and T. Maruyama. 1989. Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics 121:613–627.

Brelsford, A., B. Milá, and D. E. Irwin. 2011. Hybrid origin of Audubon's warbler. Molecular Ecology 20:2380–2389.

Case, T. J., and M. L. Taper. 2000. Interspecific competition, environmental gradients, gene flow, and the coevolution of species' borders. American Naturalist 155:583–605.

Chen, B., P. M. Pedro, R. E. Harbach, P. Somboon, C. Walton, and R. K. Butlin. 2011. Mitochondrial DNA variation in the malaria vector *Anopheles minimus* across China, Thailand and Vietnam: evolutionary hypothesis, population structure and population history. Heredity 106:241–252.

Cheviron, Z. A., and R. T. Brumfield. 2009. Migration-selection balance and local adaptation of mitochondrial haplotypes in rufous-collared sparrows (*Zonotrichia capensis*) along an elevational gradient. Evolution 63:1593–1605.

Dalziel, A. C., S. M. Rogers, and P. M. Schulte. 2009. Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology. Molecular Ecology 18:4997–5017.

Das, J. 2006. The role of mitochondrial respiration in physiological and evolutionary adaptation. BioEssays 28:890–901.

Desjardins, P., and R. Morais. 1990. Sequence and gene organization of the chicken mitochondrial genome. Journal of Molecular Biology 212:599–634.

Dhillon, R. S., and P. M. Schulte. 2011. Intraspecific variation in the thermal plasticity of mitochondria in killifish. Journal of Experimental Biology 214:3639–3648.

Doebeli, M., and U. Dieckmann. 2003. Speciation along environmental gradients. Nature 421:259–264.

Dowling, D. K., K. Chávez Abiega, and G. Arnqvist. 2007. Temperature-specific outcomes of cytoplasmic-nuclear interactions on egg-to-adult development time in seed beetles. Evolution 61:194–201.

Dowling, D. K., U. Friberg, and J. Lindell. 2008. Evolutionary implications of non-neutral mitochondrial genetic variation. Trends in Ecology & Evolution 23:546–554.

Drummond, A. J., G. K. Nicholls, A. G. Rodrigo, and W. Solomon. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161:1307–1320.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7: 214.

Fangue, N. A., J. G. Richards, and P. M. Schulte. 2009. Do mitochondrial properties explain intraspecific variation in thermal tolerance? Journal of Experimental Biology 212:514–522.

Felsenstein, J. 1975. A pain in the torus: some difficulties with models of isolation by distance. American Naturalist 109:359–368.

Garvin, M. R., J. P. Bielawski, and A. J. Gharrett. 2011. Positive Darwinian selection in the piston that powers proton pumps in Complex I of the mitochondria of Pacific salmon. PLoS ONE 6: e24127, doi:10.1371/journal.pone.0024127.

Hein, J., M. H. Schierup, and C. Wiuf. 2005. Gene genealogies, variation, and evolution: a primer in coalescent theory. Oxford University Press, Oxford.

Hey, J. 2005. On the number of new world founders: a population genetic portrait of the peopling of the Americas. PLoS Biology 3: e193, doi:10.1371/journal.pbio.0030193.

Hickerson, M. J., B. C. Carstens, J. Cavender-Bares, K. A. Crandall, C. H. Graham, J. B. Johnson, L. Rissler, P. F. Victoriano, and A. D. Yoder. 2010. Phylogeography's past, present, and future: 10 years after Avise, 2000. Molecular Phylogenetics and Evolution 54:291–301.

Howell, N., C. B. Smejkal, D. A. Mackey, P. F. Chinnery, D. M. Turnbull, and C. Herrnstadt. 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. American Journal of Human Genetics 72:659–670.

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

Hurst, G. D. D., and F. M. Jiggins. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proceedings of the Royal Society B: Biological Sciences 272:1525–1534.

Irwin, D. E. 2002. Phylogeographic breaks without geographic barriers to gene flow. Evolution 56:2383–2394.

Irwin, D. E., S. Bensch, J. H. Irwin, and T. D. Price. 2005. Speciation by distance in a ring species. Science 307:414–416.

Irwin, D. E., A. S. Rubtsov, and E. N. Panov. 2009. Mitochondrial introgression and replacement between yellowhammers (*Emberiza citrinella*) and pine buntings (*E. leucocephalos*; Aves, Passeriformes). Biological Journal of the Linnean Society 98:422–438.

Kapralov, M. V., and D. A. Filatov. 2006. Molecular adaptation during adaptive radiation in the Hawaiian endemic genus *Schiedea*. PLoS ONE 1:e8, doi:10.1371/journal.pone.0000008.

Kingman, J. F. C. 1982*a*. The coalescent. Stochastic Processes and Their Applications 13:235–248.

———. 1982*b*. On the genealogy of large populations. Journal of Applied Probability 19:27–43.

Kirkpatrick, M., and N. H. Barton. 1997. Evolution of a species' range. American Naturalist 150:1–23.

Kuo, C.-H., and J. C. Avise. 2005. Phylogeographic breaks in low-dispersal species: the emergence of concordance across gene trees. Genetica 124:179–186.

Lane, N. 2008. Low variability on the W chromosome in birds is more likely to indicate selection on mitochondrial genes. Heredity 100:444–445.

Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard. 2009. Bayesian phylogeography finds its roots. PLoS Computational Biology 5:e1000520, doi:10.1371/journal.pcbi.1000520.

Lynch, M. 1996. Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. Molecular Biology and Evolution 13:209–220.

Maia, L. P., A. Colato, and J. F. Fontanari. 2004. Effect of selection on the topology of genealogical trees. Journal of Theoretical Biology 226:315–320.

Myouga, F., K. Akiyama, R. Motohashi, T. Kuromori, T. Ito, H. Iizumi, R. Ryusui, T. Sakurai, and K. Shinozaki. 2010. The chloroplast function database: a large-scale collection of *Arabidopsis Ds/Spm*- or T-DNA-tagged homozygous lines for nuclear-encoded chloroplast proteins, and their systematic phenotype analysis. Plant Journal 61:529–542.

Neuhauser, C., and S. M. Krone. 1997. The genealogy of samples in models with selection. Genetics 145:519–534.

O'Fallon, B. D., J. Seger, and F. R. Adler. 2010. A continuous-state coalescent and the impact of weak selection on the structure of gene genealogies. Molecular Biology and Evolution 27:1162–1172.

Pease, C. M., R. Lande, and J. J. Bull. 1989. A model of population growth, dispersal and evolution in a changing environment. Ecology 70:1657–1664.

Rand, D. M. 2008. Mitigating mutational meltdown in mammalian mitochondria. PLoS Biology 6:e35, doi:10.1371/journal.pbio.0060035.

Rauch, E. M., and Y. Bar-Yam. 2004. Theory predicts the uneven distribution of genetic diversity within species. Nature 431:449–452.

Ribeiro, Å. M., P. Lloyd, and R. C. K. Bowie. 2011. A tight balance between natural selection and gene flow in a southern African arid-zone endemic bird. Evolution 65:3499–3514.

Santos, C., R. Montiel, A. Arruda, L. Alvarez, M. P. Aluja, and M. Lima. 2008. Mutation patterns of mtDNA: empirical inferences for the coding region. BMC Evolutionary Biology 8:167, doi:10.1186/1471-2148-8-167.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, Oxford.

Slatkin, M. 1973. Gene flow and selection in a cline. Genetics 75:733–756.

Stewart, J. B., C. Freyer, J. L. Elson, A. Wredenberg, Z. Cansu, A. Trifunovic, and N.-G. Larsson. 2008. Strong purifying selection in transmission of mammalian mitochondrial DNA. PLoS Biology 6:e10, doi:10.1371/journal.pbio.0060010.

Takahata, N. 1990. A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proceedings of the National Academy of Sciences of the USA 87:2419–2423.

Templeton, A. R. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Molecular Ecology 7:381–397.

Towes, D. P. L., and A. Brelsford. Forthcoming. The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology.

Wakeley, J. 2008. Coalescent theory: an introduction. Roberts, Greenwood Village, CO.

Wallace, D. C. 2007. Why do we still have a maternally inherited mitochondrial DNA? insights from evolutionary medicine. Annual Review of Biochemistry 76:781–821.

Wilkins, J. F. 2004. A separation-of-timescales approach to the coalescent in a continuous population. Genetics 168:2227–2244.

Editor: Ruth G. Shaw

A greenish warbler *Phylloscopus trochiloides*. Photograph by Darren Irwin.